

Structural insights into spike protein and its natural variants of SARS-CoV-2 found on Mexican population.

Yudibeth Sixto-López

Laboratorio de Diseño y Desarrollo de Nuevos Fármacos e Innovación Biotecnológica (Laboratory for the Design and Development of New Drugs and Biotechnological Innovation), Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, Plan de San Luis y Salvador Díaz Mirón s/n, Casco de Santo Tomás, Ciudad de México 11340, México

<https://orcid.org/0000-0001-8903-3834>

Martiniano Bello

Laboratorio de Diseño y Desarrollo de Nuevos Fármacos e Innovación Biotecnológica (Laboratory for the Design and Development of New Drugs and Biotechnological Innovation), Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, Plan de San Luis y Salvador Díaz Mirón s/n, Casco de Santo Tomás, Ciudad de México 11340, México

<https://orcid.org/0000-0002-9686-0755>

José Correa-Basurto

Laboratorio de Diseño y Desarrollo de Nuevos Fármacos e Innovación Biotecnológica (Laboratory for the Design and Development of New Drugs and Biotechnological Innovation), Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, Plan de San Luis y Salvador Díaz Mirón s/n, Casco de Santo Tomás, Ciudad de México 11340, México

<https://orcid.org/0000-0002-4973-5265>

Jose Antonio Garzón-Tiznado

2Laboratorio de Bioinformática y simulación molecular, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Culiacán Sinaloa México <https://orcid.org/0000-0003-0423-3349>

Sarita Montaña (✉ mmontano@uas.edu.mx)

Laboratorio de Bioinformática y simulación molecular, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Culiacán Sinaloa México <https://orcid.org/0000-0002-0420-4268>

Research Article

Keywords: Spike protein, SARS-CoV-2, molecular dynamics simulation, docking

Posted Date: August 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-57627/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a newly emerged coronavirus responsible for COVID-19; it becomes a pandemic since March 2020. To date, there are described three lineages of SARS-CoV-2 circulating worldwide, in Mexican population are found two of them, within this, we observed three variants of Spike (S) protein located at H49Y, D614G, and T573I. In order to understand if these mutations could affect the structural behavior of S protein of SARS-CoV-2, as well as the binding with three experimental describe inhibitors (Cepharanthine, Nelfinavir, and Hydroxychloroquine), molecular dynamic simulation and molecular docking were employed. It was found that in spite, these punctual mutations affect considerably the structural behavior of the S Protein, which also affect the binding of the inhibitors into their respective binding site. Thus, further experimental studies need to be done in order to explore if these affectations have an impact on drug-S protein binding and the possible clinical effect.

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a newly emerged coronavirus responsible for COVID-19; it becomes a pandemic since March 2020. To date, more than 18,902,735 million people have been infected and killed more than 709,511 as of August 6, 2020. The coronavirus entry into the host cell is mediated by the transmembrane spike (S) glycoprotein. Homotrimers of S proteins are surface-exposed and are responsible for the attachment to the host receptor, which turns it into the main target of neutralizing antibodies^{1,2}. The fusion capacity of the S protein is a leading indicator of the infectivity of the virus. S protein consists of two functional subunits, receptor-binding (S_1) and fusion the viral and cellular membrane (S_2)^{2,3}. S_1 comprises the receptor-binding domain (RBD) that recognizes angiotensin-converting enzyme-2 (ACE2) as its receptor^{4,5}, S_1 contributes to the stabilization of the prefusion state of the membrane-anchored S_2 subunit, that contains the fusion machinery. The S1/S2 processing sites exhibit different motifs among coronaviruses⁶. The priming process is ensured by different host cell proteases depending on the S1/S2 sequence cleavage site. In the case of the SARS-CoV-2, S protein contains a canonical furin-like cleavage site: 681-PRRAR↓SV-688, which is absent in SARS-CoV and others SARS-related coronaviruses^{2,3,6,7}, and it may provide a gain of function for efficient spreading in the human population compared to other lineages of beta- coronaviruses⁶.

