ACE2 variants underlie interindividual variability and susceptibility to COVID-19 in Italian population

ACE2 gene variants in Italian population

Benetti Elisa[1ι], Tita Rossella[2ι], Spiga Ottavia[3], Ciolfi Andrea[4], Birolo Giovanni[5], Bruselles Alessandro[6], Doddato Gabriella[7], Giliberti Annarita[7], Marconi Caterina[8], Musacchia Francesco[9], Pippucci Tommaso[10], Torella Annalaura[11], Trezza Alfonso[3], Valentino Floriana[7], Baldassarri Margherita[7], Brusco Alfredo[5,12], Asselta Rosanna[13,14], Bruttini Mirella[2,7], Furini Simone[1], Seri Marco[8,10], Nigro Vincenzo[9,11], Matullo Giuseppe[5,12], Tartaglia Marco[4], Mari Francesca[2,7], Renieri Alessandra[2,7]*, Pinto Anna Maria[2]

[1] Department of Medical Biotechnologies, University of Siena, Siena, Italy

[2] Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy

[3] Dept. of Biotechnology, Chemistry and Pharmacy, University of Siena, Siena, Italy

[4] Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, Rome, Italy

[5] Department of Medical Sciences, University of Turin, Turin, Italy

[6] Department of Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, Italy

[7] Medical Genetics, University of Siena, Siena, Italy

[8] Department of Medical and Surgical Sciences, University of Bologna

[9] Telethon Institute of Genetics and Medicine, Pozzuoli, Italy

[10] Sant'Orsola-Malpighi University Hospital, Bologna

[11] Dipartimento di Medicina di Precisione, Università della Campania "Luigi Vanvitelli",
Napoli, Italy

[12] Genetica Medica, Città della salute e della scienza, Torino, Italy

[13] Department of Biomedical Sciences, Humanitas University, Rozzano, Milan, Italy
[14] Humanitas Clinical and Research Center - IRCCS, Rozzano, Milan, Italy

ɿ These authors contributed equally to this work

***Corresponding author**

Professor Alessandra Renieri
Medical Genetics Unit
University of Siena
Policlinico Le Scotte
Viale Bracci, 2
53100 Siena, Italy
Phone: 39 0577 233303
FAX 39 0577 233325
E.mail: alessandra.renieri@unisi.it

ABSTRACT

In December 2019, an initial cluster of unexpected interstitial bilateral pneumonia emerged in Wuhan, Hubei province. A human-to-human transmission was immediately assumed and a previously unrecognized entity, termed coronavirus disease 19 (COVID-19) due to a novel coronavirus (2019-nCov) was suddenly described. The infection has rapidly spread out all over the world and Italy has been the first European Country experiencing the endemic wave with unexpected clinical severity in comparison with Asian countries.

It has recently been shown that 2019-nCov utilizes host receptors namely angiotensin converting enzyme 2 (ACE2) as host receptor and host proteases for cell surface binding and internalization. Thus, a predisposing genetic background can give reason for interindividual disease susceptibility and/or severity. Taking advantage of the Network of Italian Genomes (NIG), here we mined around 7000 exomes from 5 different Centers looking for *ACE2* variants. A number of variants with a potential impact on protein stability were identified. Among these, three missense changes, p.Asn720Asp, p.Lys26Arg, p.Gly211Arg (MAF 0.002 to 0.015), which have never been reported in the Eastern Asia population, were predicted to interfere with protein u and stabilization. Rare truncating variants likely interfering with the internalization process and one missense variant, p.Trp69Cys, predicted to interfere with 2019-nCov spike protein binding were also observed.

These findings suggest that a predisposing genetic background may contribute to the observed inter-individual clinical variability associated with COVID-19. They allow an evidence-based risk assessment opening up the way to personalized preventive measures and therapeutic options.

Keywords:

COVID-19

ACE2 variants

Italian population

Molecular dynamics

INTRODUCTION

In December 2019, a new infectious respiratory disease emerged in Wuhan, Hubei province, China (1–3). An initial cluster of infections likely due to animal-to-human transmission was rapidly followed by a human-to-human transmission (4). The disease was recognized to be caused by a novel coronavirus (2019-nCov) and termed coronavirus disease 19 (COVID-19). The infection spread within China and all over the world, and it has been declared as pandemic by the World Health Organization (WHO) on 2nd March 2020. The symptoms of COVID-19 range from fever, dry cough, fatigue, u congestion, sore throat and diarrhea to severe interstitial bilateral pneumonia with a ground–glass image at the CT scan. While recent studies provide evidence of a high number of asymptomatic or paucisymptomatic patients who represent the main reservoir for the infection progression, the severe cases can rapidly evolve towards a respiratory distress syndrome which can be lethal (5). Although age and comorbidity have been described as the main determinants of disease progression towards severe respiratory distress, the high variation in clinical severity among middle-age adults and children would likely suggest a strong role of the host genetic asset.

A high sequence homology has been shown between SARS-associated coronavirus (SARS-CoV) and 2019-nCov (6). Recent studies modelled the spike protein to identify the receptor for 2019-nCov, and indicated that angiotensin converting enzyme 2 (ACE2) is the receptor for this novel coronavirus (7,8). Zhou et. al. conducted virus infectivity studies and showed that ACE2 is essential for 2019-nCov to enter HeLa cells (9). Although the binding strength between 2019-nCov and ACE2 is weaker than that between SARS-CoV and ACE2, it is considered as much high as threshold necessary for virus infection. The spike glycoprotein (S-protein), a trimeric glycoprotein in the virion surface (giving the name of crown -*corona* in latin-), mediates receptor recognition throughout its receptor binding domain (RBD) and membrane fusion (10,11). Based on recent reports, 2019-nCov protein binds to ACE2 through Leu455, Phe486, Gln493, Asn501, and Tyr505. It has been postulated that residues 31, 41, 82, 353, 355, and 357 of the ACE2 receptor map to the surface of the protein interacting with 2019-nCov spike protein (12), as previously documented for SARS-CoV. Following interaction, cleavage of the C-terminal segment of ACE2 by proteases, such as transmembrane protease serine 2 (TMPRSS2), enhances the spike protein-driven viral entry (13,14). Thus, it is possible, in principle, that genetic variability of the ACE2 receptor is one of the elements modulating virion intake and thus disease severity.

