International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Investigating Stability of a Mutant Version of the REG3A Protein using FoldX and PIC

Aryan Prakash

Abstract: ***Quantitative analysis of stability of proteins due to different versions - the mutant and wild type versions - of the same gene would give us insight into the possible physiological effects of the particular mutation. This analysis was carried out using online tools such as PIC server and FoldX. The stability of REG3A protein and its mutant versions were compared to find that the mutant version due the replacement of nucleotide C with nucleotide T was a stable protein relative to the wild type REG3A.***

Keywords: Protein stablity, mutant protein, wild-type protein, Gibbs free energy, mutations

1. Introduction

The Avesthagenome Project is a project which studies and analyses the Parsi genome based on genetic samples collected from over 5000 Parsis in India. This genetic data is very significant for a few reasons. Firstly, the Parsis have a very small population and a sample size of 5000 people is nearly 1/10th of the Parsi population in India, hence, it is a relatively large sample size. Secondly, the Parsis have always maintained a tradition of in-breeding within the community, which means that Parsis only mate with Parsis and with nobody from an outside community. This is very significant biologically, for the reason that in-breeding in a small community can cause many traits to be passed on and expressed very frequently within the community's offspring. As a result, recessive disorders and traits are also expressed very frequently. The samples were collected in various centres across India and then were genetically sequenced through a technology known as Next-Generation Sequencing (NGS). There are primarily two methods of genome sequencing in use today (although there have been other types in the past). The two types in use today are short read sequencing and long read sequencing. In the short read sequencing, the genome is broken up into many various small parts and then the DNA is read through a series of steps which involves sequencing by synthesis (SBS). Here, the DNA is read by labelling nucleotides which are then read using Illumina. A particular challenge in short read sequencing is rearranging the fragments to get the sequence of the DNA nucleotides. This is one of the drawbacks of NGS.

The genetic data was collected on the Congenica platform to facilitate an organised analytical study of the sequenced genes for Avesthagen Limited. In this project, the gene REG3A has been studied and the proteins for which the wild-type (WT) gene and one of the particular mutant (MT) gene code for have been studied, their 3D structures have been predicted with the help of various softwares (described below) and a comparative study of the local thermodynamic features of the WT and MT proteins were carried out to predict which protein was more stable, and hence predict what kind of a physiological effect would this mutation have on the body.

2. Materials and Methods

The amino acid sequence of the WT protein was obtained from the UniProt online database (1) which gave the amino acid sequence based on the prediction by the AlphaFold software (2). The Congenica server provided us with the information collected on the amino acid change which would take place (or not) due to the SNP (single nucleotide polymorphism) in the REG3A gene in the particular patient the sample was collected from. The particular SNP being studied in this case was a change in the base at position 185 from C (cytosine) to T (thymine). The change in the amino acid due to this SNP was from amino acid threonine to the amino acid isoleucine.

We first downloaded the PDB (protein database) file of the original REG3A protein from the UniProt database. This file could be visualised using the PyMol software (3) as shown below in *Fig 1.*

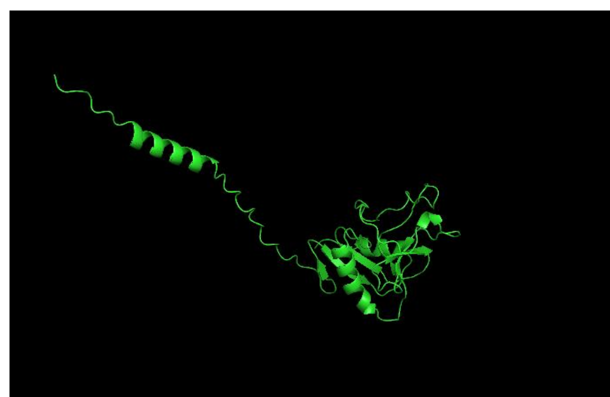

Figure 1: A representation of the WT protein as modelled by AlphaFold visualised in the PyMol software.