Recent studies show SARS-CoV-2 is mutating during the continuous transmissions among the population^{8,9}. To date, the Mexican population is circulating two of three viral lineages reported¹⁰. The principal mutations detected worldwide in the SARS-CoV-2 are on the viral S protein^{9,11-14}, which is critical for the virus attachment to the host cell¹¹. However, the mutation effects on the S protein are poorly understood and possible impact at 3D structural or functional level. These variations can be used to identifies epitopes as a target for vaccine development, as well as antibody and antiviral drugs design, to find treatments against SARS-COV-2^{9,14}.

Insights into the sequences mutations of S protein among available genomes on the GISAID database, we found three principal variations, H49Y, D614G, and T573I in the Mexican population. The 3D crystal structures of S protein are available at PDB database^{2,15-17}. However, to capture conformational changes in the 3D atomistic model, Molecular Dynamic simulation (MD) needs to be performed¹⁵. To determine if the variants H49Y, D614G, and T573I on S protein founded on the Mexican population, could affect the 3D structure, here we employed molecular dynamics simulations (MD) for a full-length atomistic 3D structure of S protein and its variants. Molecular docking calculations with the most populated cluster conformations of the variants were done with Cepharanthine, Hydroxychloroquine, and Nelfinavir, reported as S protein inhibitors^{27,28,29}, in order to explore how the binding mode or free energy of binding was affected due to possible structural modifications because of punctual mutations on S protein. The results provide evidence of affectations of both binding free energy and binding mode of the compounds with the S protein WT and its variants (H49Y, D614G, and T573I) supported by *in silico* studies that can be further investigated with experiments.

Results And Discussion

Multiple sequence alignments of the SARS-CoV-2 S proteins show three variants present on Mexican population (H49Y, T573I, and D614G), the rest of the residues are conserved (Fig. 1). These variants above mentioned are far from the RBD domain of S protein.

D614G change is caused by an A/G nucleotide mutation at position 23,403 in the reference strain³², this variant is being associated with higher viral loads and enhances viral effectivity in patients^{8,32,33} despite the higher viral loads, G614 is neutralizing similarly as D614 by polyclonal antibody³². To date, this variant becomes the dominant form displacing the WT, according to the levels of mutation in the world, presented on the nextstrain database (www.nextstrain.org). H49Y variant is produced by C/T change at position 21,707¹⁰. The properties of residues H/Y changes from positive to neutral charge⁹, this variation causes a reduction in the total free energy, while the mutants with D614G substitutions showed stabilizing structure, suggesting a prevalence role in S protein evolution¹³. Finally, regarding T573I mutants, until now, there is no information available, but the change by T/I implies a modification from polar residue to non-polar hydrophobic residue changing the chemical environment to more hydrophobic site. Even though these are punctual modifications due to the chemical nature of the substitution, modification at the structural level is expected.

Molecular dynamics simulations

The mutations on the 3D structures of S were done using PDB: 6VSB crystal on PyMol (Fig. 2), and proceed to develop the Molecular dynamic simulation with Amber. RMSD was calculated to determine the average deviation in the atomic stability under MD simulations. All RMSD trajectories reached the equilibrium after 30 ns of MD simulation (Fig. 3A). The Rg values for WT showed expansion at the last 30

ns of MD with values from 47.5 to 50 Å. D614G and T573I show compactness since 20 ns of trajectory with values from 49.5 to 44.8 Å and 50.4 to 43.79 Å, respectively. H49Y presents compactness from 48.47 to 43.9 Å in the whole simulation (Fig. 3B). Rg values, suggests that variants have a similar degree of compactness while the WT shows an opposite behavior. We explored protein flexibility by RMSF values of the Ca through MD simulations of the S proteins and its variants (Fig. 3C). Seven principal peaks of fluctuation were located on WT showed up as the most flexible areas situated at 223-288, 688-729, 730-776, 823- 869,

946-1024, and at the C- terminus from 1060-1143; for D614G: 57-77, 395-594, 722- 780, 823-870, 948-1032, and the C- terminus from 1064-1143; H49Y presented a different pattern of flexibility showing punctual fluctuations, where the most flexible residues were 67, 112, 47, 182, 255, 444, 476, 557, 632, 701, 838, 845 and the C-terminus from 1135-1143 T573I showed three flexible regions located from 55-94, 401-525, 947- 1026 plus the C-terminus from 1138-1143.