Taking advantage of the Network of Italian Genomes (NIG), a consortium established to generate a public database (NIG-db) containing aggregate variant frequencies data for the Italian population (http://www.nig.cineca.it/), here we describe the genetic variation of *ACE2* in the Italian population, one of the newly affected countries by the 2019-nCov outbreak causing COVID-19. Three common (p.Lys26Arg, p.Gly211Arg, p.Asn720Asp) and 30 rare missense variants were identified, 12 of which had not

previously been reported in public databases. We show that p.Asn720Asp, which affect a residue located close to the cleavage sequence of TMPRSS2, likely affects the cleavage-dependent virion intake. Along with the other two common variants, this substitution is significantly represented in the Italian and European populations but is extremely rare in the Asian population. We also show that three rare variants, namely, p.Trp69Cys, p.Leu351Val and p.Pro389His are predicted to cause conformational changes impacting RBD interaction. As the uncertainty regarding the transmissibility and severity of disease rise, we believe that a deeper characterisation of the host genetics and functional characterization of variants may help not only in understanding the pathophysiology of the disease but also in envisaging risk assessment.

Materials and Methods

Population

Parents provided signed informed consents at each participating center for exome sequencing analysis, and clinical and molecular data storage and usage, for both diagnostic and research purposes. The work has been realized in the context of NIG, with the contribution of centers: Azienda Ospedaliera Universitaria Senese, Azienda Ospedaliera-Universitaria Policlinico Sant'Orsola-Malpighi di Bologna, Città della Salute e della Scienza di Torino, Università della Campania "Luigi Vanvitelli", Ospedale Pediatrico Bambino Gesù. All subjects were unrelated, apparently healthy, and of Italian ancestry.

Whole Exome Sequencing

Targeted enrichment and massively parallel sequencing were performed on genomic DNA extracted from circulating leukocytes of 6984 individuals. Genomic DNA was

extracted from peripheral blood samples using standard procedures. Exome capture was carried out using SureSelect Human All Exon V4/V5/V6/V7 (Agilent Technologies, Santa Clara, CA), Clinical Research Exome V1/V2 (Agilent), Nextera Rapid Capture v.1.2 (Illumina, San Diego, CA), TruSeq Exome Targeted Regions (Illumina, San Diego, CA), TruSight One Expanded V2 (Illumina, San Diego, CA), Sequencing-by-Synthesis Kit v3/v4 (Illumina, San Diego, CA) or HiSeq 2000 v2 Sequencing-by-Synthesis Kit (Illumina, San Diego, CA) and sequencing was performed on Genome Analyzer (v3/v4)/HiSeq2000/NextSeq550/NextSeq500/Novaseq6000 platforms (Illumina, San Diego, CA). A subset of WES had been outsourced (BGI, Shenzen China; Mount Sinai, New York, USA; Broad Institute, Harvard, USA) . Reads alignment and variant calling and annotation were attained using in-house pipelines implemented from which previously reported ones (15–18), which mainly take advantage of the Genome Analysis Toolkit (GATK V.3.7) framework (19).The genome aggregation database gnomAD (https://gnomad.broadinstitute.org/) was used to assess allele frequency for each variant among different populations.

Computational studies

The structure of native human angiotensin converting enzyme-related carboxypeptidase (ACE2) was downloaded from Protein Data Bank (https://www.rcsb.org/) (PDB ID code 1R42) (20). The DUET program (21) was used to predict the possible effect of amino acids substitutions on the protein structure and function, based on the use of machine-learning algorithms exploiting the three dimensional structure to quantitatively predict the effects of residue substitutions on protein functionality. Molecular dynamics (MD) simulations of wild-type and variant ACE2 proteins were carried out in GROMACS 2019.3 (22) to calculate root-mean-square deviation (RMSD) to define

structural stability. The graphs were plotted by the XMGrace software (23). MD simulations were performed by LINUX Cluster having 660 cpu on 21 different nodes, 190T of RAM, 30T hard disk partition size and 6 NVIDIA TESLA gpu with CUDA support. PyMOL2.3 was used as a molecular graphic interface. The protein structures were solvated in a triclinic box filled with TIP3P water molecules and Na +/Cl - ions were added to neutralize the system. The whole systems were then minimized with a maximal force tolerance of 1000 kJ mol-1 nm-1 using the steepest descendent algorithm. The optimized systems were gradually heated to 310 K in 1 ns in the NVT ensemble, followed by 10 ns equilibration in the NPT ensemble at 1 atm and 310 K, using the V-Rescale thermostat and Berendsen barostat (24,25). Subsequently, a further 100 ns MD simulations were performed for data analysis.

RESULTS

The extent of variability along the entire *ACE2* coding sequence and flanking intronic stretches was assessed using 6984 Italian exomes. Identified variants and predicted effects on protein stability are summarized in Table 1, Table S1 and Table 2, and represented in Fig. 1. Three variants with MAF between 0.015 and 0.002 (p.Lys26Arg, p.Gly211Arg and p.Asn720Asp) were identified. The p.Asn720Asp substitution was estimated to have a frequency of 0.01 (105/6984), which is quite in line with the frequency of the variant reported in the EXAC database (0.016), and is lower than the frequency reported for the European non-Finnish population, in gnomAD (0.025). All analysed females were heterozygotes for the variant. Notably, this variant has not been reported in the Eastern Asia population (13,784 exomes). p.Lys26Arg and p.Gly211Arg were found with a frequency of 0.001 (lower than the frequency in the European-non