Then we used a series of commands in the terminal command of the computer with the FoldX software to give us a repaired version of the PDB, to optimise the downloaded protein structure and bring it to its minimum energy state. After generating this "repaired" protein structure, we input the specific change in the amino acid sequence (i.e, cytosine to thymine) in the downloaded file of the amino acid sequence. Thereafter, we entered a series of commands and used the software FoldX (4) to give us a visualisation in the form of a PDB file of the new mutant protein structure. The FoldX, upon being prompted, generated a PDB file of the mutant which could be visualised in PyMol. Furthermore, upon entering another

Licensed Under Creative Commons Attribution CC BY

series of computer commands, we obtained a file generated by the FoldX software, the "diff file", which gave us a detailed comparison of the change in Gibbs free energy of the WT and mutant proteins. The FoldX software calculated the ΔΔ*G (change in Gibbs free energy)* of the structures, which is simply ΔΔG = ΔGmutant - ΔGwild type. The change in Gibbs free energy of a protein gives us insight into the stability of the protein. For the individual protein, the more negative the value Δ*G,* the more stable a protein is. Therefore, in case the value of ΔΔG is negative, then we can conclude that the mutant protein is more thermodynamically stable than the WT protein, as the Δ*G* of the mutant protein is more negative than the WT protein, hence it would be a stabilising mutation. In case it is positive we can say the vice versa.. The two structures of the WT and mutant protein can also be easily visualised in the PyMol software as shown in Fig 1.

Additionally, we also used the PIC or Protein Interactions Calculator web server (5) developed by IISC Bangalore to model the intra protein interactions and their effect on the protein's stability. The PIC server looks at various parameters such as hydrogen bond interactions with respect to various different chains, disulphide bridges, ionic interactions and so on in order to determine the energy of each intraprotein interatomic bond. Using the PIC server, we can also further confirm the veracity of the prediction which we make on the protein's stability vis-a-vis the ΔΔG calculated by the FoldX software.

3. Results

Using FoldX -
The diff file output using the FoldX gave us the value of ΔΔG as -0.975801 J mol^-1 meaning that the mutation was indeed stabilising because the value of the ΔG for the mutant was lesser than the value of ΔG for the wild type protein.

Using PIC web server
The results show, only amino acid variations have occurred, and there is no change in the values of distances between 2 atoms and no gain/loss of interactions of any form have occurred as a result of this missense mutation.

4. Conclusion

Hence, using the results from FoldX and PIC, we can reasonably predict that this mutation stabilised the REG3A protein further. This has important physiological implications as the REG3A (Regenerating Islet Derived 3-alpha) is shown to have bactericidal activity, specifically against Gram positive bacteria. It has been shown (5) that the human REG3A binds to phospholipids in bacteria and creates pores in their membrane to kill the bacteria. The effects of a stabilising mutation on this protein must be investigated further to further elaborate on its physiological impacts.

References

[1] Wang Y, Wang Q, Huang H, Huang W, Chen Y, McGarvey PB, Wu CH, Arighi CN; UniProt Consortium. A crowd sourcing open platform for literature curation in UniProt. PLoS Biol. 2021 Dec 6;19(12):e3001464. doi: 10.1371/journal.pbio.3001464. PMID: 34871295; PMCID: PMC8675915.

[2] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583-589. doi: 10.1038/s41586-021-03819-2. Epub 2021 Jul 15. PMID: 34265844; PMCID: PMC8371605.

[3] Buß O, Rudat J, Ochsenreither K. FoldX as Protein Engineering Tool: Better Than Random Based Approaches? Comput Struct Biotechnol J. 2018 Feb 3;16:25-33. doi: 10.1016/j.csbj.2018.01.002. PMID: 30275935; PMCID: PMC6158775.

[4] Tina KG, Bhadra R, Srinivasan N. PIC: Protein Interactions Calculator. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W473-6. doi: 10.1093/nar/gkm423. Epub 2007 Jun 21. PMID: 17584791; PMCID: PMC1933215.

[5] Mukherjee S, Zheng H, Derebe MG, Callenberg KM, Partch CL, Rollins D, Propheter DC, Rizo J, Grabe M, Jiang QX, Hooper LV. Antibacterial membrane attack by a pore-forming intestinal C-type lectin. Nature. 2014 Jan 2;505(7481):103-7. doi: 10.1038/nature12729. Epub 2013 Nov 20. PMID: 24256734; PMCID: PMC4160023.

Licensed Under Creative Commons Attribution CC BY