Clustering analysis

Representative ensembles were calculated from the last 30 ns of the MD simulation. 70% of the most populated conformations were grouped into the first 10, 13, 15, and 13 clusters for WT, D614G, H49Y, and T573I, respectively (Table 1), this clustering dispersion indicates that proteins possess a complex structural behavior. Although, it was observed that WT protein depicted lesser clustering dispersion than mutated protein, which pointed out that mutated spike is more complexed in their structural behavior. Only the most populated cluster conformation was retrieved from each one of the systems, and it is represented in Figure 4. The most populated cluster conformation of the WT showed more difference regarding the native structure (RMSD = 9.625 Å, Fig. 4A), D614G (RMSD = 9.998 Å, Fig.4B) showed a higher RMSD than WT while, the other mutated proteins depicted lesser difference in decreasing order as follow, H49Y (RMSD = 6.965 Å, Fig. 4C) >T573I (RMSD = 6.858 Å, Fig. 4D). Specifically, WT structural difference was observed in most of the structure, and little differences were observed at 270-318 that belongs to N- terminal domain (NTD) and 531-700, that belongs to S1; however, D614G mutant depicted difference in this region which indicate that a single residue mutation affected the stability of this “central portion of the WT protein. In both cases, WT and D614G difference into secondary structures were observed as well as displacements in loops, α -helices, and β - sheets structures. While in H49Y variant, not greater structural changes were observed in comparison to WT, thus it seems that H49Y induces lesser structural changes that are in agreement with RMSD and RMSF observations, where punctual fluctuations were observed. These observations are in agreement with *in silico* predictions of energetic stability induce by mainly by H49Y mutations and in lesser grade by D614G¹³. In T573I variant, fluctuation from 535-580 was observed reduced, therefore T573I mutation seems to induce structural stability into WT protein as well as in the zone around where the mutation is found.

Table 1. Cluster analysis using RMSD cut-off of 2.0 Å.

Cluster	Cluster population (%)			
	WT	D614G	H49Y	T531I
1	10.6	10.4	6.8	8.2
2	9.6	7	6	7.4
3	9	6.6	6	7.4
4	7	5.4	5.4	7.2
5	6.6	5.2	5.4	5.8
6	6.6	5.2	5.4	5.4
7	6.4	4.6	4.6	5.4
8	6.4	4.6	4.4	5
9	6	4.6	4.2	4.6
10	6	4.2	4.2	4.6
11	5.2	4.2	4.2	4.4
12	4.8	4.2	4	4.4
13	4.2	4	3.6	4
14	3.8	4	3.6	4
15	3.4	3.8	3.6	3.6
16	2.6	3.4	3.4	3
17	1.8	3.4	3.2	3
18		3.4	3.2	3
19		3.2	2.8	3
20		3.2	2.6	2.8
21		3	2.6	2
22		2.4	2.4	1.8
23			2.4	
24			2.4	
25			1.8	
26			1.8	

Principal component analysis

Exploration of the principal components that contribute to the global motions of WT protein and mutant systems was performed (Fig. 5A) on the MD simulations. According to PCA analysis, the first 15 eigenvectors captured 92-96% of the total motions (94.8, 96.2, 92.8, and 96.1 % for WT, D614G, H49Y, and T563I, respectively) (Fig. 5B). Whereas, the projections of the first two eigenvectors (PC1 vs PC2) contributed to 48-71% of the collective motions (67.7, 74.4, 48.1, and 71.3 % for WT, D614G, H49Y, and T563I, respectively) (Fig. 5 C-F). By projections onto the essential subspace of PC1 vs. PC2, it was observed that the WT system showed different mobility behavior in comparison to the variant systems. Similar cluster distribution was observed for WT, D614G, and T563I variants, having slightly higher conformational mobility for D614G and T563I than WT (Fig. 5 C, D, and F). In comparison, H49Y showed a more compact cluster distribution than the other systems, which suggest a reduction in conformational mobility due to single residue mutation (Fig 5D). T563I mutant showed dissimilarly conformed distribution along the subspace in comparison to the others systems, this behavior suggest that the trajectories sampled different regions of the phase space with different minima but small energy barrier,