Finnish population, 0.0058) and 0.0011 (European-non Finnish population frequency, 0.0019), respectively. Out of ~ 41,000 individuals, one homozygous female has been reported for the p.Lys26Arg while no homozygous females have been reported for the p.Gly211Arg. According to gnomAD, the allele frequency of the p.Lys26Arg variant in the Eastern Asia population was 0.00006, while the p.Gly211Arg has not been reported in 14,822 exomes. In addition to these variants, 30 rare missense variants were identified, out of which 12 had not previously been reported in GnomAD database and 13 truncating variants that had not been reported in gnomAD database (Tab 1, Supplementary Table 1). Out of these variants, two fall in the neck domain, which is essential for dimerization and one in the intracellular domain. Many of them truncate the protein in different positions of the Protease domain embedded in the extracellular domain, which contains the receptor binding site for COVID-19. Only three truncating mutations have been previously described for *ACE2* likely due to a low-tolerance for loss-of function variants, in line with this quite all of them (11/13) were detected in women in heterozygous state. Three missense changes p.Val506Ala, p.Val209Gly, p.Gly377Glu were predicted to have a destabilizing structural consequences (Figure 1, Table 2); among these, p.Val506Ala is indeed the only amino acid change reported in the European non-Finnish population (rs775181355; frequency: 0.000006561, cadd 27,2) and is predicted as probably damaging for the protein structure by Polyphen and deleterious by SIFT .

Similarly, p.Trp69Cys, p.Leu351Val and p.Pro389His, which affects a highly hydrophobic core, were predicted to induce conformational changes influencing the interaction with spike protein. The aminoacidic substitution p.Pro389His (rs762890235, European-non Finnish population allele frequency: 0.00002453, cadd 24,8) was

predicted to be probably damaging by Polyphen and deleterious by SIFT. Also this rare variant has never been reported in Asian populations. We analysed by MD simulations the structural changes of two specific missense mutations: the V506A that showed the higher destabilizing effect and the W69C whose surrounding residues, (30-41) segment, are involved in 2019-nCOV Spike protein interaction. To analyse differences in protein structure between wild type and mutants (V506A and W69C), we performed 100ns MD simulations. The comparison was performed by RMSD. The global effects of the residue substitutions on flexibility and global correlated motion of ACE2 protein are represented in Figure 2 and the simulation is provided in Supplementary Figure 1 and 2. While a similar trend for wild-type and V506A is observed with a steady course in the RMSD value, which stabilizes at an average of 0.2 nm and 0,3 nm, respectively (Figure 3, panel A) the W69C variant shows a difference in comparison with the native protein with a gradual increase in RMSD value, which stabilizes at an average of 0.5 nm (Figure 3, panel B). During MD simulations, we have also investigated the surrounding region of ACE2 WT and W69C variant by calculating change in Solvent Accessibility Surface Area (SASA). Structural analysis between WT and V506A mutant MD simulations showed that Val506 forms a hydrophobic centre together with Leu456, Leu503 and Phe516 with minimum differences in protein rearrangements when the residue is mutated in Ala as reported in Figure 3, panel A and confirmed by SASA superimposition (Figure 3, panel B). Trp69 together with Phe70 and Leu71 is normally involved in a hydrophobic core with Leu38 and Phe39 of (30-41) 2019-nCOV interaction region and, on the contrary, a drastic rearrangement of protein structure is observed when, as consequence of the amino acid change (Trp in Cys), its central role is substituted by Phe39 as confirmed by SASA superimposition (Figure 3, panel B,

Supplementary Figure 2). Differences in average SASA value would suggest for the native protein a wider surface exposed to solvent and subsequently a different ability to interact with spike 2019-nCov in comparison with the variants.

DISCUSSION

According to recent reports, ACE2 is essential for 2019-nCov to enter cells. Recent single-cell RNA studies have also shown that ACE2 is expressed in human lung cells (26). The majority of ACE2-expressing cells are alveolar type 2 (AT2) cells. Other ACE2 expressing cells include alveolar type 1 (AT1) cells, airway epithelial cells, fibroblasts, endothelial cells, and macrophages although their ACE2-expressing cell ratio is low and variable among individuals. The expression and distribution of the ACE2 receptor can thus justify the route of infection and the main localization at the alveolar level. Although the different density of ACE2 receptors in the upper respiratory tract among individuals can partially give reason for the clinical variability, which ranges from asymptomatic/paucisymptomatic patients to severely affected ones, it cannot be the only reason for this variability. In addition, recent works did not observe significantly different viral loads in nasal swabs between symptomatic and asymptomatic patients (27). Italy has been the first European country that experienced the COVID-19 outbreak with a rapid increase in the positive cases in a very short-time period and a morbidity and lethality (~10%) definitely higher in comparison with Asian countries, such as China (4%) and South Chorea (1.2%) according to the OMS reports. These considerations raise the the possibility of a predisposing genetic background accounting for or contributing to the wide inter-individual clinical variability, as well

for the differential morbidity and lethality observed among different countries, population awareness and constrictive measures apart.

We integrated genomic data from 5 Italian centers (Siena, Naple, Turin, Bologna, Rome) interconnected by the Network for Italian Genome (NIG) in the attempt to identify variation encompassing the *ACE2* gene, which could account for a difference in 2019-nCov spike binding affinity, processing or internalization. Previous studies showed that the residues near lysine 31, and tyrosine 41, 82–84, and 353–357 in human ACE2 are important for the binding of S-protein to coronavirus (12). In line with previous reports (28), we did not find polymorphism or rare variants in these residues in the Italian population. However, we identified three variants p.Asn720Asp, p.Lys26Arg, p.Gly211Arg, one of which polymorphic (p.Asn720Asp), moderately expressed in the Italian and European-non Finnish populations and with a very low allele frequency or not occurring in the Eastern Asia population. These variants which surround residual essentials for the 2019-nCov spike binding are predicted to likely affect the cleavage dependent virion intake, such as the Asn720Asp which lies 4 amino acids from the cleavage sequence of TMPRSS2. Give their rarity in other populations, we cannot exclude that these variants can partially account for the higher morbidity and lethality observed in the Italian and European-non Finnish population. Indeed, while this manuscript was in preparation, COVID-19 was drastically spreading all over Europe severely impacting other Latin countries such as Spain (6% lethality) and France (4%). Along with these more common variants we also identified very rare variants, some of which only described in the non-Finnish European population. Some of them are able to affect a highly hydrophobic core inducing a conformational change impacting in RBD interaction and could give reason for a higher affinity for the 2019-nCov spike protein

