# Principal Component Analysis to evaluate the stability impact of protein mutations: the case of SARS-CoV2 K417T mutation

14th Central European Symposium on Pharmaceutical Technology - CESPT 2023

**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, Stip 2000, N. Macedonia*

[2] *Faculty of Pharmacy, Ss. Cyril and Methodius University in Skopje, Majka Tereza 47, 1000 Skopje, N. Macedonia*

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. Several RBD residues mutated independently in multiple lineages. SARS-CoV-2 variants possess strong virulence and infectivity and can produce immune escape.

## Purpose

**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 $4.45\times10^{-5}$ $bar^{-1}$. 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 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 projection of the first two principal components: PC1 and PC2 is plotted on Fig. 1.

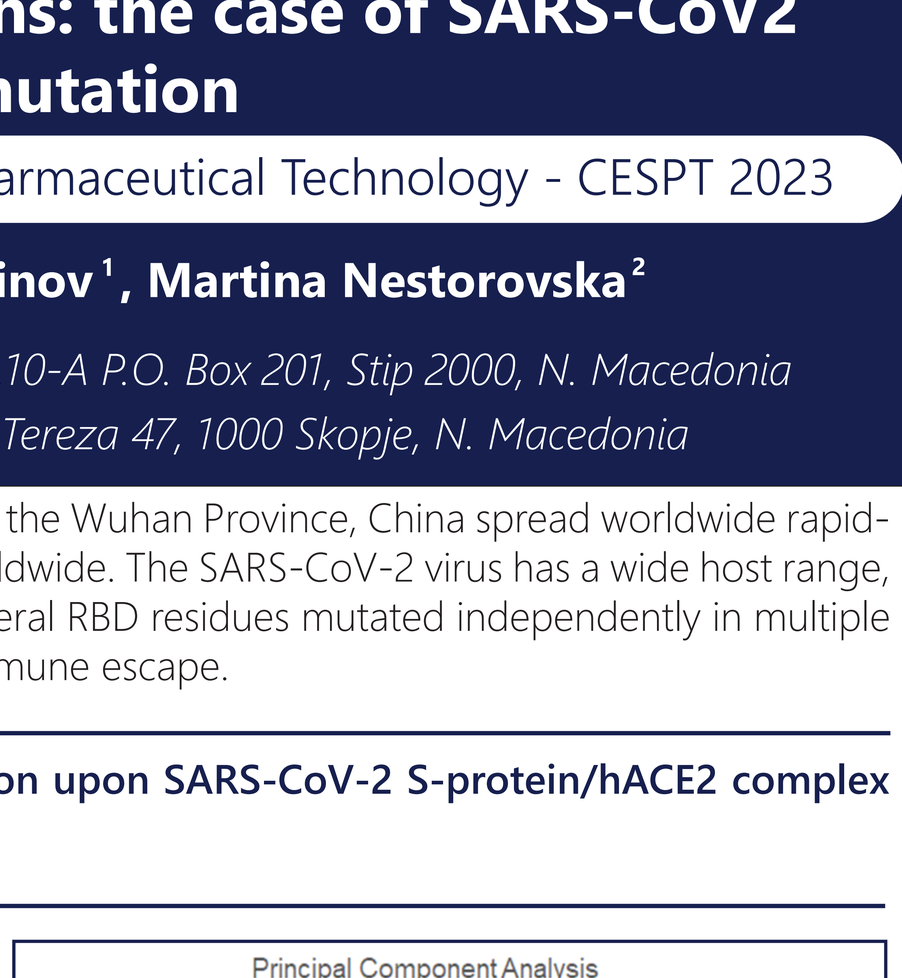


**Fig. 1. PCA plots: K417 wild-type system (blue scatter plot) and T417 mutant complex (orange scatter plot).**

## Results and discussion

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

**Table 1. PC1, PC2 standard deviation (nm)**

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). 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.

## 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.

## Related research

*Stojanov, D. (2023). Structural implications of SARS-CoV-2 Surface Glycoprotein N501Y mutation within receptor-binding domain [499-505]–computational analysis of the most frequent Asn501 polar uncharged amino acid mutations. Biotechnology & Biotechnological Equipment, 37(1), 2206492.*

*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, 100967.*

*Stojanov, D. (2021). Data on multiple SARS-CoV-2 surface glycoprotein alignments. Data in Brief, 38, 107414.*

*Goce Delcev University of Stip, Republic of N. Macedonia*
*Faculty of Computer Science*
**www.ugd.edu.mk**
**fi.ugd.edu.mk**

*Ss Cyril and Methodius University, Skopje, Republic of N. Macedonia Faculty of Pharmacy*
**www.ukim.edu.mk**
**ff.ukim.edu.mk**