in this sense T563I mutant affects the structural behavior of the protein with is also reflected into the conformations observed during MD simulation (Fig 5F). These results can be corroborated from the analysis of the graphical representation of the full mobility along PC1 and PC2 (Figure 6), which allows us to study the direction and magnitude of the motions that contribute to the total system's mobility. According to this projection, the movement on WT was uniformed, and by regions, they showed the same direction, i.e., RBD region and S1 and S2 (Fig. 6 A). However, for D614G the direction of the movements change mainly on S2 region and in NTD segment (Fig. 6 B), which is in agreement with the cluster distribution of PC1 vs PC2 projection (Fig. 5D), this modification not only increases the flexibility of the hinge but also it changes the direction which might contribute to the increase virulency of this mutant because of might alter the binding to ACE protein¹³. Previously, it was found that D614 formed a hydrogen bond with T859 and a salt bridge with K854 placed on the subunit two of the other protomer, thus the change of D by G might provide a flexible space between the monomers and more conformational flexibility as our MD simulations suggest which might lead to improved accessibility to ACE2, which are consistent with more cell entry of the virus that carries this mutation^{8,34}. This motion perturbation is explained because D614 residue is located on a loop exposed side chain, has a negative charge, and that will be a loss on the variant G614, this causes loss of interactions with other molecules or residues. The G614 has a neutral charge, is smaller, and introduces a more hydrophobic residue at this position, and this can result in loss of hydrogen bonds and/or disturb correct folding. Glycines are very flexible and can disturb the required rigidity of the protein at this position¹³. In the case of H49Y mutant, the whole mobility decreased, which was reflected in the magnitude of the porcupine representation in Figure 6C, besides the direction of the movement was completely altered in comparison to WT. It was significantly decreased on S2 region (816-1106), suggesting that this mutant not only confers more structural stability as we found but also energetic stability as it was suggested by other authors¹³. Aside, it was found that the H49Y mutant virus has increased cell entry in comparison to WT S protein⁸.

Finally, on the T563I mutant at S1 region, an increment on the magnitude of the movements was observed, while on S2 region, the direction of the movements was on the contrary direction to that observed on WT (Fig 6D). So, this mutation altered both magnitudes a direction of the mobility of the protein. In this case, until now, there is not much more evidence about the behavior of the virus that possesses this mutation, so it should be explored to correlate with our findings.

Molecular docking with known ligands

Finally, to explore how these punctual mutations affect the protein-ligand binding, a molecular docking with Cepharenthine, Hydroxychloroquine, and Nelfinavir which are experimental reported spike ligands were performed²⁹⁻³¹

Cepharanthine showed potent antiviral activity *In vitro* against SARS-CoV-2, and by molecular docking, spike protein was suggested as its target. It is suggested that cepharanthine is bound to RBD, and might

interfere with human ACE2 binding avoiding the anchorage of the virus to the cell^{30,35}. In wild type complex, cepharanthine formed a hydrogen bond with E484 and S494 and hydrophobic interactions with L455, F456, Y489, F490, L492, and Q493 which are residues involved in the Interactions between viral spike and ACE2 (7A & D)³⁶. Cepharanthine was bound to WT spike protein with the highest binding free energy (-6.57 kcal/mol), followed by H49Y (-6.42 kcal/mol) which also interacted with residues belonging to RBD by only hydrophobic interactions with Y449, N450, Y451, L452, F490, L492, S494 and Q493 (Fig. 7A and F). Cepharanthine was a little bit displaced regarding wild type-cepharanthine complex, which might be responsible for the binding energy detriment. While D614G and T563I showed lesser binding free energy (-5.95 and -5.39, respectively). Cepharanthine also interacted with residues of the RBD domain of D614G mutant, by hydrophobic interactions with Y449, N450, Y451, L452, E484, F490, L492, Q493, and S496 (Fig. 7E). While within the T563I mutant, Cepharanthine was moved away from the typical site where the other complexes interacted, and formed hydrogen bond with R509 and hydrophobic interactions with F342, N343, A344, W436, N437, S438, N439, and N440 none of these residues belong to ACE2-RBD interacting residues (Fig. 7G). It seems that mutation on T563I might affect the binding and maybe the biological effect of cepharanthine as anti-SARS-CoV-2 agent³⁷.