(Figure 2). Among these the p.Trp69Cys has never been reported before and it is likely to be an Italian population-private rare variant. Other rare variants are predicted to truncate the protein in different positions of the Protease domain thus likely acting on the internalization process. These rare variants would likely account for the inter-individual clinical variability and likely explain severity even in young adults. Notably, morbidity and lethality have been reported definitely higher in men compared to women (~70% vs 30%, 20th March 2020 ISS report). Although several parameters have been brought to case to explain this difference, i.e. smoking, differences in ACE2 localization and/or density in alveolar cells, hormonal asset, it is noteworthy that ACE2 is located on chromosome X and that given the low allele frequency of the identified variants the rate of homozìgous women is extremely low (see Results section). Therefore, the impact of X-inactivation on the alternate expression of the two alleles would guarantee a heterogeneous population of ACE2 molecules, some of which protective towards the infection until the point of a complete or almost complete protection in the case of a X-inactivation skewed towards the less 2019-nCov -binding prone allele. This hypothesis would justify the high rate of asymptomatic or paucisymptomatic patients. *ACE2* is definitely one of the main molecules whose genetic heterogeneity can modulate infection and disease progression; however a deeper characterisation of the host genetics and functional variants in other pathway-related genes may help in understanding the pathophysiology of the disease opening up the way to a stratified risk assessment and to tailored preventive measures and treatments.

Acknowledgements

This work was, in part, supported by: Telethon Network of Genetic Biobanks (project no. GTB12001), funded by Telethon Italy; Fondazione Bambino Gesù (*Vite Coraggiose*

to M.T.); Mount Sinai, NY (USA) in the context of the international project ASC (Autism sequencing consortium); SOLVE-RD and MIUR project "Dipartimenti di Eccellenza 2018 – 2022" (n° D15D18000410001) to the Department of Medical Sciences, University of Torino (G.M., A.B.). We thank the CINECA consortium for providing computational resources.

Ethical Approval

The study was approved by the Ethical Commettee of each center , Prot Name NIG Prot n 10547_2016 , approved CEAVSE on 18.07.2016 n. 88/16

Conflict of interest

The authors declare that they have no competing interests.

Author's contribution

EB, RT, OS, AMP and AR have made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data and have been involved in drafting the manuscript. RA, GB, AB, AB, AT, GD, AG, FM, TP, AT, AT and FV has made substantial contributions to acquisition and analysis of the data. MB, MB, AC, SF, FM, GM, VN, MS and MT have made substantial contributions to interpretation of data and clinical evaluation. All authors have been involved in drafting the manuscript; have given final approval of the version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, *et al*. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;

2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;

3. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet. 2020.

4. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020;

5. Lai C-C, Liu YH, Wang C-Y, Wang Y-H, Hsueh S-C, Yen M-Y, *et al.* Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. J Microbiol Immunol Infect. 2020;

6. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 2020;

7. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, *et al*. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med. 2020;

8. Liu Y, Yan L-M, Wan L, Xiang T-X, Le A, Liu J-M, *et al.* Viral dynamics in mild and severe cases of COVID-19. Lancet Infect Dis [Internet].

2020;2019(20):2019–20. Available from: https://doi.org/10.1016/S1473-3099(20)30232-2

9. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;

10. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, *et al*. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science (80-). 2020;

11. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science. 2020;

12. Li W, Kuhn JH, Moore MJ, Wong SK, Huang IC, Farzan M, *et al*. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. 2005;

13. Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T. A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. J Virol. 2011;

14. Heurich A, Hofmann-Winkler H, Gierer S, Liepold T, Jahn O, Pohlmann S. TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. J Virol. 2014;

15. Magini P, Smits DJ, Vandervore L, Schot R, Columbaro M, Kasteleijn E, *et al*. Loss of SMPD4 Causes a Developmental Disorder Characterized by Microcephaly and Congenital Arthrogryposis. Am J Hum Genet. 2019;

16. Del Dotto V, Ullah F, Di Meo I, Magini P, Gusic M, Maresca A, *et al*. SSBP1 mutations cause mtDNA depletion underlying a complex optic atrophy disorder. J Clin Invest. 2020;

17. Flex E, Martinelli S, Van Dijck A, Ciolfi A, Cecchetti S, Coluzzi E, *et al.* Aberrant Function of the C-Terminal Tail of HIST1H1E Accelerates Cellular Senescence and Causes Premature Aging. Am J Hum Genet. 2019;

18. Kortüm F, Caputo V, Bauer CK, Stella L, Ciolfi A, Alawi M, *et al*. Mutations in KCNH1 and ATP6V1B2 cause Zimmermann-Laband syndrome. Nat Genet. 2015;

19. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GA Van der, *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2017;

20. Towler P, Staker B, Prasad SG, Menon S, Tang J, Parsons T, *et al*. ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis. J Biol Chem. 2004;

21. Pires DEV, Ascher DB, Blundell TL. DUET: A server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res. 2014;

22. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, *et al*. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;

23. Turner P. XMGRACE, Version 5.1. 19. Cent Coast Land-Margin Res Oregon Grad Inst Sci Technol Beavert. 2005;

24. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;

25. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;

26. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. bioRxiv. 2020;

27. D C, M T, F R, V D, M A, P P, *et al*. The early phase of the COVID-19 outbreak in Lombardy, Italy. 2020; Available from: http://arxiv.org/abs/2003.09320

28. Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, *et al*. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discovery. 2020.

