Macedonian pharmaceutical bulletin, 69 (Suppl 1) 273 - 274 (2023)
Online ISSN 1857 - 8969
DOI: 10.33320/maced.pharm.bull.2023.69.03.133
Short communication

Principal Component Analysis to evaluate the stability impact of protein mutations: the case of SARS-CoV-2 K417T mutation

Done Stojanov[1]*, Marija Glavas Dodov[2], Aleksandar Velinov[1], Martina Nestorovska[2]

[1]Faculty of Computer Science, Goce Delcev University of Stip, Krste Misirkov No.10-A P.O. Box 201, 2000 Stip, N. Macedonia
[2]Faculty of Pharmacy, Ss. Cyril and Methodius University in Skopje, Majka Tereza 47, 1000 Skopje, N. Macedonia

Introduction

The severe acute respiratory syndrome CoV-2 (SARS-CoV-2), which was initially identified in the Wuhan Province, China spread worldwide rapidly. The intense escalation forced the WHO to declare a pandemic with 6.5 million deaths worldwide.

The SARS-CoV-2 virus has a wide host range, as it uses the angiotensin-converting enzyme 2 (ACE2) as a target receptor in humans. A 211 amino acid region at the C-terminal domain of the S1 subunit of the coronavirus as the receptor/binding domain was identified through structural and biochemical analyses. This plays a crucial role in virus entry and is the main target of the host immune responses. The RBD mediates contact with the ACE2 receptor and RBD region in SARS-CoV-2 varies from the one in other SARS-CoV viruses in the five residues critical for ACE2 binding. As a result of these changes, the interaction of SARS-CoV-2 with its receptor stabilizes the two virus-binding hotspots on the surface of hACE2. Moreover, four-residue motif in the RBM of SARS-CoV-2 leads to a more compact conformation of its hACE2-binding bridge. While the SARS-CoV-2 S protein lost some of its key mutations which are associated with higher infectivity (Stojanov, 2021), many SARS-CoV-2 variants possess stronger virulence and infectivity and can produce immune escape. Several RBD residues mutated independently in multiple lineages. The RBD residues 331–524 of the Spike protein have been a prime focus in many studies. These mutations include N501Y in the Alpha, Beta, Gamma, and Omicron variants; K417 mutated to N in the Beta, Delta, and Omicron variants and to T in the Gamma variant, and E484 mutated to K in the Beta and Gamma variants and to A in the Omicron variant.

The purpose of this study was to examine and analyse the impact of K417T mutation upon SARS-CoV-2 S-protein/hACE2 complex stability through the process of PCA analysis.

Materials and methods

In order to evaluate the impact of K417T mutation upon SARS-CoV-2 S-protein/hACE2 complex stability, we induced K417T mutation in SARS-CoV-2 S-protein, PDB heterodimer 6M0J (https://www.rcsb.org/structure/6m0j), using the PyMol software (https://pymol.org/2/). We used SPC216 water solvent model, having placed both systems: K417 wild type and T417 mutant into a cubic solute box. Both systems were brought to a neutral net charge and they were energetically optimized ($F_{max} < 1000\ kJ\ mol^{-1}nm^{-1}$) by applying the steepest descent minimization algorithm. We used V-rescale thermostat to equilibrate systems' temperature at 310 K. The referent coupling pressure was set up to 1 bar, assuming for water isothermal compressibility 4.45×10^{-5} bar^{-1}. Each preparation step lasted 100 ps. Following successful preparation, systems underwent 50 ns molecular dynamics simulation in Gromacs molecular dynamics software (Abraham et al., 2015). We used simulation output files: xtc and tpr as fundamentals of our analysis. We applied PCA (The Principal Component Analysis) to evaluate the stability impact of K417T S-protein mutation. PCA is a dimensionality reduction technique, which can be applied for extraction of dominant modes in the overall molecule

*done.stojanov@ugd.edu.mk

motion (Wang et al., 2019). All atoms, slow, collective molecule motions are distinguished out of fast, local fluctuations by the means of PCA. PCA maps integral molecule movements per frame into linear vectors of orthogonal values, called principal components: PC1, PC2, that stand for the largest uncorrelated movements in the trajectory. Plots of the principal component values are used to evaluate molecule stability and detection of significant conformational shifts. Two Gromacs modules were used for the principal component analysis: gmx covar and gmx anaeig. The module gmx covar was used to compute covariance matrix upon C_α alpha-carbon atoms, while gmx anaeig to analyze eigenvectors. The projection of the first two principal components: PC1 and PC2 is plotted on Fig. 1.