Hydroxychloroquine is also a compound with effective *in vitro* inhibition activity against SARS-CoV-2 infection, besides it elevates the endosomal pH and affects the ACE-2 terminal glycosylation avoiding the virus binding³⁸. Two independent computational work suggested that hydroxychloroquine might interact with the RBD domain of spike protein, through it can disrupt the viral-host recognition to avoid viral infection^{29,30}. Therefore, in this study, we selected hydroxychloroquine to performed docking with WT spike protein and with mutants. Similar behavior with cepharanthine was observed, hydroxychloroquine reaches the binding site on RBD WT spike protein as well as D614G and H49Y, while in T563I mutant hydroxychloroquine was moved away from this site, and it was reflected on the detriment of the binding free energy (Table 2, Fig. 7B). Hydroxychloroquine in wild type formed a hydrogen bond with Y489 and L492, and hydrophobic interactions with Y449, L452, F456, E484, G485, C488, F490, E493 and S494 (Fig. 7H). In the interaction with D614G, hydroxychloroquine formed hydrogens bond with E484, F490, L492, and Q493 and hydrophobic interactions with S494, Y449, L452, L455, and P491 (Fig. 7I). And with H49Y formed hydrogen bonds with S349, Y449, N450, and S494 and hydrophobic interactions with R346, F347, A348, Y351, A352, Y451 and L452 (Fig. 7J). in wild type spike protein and both D614 and H49Y mutant Hydroxychloroquine interacted with residues of the RBD that participate in ACE-RBD recognition^{36,37}. At the same time, Hydroxychloroquine in T563I mutant interacted with F342, W436, and R509, and hydrophobic interactions with S438, L441, N343, A344, S373, F374, N437, S438, which are residues that are not involved in ACE2-RBD interaction³⁶ (Fig. 7K).

Nelfinavir is currently marketed as an anti-HIV drug, and recently it was demonstrated that inhibit the cell fusion mediated by spike protein of SARS-CoV-2, by molecular docking, it was identified that nelfinavir is placed between fusion peptide and HR1 helix near the S1/S2 cleavage site³¹. Therefore, nelfinavir was assayed by molecular docking against WT and mutants forms of the spike protein. Nelfinavir was bound to wild type spike protein with the highest binding free energy (-7.88 kcal/mol) followed by H49Y (-7.52

kcal/mol), D614G (-6.10 kcal/mol) and finally, T563I (-5.38 kcal/mol). Regarding the binding site, all ligands were allocated around HR1, FP region, and S1/S2 cleavage site, but Nelfinavir reaches different binding sites in every protein studied (Fig. 7C). Nelfinavir in WT spike protein was placed in the HR1 region and formed hydrogens bond with K310, I312, and Q314, and hydrophobic interactions with L303, E309, G311, Y313, I664, P665 and I666 as previously reported (Fig. 7L)³¹. While Nelfinavir with D614 mutant was also placed on the HR1 region but interacted with some different residues than in the wild type, formed a hydrogen bond with N317, S596, K947 and Q1010 and hydrophobic interactions with Q314, F592, G593, I1013 (Fig. 7M). Whereas Nelfinavir with H49Y only interacts with a couple of residues of HR1 region by hydrophobic interactions (A942 and S943), and with residues of the N terminal domain (S13-L303) by hydrogens bond T302, L303, K304 or hydrophobic interactions V47, L48, Y49, S305, in this case, the displacement outer from the HR domain was more evident than with D614G. Still, these interactions were energetically favorable (Table 2, Fig. 7N). In the case of Nelfinavir-T563I mutant complex, Nelfinavir interacted mainly with residues of the HR1 region by hydrogen bonds S943, D950 and D954, and hydrophobic interactions with K947, V951, R1014, it also formed hydrogen bond with K310 and hydrophobic interaction with P665 (Fig. 7O), eventhough this binding mode was energetically lesser favorable than those observed with mutants and WT protein.