Titles and legends to figures

Figure 1. ACE2 Crystal structure with PDB ID 1R42. ACE2 Crystal structure with PDB ID 1R42. Surface and cartoon representations of protein in grey. In blue single mutated positions, while in yellow (30-41), green (353-357) and red (82-84) respectively, residues of interaction with 2019-nCOv spike glycoprotein [Uniprot ID Q9BYF1

]**Figure 2: Structure superimposition snapshot between wild type protein and V506A and W69C mutant proteins**

Panel A. ACE2 wild type and mutant V506A superimposed structure after 100ns MD simulation. Cartoon representation of ACE2 wild type (green), in red sticks Val506 involved in a hydrophobic core with Leu456, Phe512 and Leu503 (green sticks). During simulation no particular differences in protein structure were detected. **Panel B.** ACE2 wild type and mutant W69C superimposed structure after 100ns MD simulation. Cartoon representation of ACE2 wild type (green), in red sticks Trp69 involved in a hydrophobic core with Leu38 and Phe39 (green sticks) and Phe70 and Leu71 (blue sticks). Cartoon representation of ACE2 W69C mutant (orange) with Cys69 (yellow sticks), Leu38 and Phe39 (green sticks) and Phe70 and Leu71 (light blue sticks). During simulation, Phe39 substituted the central role of Trp69 with consequently rearrangement of protein structure.

Figure 3. Structural analysis and MD simulations of WT, V506A and W69C mutants.

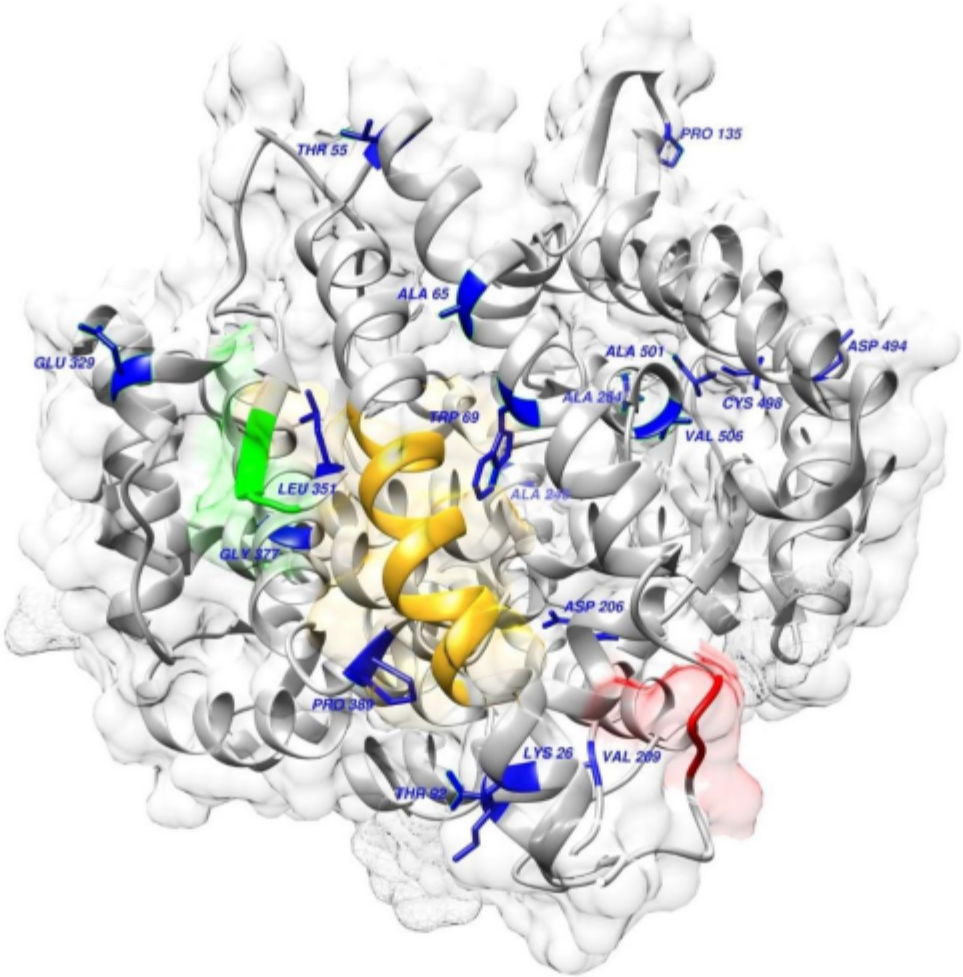
Panel A. Root Mean Square Deviation (RMSD) trends for the backbone of ACE2 WT (black line), V506A (green line) and W69C mutation (red line) during 100 ns of simulation. The molecular dynamics simulation showed a good stability for all systems. RMSD profile of ACE2 WT, V506A and W69C variant was used to explore their structural stability and mobility. Wild type and V506A show a steady course in the RMSD value, stabilizing at an average of 0.2 nm and 0,3 nm, respectively, while the W69C variant shows a gradual increase in RMSD value, stabilizing at an average of 0.5 nm. ACE2 W69C shows an initial stability change in the first 20ns probably due to a structural local rearrangement likely related to the transition from an hydrophobic amminoacid to a polar one. **Panel B** SASA graphical representation of ACE2 WT (Black line), W69C variant (red line) and ACE2 V506A variant (green line). ACE2 WT and ACE2 V506A variant showed a total average SASA higher than ACE2 W69C, along MD run. While, ACE2 V506A shows a similar SASA profile compared to ACE2 WT, a significant increment in average SASA has been observed in residues involved in 2019-nCoV S protein binding domain interaction for ACE2 WT compared to ACE2 W69C variant.