Results and discussion

Fig. 1 shows the two-dimensional PCA plots for K417 (wild type system) and T417 (mutant complex). Data distribution over the first principal component PC1, represents most of the variance of molecule collective motion and is considered to be the most important factor when considering molecule stability. Narrow PC1 distribution stands for stable molecule behavior (restricted global molecule motions), while the opposite stands for destabilizing impact (certain degree of flexibility in global molecule motions observed).

Table 1. PC1, PC2 standard deviation (nm)

System	Std. dev. PC1 (nm)	Std. dev. PC2 (nm)
K417	2.082582267	1.307230045
T417	1.994043267	1.4426429

Following PCA results (Fig. 1), K417T mutation confers enhanced S-protein/hACE2 complex stability, given that PC1 distribution shrinks in T417 complex (orange scatter plot) relative to the wild type system (blue scatter plot). Apart from the visual examination of the PCA results (Fig. 1), the same conclusion can be derived empirically, by the means of standard deviation of PC1 of K417 and T417 systems, Table 1: std. dev. PC1 T417=1.994043267 nm < std. dev. PC1 K417=2.082582267 nm. Although the second principal component (PC2), in terms of the standard deviation, satisfies the opposite of the previous inequality, Table 1: std. dev. PC2 T417=1.4426429 nm > std. dev. PC2 K417=1.307230045 nm, PC1 models the vast of variance (most of the molecule uncorrelated movements) and that is the reason why we evaluated the stability impact of K417T substitution in terms of PC1 distribution.

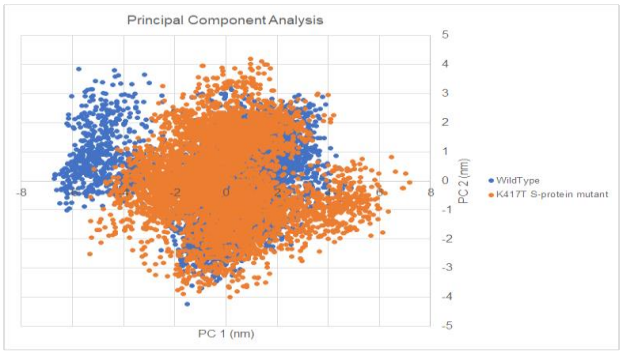

Fig. 1. PCA plots: K417 wild type system (blue scatter plot) and T417 mutant complex (orange scatter plot).

Conclusion

In this study, we have evaluated the stability impact of K417T SARS-CoV-2 S-protein mutation. We have shown that the Principal Component Analysis, which is a dimensionality reduction technique, can be successfully applied for that purpose. Although we computed/plotted the first two principal components: PC1 and PC2, the first principal component PC1, models the most of the molecule uncorrelated movements and therefore is suitable for examining the overall molecule stability due to induced mutations. Our in silico experimental results, showed that K417T substitution confers stabilizing effect upon SARS-CoV-2 S-protein/hACE2 complex.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1, 19-25. https://doi.org/10.1016/j.softx.2015.06.001.

Stojanov, D., 2021. Phylogenicity of B.1.1.7 surface glycoprotein, novel distance function and first report of V90T missense mutation in SARS-CoV-2 surface glycoprotein. Meta Gene, 30, p.100967.

Wang, Q., Mehmood, A., Wang, H., Xu, Q., Xiong, Y., Wei, D.Q., 2019. Computational screening and analysis of lung cancer related non-synonymous single nucleotide polymorphisms on the human kirsten rat sarcoma gene. Molecules 24(10), 1951. doi: 10.3390/molecules24101951.