Therefore, the ligand-protein interaction depicted a similar energy behavior than the observed with cepharanthine and hydroxychloroquine, where WT protein-ligand complex has the highest energy, while T563I-ligand complex has the lowest energy of binding, which could indicate that mutation on T563I position could be relevant for drug-spike protein interaction affecting not only the binding mode but also the binding free energy.

Table 2. binding free energy of compounds docked with Wild type spike protein and mutants.

	Binding free energy (Kcal/mol)
Cepharanthine	
D614G	-5.95
H49Y	-6.42
T563I	-5.39
WT	-6.57
Hydroxychloroquine	
D614G	-4.55
H49Y	-4.66
T563I	-4.19
WT	-4.46
Nelfinavir	
D614G	-6.1
H49Y	-7.52
T563I	-5.38
WT	-7.88

Materials And Methods

Multiple alignments of Spike (S) protein. The sequences corresponding to S protein from SARS-CoV-2 were downloaded from the Global Initiative for Sharing All Influenza Data (GISAID) database¹⁶ (www.gisaid.org). The sequence from Wuhan was used as wild type (WT), and it was aligned against twelve S sequences from the Mexican population (Table 3). Multiple alignments were carried out in ClustalX using default parameters, further, it was edited on Jalview¹⁷. The whole sequence of the spike S protein has 1282 residues, to show only the residues where the protein exhibited the variants on the Mexican population, the sequence without changes were hidden. To obtain the 3D variants of S protein, mutations were done on PyMol¹⁸, using as template the crystal structure of SARS-CoV-2 spike glycoprotein (PDB: 6VSB). Visualization of wild type and the H49Y, T573I, and D614G S variants were performed on VMD (Fig. 2). Zoom of the structural 3D alignments were done to visualize the region which contains the mutation (Fig. 3).

Table 3. Sequences of SARS-COV-2 and procedences

SECQUENCE	Type	Procedence	ID
1	hCoV-19	IPBCAMS-WH-01	EPI_ISL_402123
2	hCoV-19	CDMX-INCMNSZ_04	EPI_ISL_426364
3	hCoV-19	CDMX-INCMNSZ_03	EPI_ISL_426363
4	hCoV-19	CDMX-INCMNSZ_01	EPI_ISL_426361
5	hCoV-19	CDMX-INCMNSZ_02	EPI_ISL_426362
6	hCoV-19	CDMX-InDRE_06	EPI_ISL_424673
7	hCoV-19	Puebla-InDRE_05	EPI_ISL_424672
8	hCoV-19	Queretaro-InDRE_04	EPI_ISL_424670
9	hCoV-19	EdoMex-InDRE_03	EPI_ISL_424667
10	hCoV-19	CDMX-INER_05	EPI_ISL_424627
11	hCoV-19	CDMX-INER_04	EPI_ISL_424626
12	hCoV-19	CDMX-INER_01	EPI_ISL_424345
13	hCoV-19	CDMX-InDRE_01	EPI_ISL_412972

MD Simulation and analysis.

MD simulations were carried out with AMBER 16 software package¹⁹ using ff14SB forcefield²⁰. The systems were solvated using an explicit TIP3P water model and centered into a rectangular box of 12.0 Å; after that, all the systems were neutralized by adding 6 Na⁺ counterions. Each one of the systems was minimized through 2500 steps of steepest descent, and 2500 steps of conjugate gradients. Then, they were equilibrated through 500 picoseconds (ps) of heating and 500 ps of density equilibration with weak restraints followed by 2 nanoseconds (ns) of constant pressure equilibration at 310 K. 50 ns of MD simulations were performed under periodic boundary conditions and using an NPT ensemble at 310 K.