Table 1. Identified variants and predicted effects on protein stability

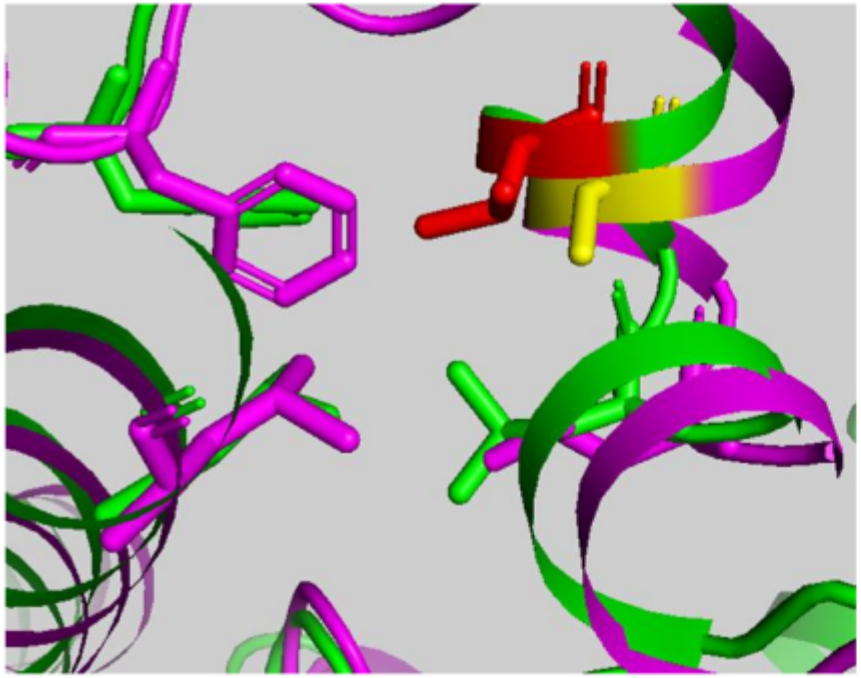
The table reports the identified variants, the corresponding allele frequency and the number of individuals hemyzigous, heterozygous or homozygous for the identified variant

Table 2. Predicted changes in ACE2 protein stability as consequence of residues changes

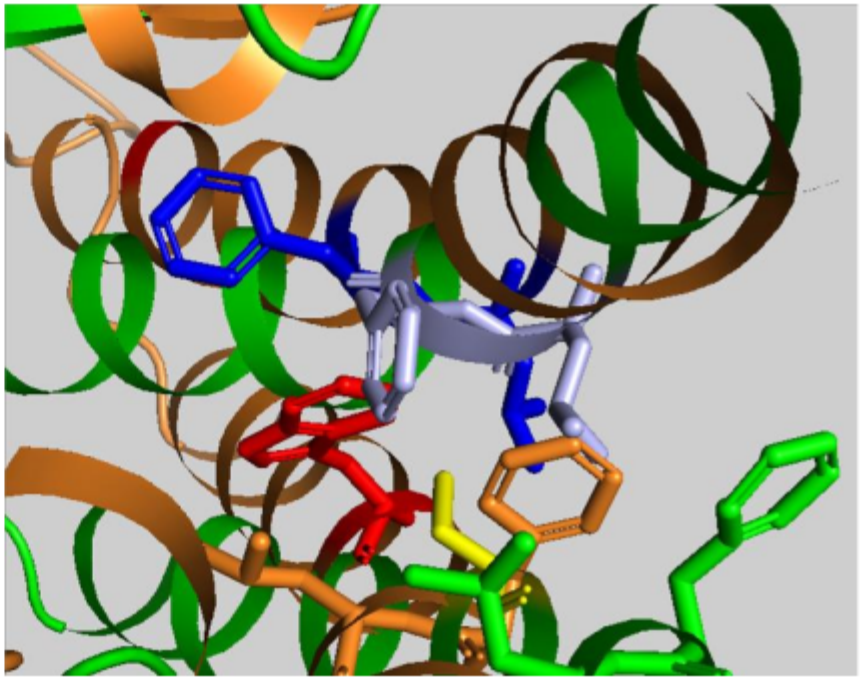
DUET program results that display predicted change in folding free energy upon ACE2 missense mutation (ΔΔG in kcal/mol). In this table are reported single missense mutations on ACE2 protein. ΔΔG analysis predict effects of missense mutations on protein stability using an integrated computational approach. The following column show the residues around 5Å, of each single missense mutations. In this column, we highlight in green residues involved in spike SARS-CoV protein interaction, in yellow residues involved in Zinc coordination and finally in magenta residues of Asn involved in N-glycosylation. The residues in the first column highlighted in grey are involved in N-glycosylation pattern NxT/S, therefore those missense mutations determine the loss glycosylation of Asparagine 53 and 90 respectively. The last column define the outcome of protein stability after each single missense mutations, much negative is ΔΔG, much higher is the destabilizing effect, while a positive sign corresponds to a mutation predicted as stabilizing.


THR 55
PRO 135
ALA 65
GLU 329
ALA 501
ASP 494
ALA 284
CYS 498
TRP 69
VAL 506
LEU 351
ALA 248
GLY 377
ASP 206
PRO 381
LYS 26
VAL 269
THR 92

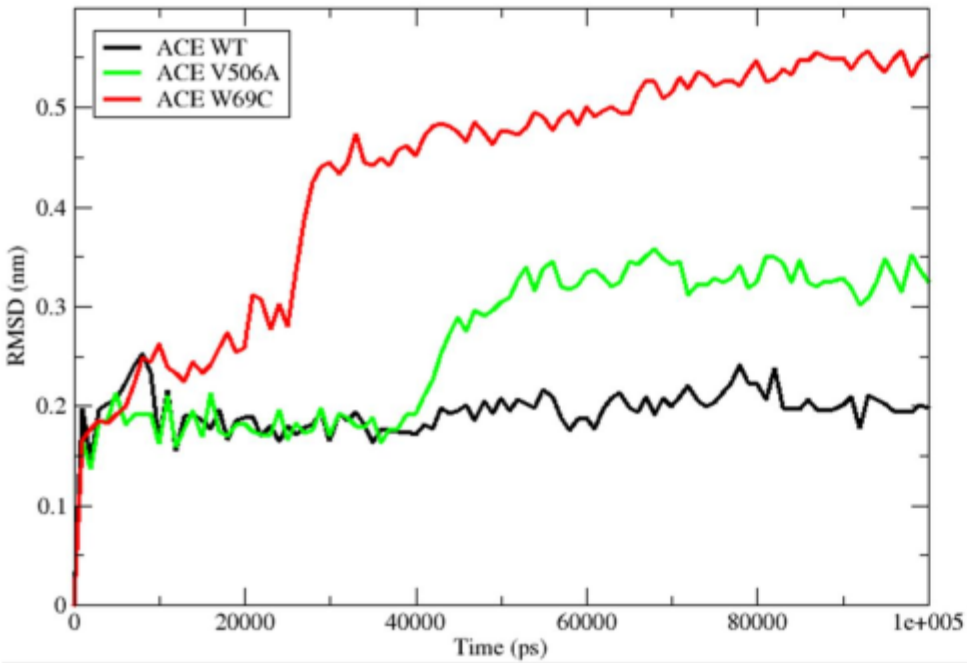
A



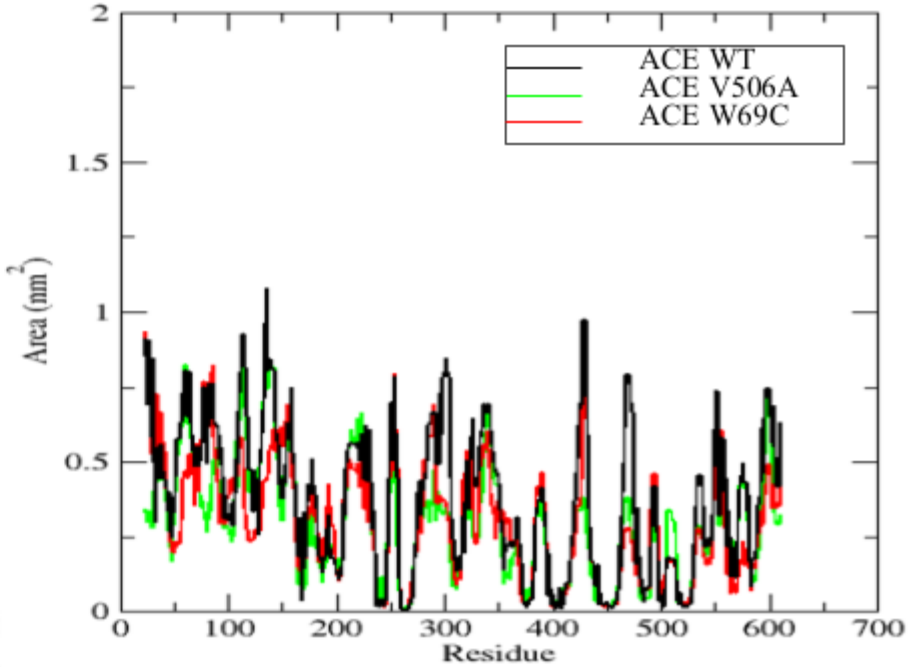
B



Root Mean Square Deviation (RMSD)



Solvent Accessible Surface Area (SASA)



Tab.1 Missens, frameshift and truncating variants										
AAAChange NM_021804.2 (hg37)										
c.	**p.**	**exon**	**CADD_phred**	**dbSNP**	**ExAC_ALL**	**Hemi M**	**Hetero F**	**Homo F**	**Num of Individuals**	**NIG allele frequency**
c.22C>T	p.Leu8Phe	exon2	14.2	rs201035388	0,000075506	1	1	-	2/6984	0,00019
c.77A>G	p.Lys26Arg	exon2	10.5	rs4646116	0,0039409	4	7	-	11/6984	0,00105
c.163A>G	p.Thr55Ala	exon2	23,8	rs775273812	0,0000056689	1	-	-	1/6984	0,00009
c.194C>T	p.Ala65Val	exon3	11.7	-	-	-	1	-	1/6984	0,00009
c.207G>T	p.Trp69Cys	exon3	23,7	-	-	1	0	-	1/6984	0,00009
c.275C>T	p.Thr92Ile	exon3	0,031	rs763395248	0,00001124	-	2	-	2/6984	0,00019
c.305A>C	p.Gln102Pro	exon3	8,311	rs1395878099	0,00001131	-	2	-	2/6984	0,00019
c.334A>G	p.Lys112Glu	exon3	18,09	-	-	-	1	-	1/6984	0,00009
c.404C>T	p.Pro135Leu	exon4	12.56	-	-	-	1	-	1/6984	0,00009
c.617A>G	p.Asp206Gly	exon6	20.8	rs142443432	0,0003	1	1	-	2/6984	0,00019
c.626T>G	p.Val209Gly	exon6	11,3	-	-	-	1	-	1/6984	0,00009
c.631G>A	p.Gly211Arg	exon6	14,7	rs148771870	0,00128	4	8	-	12/6984	0,00114
c.736G>A	p.Ala246Thr	exon7	0,28	-	-	2	-	-	2/6984	0,00019
c.791C>G	p.Ala264Gly	exon7	27,6	-	-	-	2	-	2/6984	0,00019
c.986A>G	p.Glu329Gly	exon9	12,4	rs143936283	0,00003443	-	5	-	5/6984	0,00047
c.1051C>G	p.Leu351Val	exon9	22,7	-	-	-	2	-	2/6984	0,00019
c.1129G>T	p.Gly377Trp	exon10	34	-	-	-	1	-	1/6984	0,00009
c.1130G>A	p.Gly377Glu	exon10	27,5	rs767462182	0,0000056	1	-	-	1/6984	0,00009
c.1166C>A	p.Pro389His	exon10	24,8	rs762890235	0,000039273	-	1	-	1/6984	0,00009
c.1189A>G	p.Asn397Asp	exon10	26,8	rs1365935088	0,00001464	-	1	-	1/6984	0,00009
c.1481A>T	p.Asp494Val	exon12	28,3	rs765152220	0,000050738	-	1	-	1/6984	0,00009
c.1492T>C	p.Cys498Arg	exon12	28,9	-	-	-	2	-	2/6984	0,00019
c.1501G>A	p.Ala501Thr	exon12	23,5	rs140473595	0,00002217	-	2	-	2/6984	0,00019
c.1514A>G	p.His505Arg	exon12	11,62	rs1016409802	0,00065	-	1	-	1/6984	0,00009
c.1517T>C	p.Val506Ala	exon12	27.2	rs775181355	0,000006561	1	1	-	2/6984	0,00019
c.1640C>G	p.Ser547Cys	exon13	24	rs373025684	0,00021	-	1	-	1/6984	0,00009
c.1738A>G	p.Asn580Asp	exon14	0,69	-	-	1	1	-	2/6984	0,00019
c.1918A>G	p.Met640Val	exon16	16.7	-	-	-	1	-	1/6984	0,00009
c.2012G>A	p.Arg671Gln	exon17	10.38	rs753705431	0,00001962	1	2	-	3/6984	0,00029
c.2158A>G	p.Asn720Asp	exon18	15,1	rs41303171	0,01622	43	60	-	103/6984	0,00983
c.2221A>G	p.Ile741Val	exon18	0,15	rs372923812	0,0001003	3	1	-	4/6984	0,00038
c.2345C>T	p.Ala782Val	exon19	11.3	rs147487891	0,000062374	1	-	-	1/6984	0,00009
c.2353G>T	p.Asp785Tyr	exon19	24,1	rs373153165	0,00003	-	1	-	1/6984	0,00009
c.143G>A	p.Trp48*	exon2	-	-	-	-	1	-	1/6894	0,00009
c.385del	p.Thr129Leufs*20	exon4	-	-	-	1	-	-	1/6984	0,00009
c.490del	p.Ala164Leufs*13	exon5	-	-	-	-	1	-	1/6984	0,00009
c.533del	p.Pro178Hisfs*9	exon5	-	-	-	-	1	-	1/6984	0,00009
c.563del	p.Asn188Metfs*2	exon5	-	-	-	-	1	-	1/6984	0,00009
c.670del	p.Glu224Lysfs*21	exon6	-	-	-	-	1	-	1/6984	0,00009
c.903del	p.Trp302Glyfs*22	exon9	-	-	-	-	1	-	1/6984	0,00009
c.1300del	p.Thr434Glnfs*3	exon11	-	-	-	-	1	-	1/6984	0,00009
c.1302dup	p.Glu435Argfs*31	exon11	-	-	-	-	1	-	1/6984	0,00009
c.1500del	p.Ala501Hisfs*58	exon12	-	-	-	-	1	-	1/6984	0,00009
c.1979del	p.Asn660Ilefs*3	exon16	-	-	-	1	-	1	2/6984	0,00029
c.2069del	p.Asn690Metfs*2	exon17	-	-	-	-	4	-	4/6984	0,00038
c.2306del	p.Lys769Argfs*55	exon18	-	-	-	-	1	-	1/6984	0,00009