The electrostatic term was described through the particle mesh Ewald method²¹; using a 10.0 Å cut-off, similar radio was chosen for Van der Waals interactions. The SHAKE algorithm²² was used to constrain bond lengths at their equilibrium values, and a time step was set to 2.0 fs. Temperature and pressure were maintained with the weak coupling algorithm²³ using coupling constants τ_T and τ_p of 1.0 and 0.2 ps, respectively (310 K, 1 atm). Trajectories were analyzed using cpptraj module for root mean squared deviation (RMSD), root mean squared fluctuations (RMSF), the radius of gyration (R_g). Based on the knowledge of equilibration time, clustering analysis (using a 3.0 Å cut-off), was done using a hierarchical agglomerative (bottom-up) approach employing AmberTools 16 Principal component analysis (PCA), also known as essential dynamic (ED) was performed. PCA analysis is a statistical technique that allows to obtain the large scale collective motions of the atoms on the simulations, which are frequently correlated to the biological function and biophysical properties²⁴. The performed method is described in detail elsewhere²⁵. Structural analysis of the systems and figures were performed using PyMOL v0.99¹⁸.

Molecular graphics were performed in SigmaPlot 12.0, and protein visualization was performed by VMD and Chimera^{26,27}.

Molecular docking

Autodock 4.2 software was used to docking calculations²⁸. Focused docking was performed, for Cepharanthine and hydroxychloroquine grid box was center on RBD domain with a grid box size of 80 Å x 70 Å x 70 Å^{29,30}. For Nelphinavir grid box was focused on residues L303, Y313, Q314, between HR1 and FP region. With a grid box size of 80 Å x 82 Å x 80 Å³¹ and a grid spacing of 0.375Å. Compounds and S protein were prepared using AutoDock Tools 1.5.6²⁸. Polar hydrogen atoms and Kollman charges were encompassed for the S protein. Lamarckian genetic algorithm was used as a scoring sample with a randomized population of 100 individuals, and energy evaluations of 1×10^7 , 100 runs were performed. As a selection criterion, the conformation with the lowest free energy values was chosen. Docking results were analyzed using Autodock Tools 1.5.6²⁸, and whereas figures were further processed with Pymol v.099¹⁸.

Declarations

Disclosure statement: The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Funding: This work was supported by project PROFAPI 2015/182 (SM)

Author contributions

Y.S.L Carried out theoretical studies, results analysis and wrote the manuscript; J.C.B, MB and

J.A.G.T analyzed the data and wrote the manuscript and SM, carried out theoretical studies, project design, results analysis and wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Beniac, D. R., Andonov, A., Grudeski, E. & Booth, T. F. Architecture of the SARS coronavirus prefusion spike. *Struct. Mol. Biol.* **13**, 751–752 (2006).
2. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).
3. Xia, S. *et al.* The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal Transduct. Target. Ther.* **5**, (2020).
4. Li, F. Receptor Recognition Mechanisms of Coronaviruses: a Decade of Structural Studies. *Virology* **89**, 1954–1964 (2015).
5. Wenhui Li, Michael J. Moore, Natalya Vasilieva, Jianhua Sui, Swee Kee Wong, Michael A. Berne, Mohan Somasundaran, John Sullivan, Katherine Luzuriaga, Thomas C. Greenough, H. C. & M. F. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Lett. to Nat.* **426**, 450–454 (2003).
6. Coutard, B. *et al.* The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* **176**, (2020).
7. Shang, J. *et al.* Cell entry mechanisms of SARS-CoV-2. *Natl. Acad. Sci. U. S. A.* **117**, (2020).
8. Ozono, S. *et al.* Naturally mutated spike proteins of SARS-CoV-2 variants show differential levels of cell entry. *bioRxiv* 06.15.151779 (2020). doi:10.1101/2020.06.15.151779
9. Mohammad, S. *et al.* Infection , Genetics and Evolution Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein : A computational biology approach. *Genet. Evol.* **84**, 104389 (2020).
10. Taboada, B. *et al.* Genomic Analysis of Early SARS-CoV-2 Variants Introduced in Mexico. *Virology* (2020). doi:10.1128/JVI.01056-20
11. Becerra-Flores, M. & Cardozo, T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *J. Clin. Pract.* 4–7 (2020). doi:10.1111/ijcp.13525
12. Zhang, L. *et al.* The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* 06.12.148726 (2020). doi:10.1101/2020.06.12.148726
13. Laha, S. *et al.* Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *bioRxiv* **85**, 2020.05.03.066266 (2020).
14. Ing, , Bello, I., Areiza, M. & Oliver, J. No Title No Title. *J. Chem. Inf. Model.* **53**, 45–50 (2013).