Wild Residue	Residue position	Mutant Residue	Predicted ΔΔG	Interaction Network around (5 Å)	Outcome
V	**506**	**A**	-2.456	*Y180,L456,R460,P500,A501,S502,L503,F504,H505,N506,S507*	*Highly Destabilizing*
V	**209**	**G**	-2.353	*Y207,E208,V209,N210,G211,V212,Y215,D216,Y217,P565,T567*	*Highly Destabilizing*
G	**377**	**E**	-2.231	*H373,* **H374**,**E375**,M376,G377,**H378**,I379,A380,Y381,F315,H401,V404,G405,M408	*Highly Destabilizing*
A	**501**	**T**	-1.669	*C498,D499,P500,A501,S502,L503,V506,S507,G173,L176,R177,Y180*	*Destabilizing*
A	**264**	**G**	-1.555	*L262,P263,A264,H265,L266,L267,W271,W478,V487,V488,E489,P490,W165*	*Destabilizing*
C	**498**	**R**	-1.539	*Y497,C498,D499,P500,A501,S502,G173,R177,L176,Y180,W459,W473,M474*	*Destabilizing*
A	**246**	**T**	-1.454	*A242,Y243,V244,R245,A246,K247,L248,M249*	*Destabilizing*
G	**377**	**W**	-1.318	*H373* ,**H374**,**E375**,M376,G377,**H378**,I379,A380,Y381,F315,H401,V404,G405,M408	*Destabilizing*
W	**69**	**C**	-1.246	*F72,L73,* **F32**,**E35**,**A36**,**E37**,**L39**,**F40**,F390,A65,G66,K68,S70,S43	*Destabilizing*
L	**351**	**V**	-1.173	*W349,D350,L351,G352,***D355**,R357,Y41,S44,L45,W48	*Destabilizing*
P	**389**	**H**	-1.161	*A387,Q388,P389,F390,L391,L392,* **N33**,T92,Q96	*Destabilizing*
T	**55**	**A**	-0,948	**N53**,I54,T55,E56,E57,N58,V59	*Destabilizing*

D	**206**	**G**	-0,87	*W203,G205,D206,Y207,A396,N397,E398,G399*	*Destabilizing*
K	**26**	**R**	-0,79	*E22,E23,Q24,A25,K26,T27,L29,N90,V93*	*Destabilizing*
N	**580**	**D**	-0,629	*M579,N580,V581,R582,P583,Q524*	*Destabilizing*
S	**547**	**C**	-0,611	*I544,S545* ,N546,S547,T548,E549,A550,G551	*Destabilizing*
A	**65**	**V**	-0,423	*N61,M62,N63,N64,A65,G66,D67,K68,Q42,S43,S44,A46*	*Destabilizing*
P	**135**	**L**	-0,388	C133,N134,P135,D136,W163	*Destabilizing*
H	**505**	**R**	-0,345	*L503,F504,M505,F512,Y515,Y510,S511,R273*	*Destabilizing*
T	92	**V**	-0,322	N90,L91,T92,V93,L95,Q96,P389,L392,S563,E564	*Destabilizing*
E	**329**	**G**	-0,302	*Q325,G326,F327,W328,E329,N330,S331*	*Destabilizing*
G	**211**	**R**	-0,283	*V209,N210,G211,V212,D213,D216*	*Destabilizing*
T	92	**I**	-0,155	N90,L91,T92,V93,L95,Q96,P389,L392,S563,E564	*Destabilizing*
D	**494**	**V**	-0,041	*H493,D494,E495,T496,Y497*	*Destabilizing*
Q	**102**	**P**	0,036	Q98,A99,Q102,N103,G104	(Stabilizing)