15. Sikora, M. *et al.* Map of SARS-CoV-2 spike epitopes not shielded by *bioRxiv* 2020.07.03.186825 (2020). doi:10.1101/2020.07.03.186825
16. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, 2–4 (2017).
17. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
18. Schrödinger, LLC. *The {PyMOL} Molecular Graphics System, Version~1.8.* (2015).
19. Case D.A. , Brozell S.R., Cerutti D.S. , Cheatham T.E. , Cruzeiro V.W.D. , Darden T.A. , Duke R.E. , Ghoreishi D. , Gohlke H. , Goetz A.W. , Greene D. , Harris R , Homeyer N. , Izadi S. , Kovalenko A. , Lee T.S. , LeGrand S. , Li P. , Lin C. , Liu J. , Luch, P. A. K. AMBER. (2018).
20. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Comput. Chem.* **24**, 1999–2012 (2003).
21. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *Chem. Phys.* **98**, 10089–10092 (1993).
22. van Gunsteren, W. F. & Berendsen, H. J. C. Algorithms for macromolecular dynamics and constraint dynamics. *Phys.* **34**, 1311–1327 (1977).
23. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684– 3690 (1984).
24. Amadei, A., Limddrn, A. B. M. & Berendsen, H. J. C. Essential dynamics of peotenis Priteins. *Proteins* **17**, 412–425 (1993).
25. Sixto-López, Y. *et al.* Searching the conformational complexity and binding properties of HDAC6 through docking and molecular dynamic simulations. *Biomol. Struct. Dyn.* **35**, 2794–2814 (2017).
26. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Mol. Graph.* **14**, 33–38 (1996).
27. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *Comput. Chem.* **25**, 1605–1612 (2004).
28. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Comput. Chem.* **30**, 2785–2791 (2009).
29. Nimgampalle, M., Devanathan, V. & Saxena, A. Screening of Chloroquine, Hydroxychloroquine and its derivatives for their binding affinity to multiple SARS- CoV-2 protein drug targets. *Biomol. Struct. Dyn.* 1–13 (2020). doi:10.1080/07391102.2020.1782265
30. M, S. & JC, S. Repurposing Therapeutics for COVID-19: Supercomputer- Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral SpikeProtein- Human ACE2 Interface. *chemRxiv* (2020).

31. Musarrat, F. *et al.* The anti-HIV drug nelfinavir mesylate (Viracept) is a potent inhibitor of cell fusion caused by the SARSCoV-2 spike (S) glycoprotein warranting further evaluation as an antiviral against COVID-19 infections. *Med. Virol.* jmv.25985 (2020). doi:10.1002/jmv.25985
32. Korber, B. *et al.* Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* (2020). doi:10.1016/j.cell.2020.06.043
33. Eaaswarkhanth, M., Madhoun, A. Al & Al-mulla, F. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information. (2020).
34. Korber, B. *et al.* Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv* 04.29.069054 (2020). doi:10.1101/2020.04.29.069054
35. Hirofumi Ohashi, Koichi Watashi, Wakana Saso, Kaho Shionoya, Shoya Iwanami, Takatsugu Hirokawa, Tsuyoshi Shirai, Shigehiko Kanaya, Yusuke Ito, Kwang Su Kim, Kazane Nishioka, Shuji Ando, Keisuke Ejima, Yoshiki Koizumi, Tomohiro Tanaka, Shin Aoki, Kouji Kur, T. W. Multidrug treatment with nelfinavir and cepharanthine against COVID-19. *chemRxiv*
36. Ortega, J. T., Serrano, M. L., Pujol, F. H. & Rangel, H. R. Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis. *EXCLI J.* **19**, 410–417 (2020).
37. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
38. Liu, J. *et al.* Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. *Cell Discov.* **6**, 16 (2020).

Figures

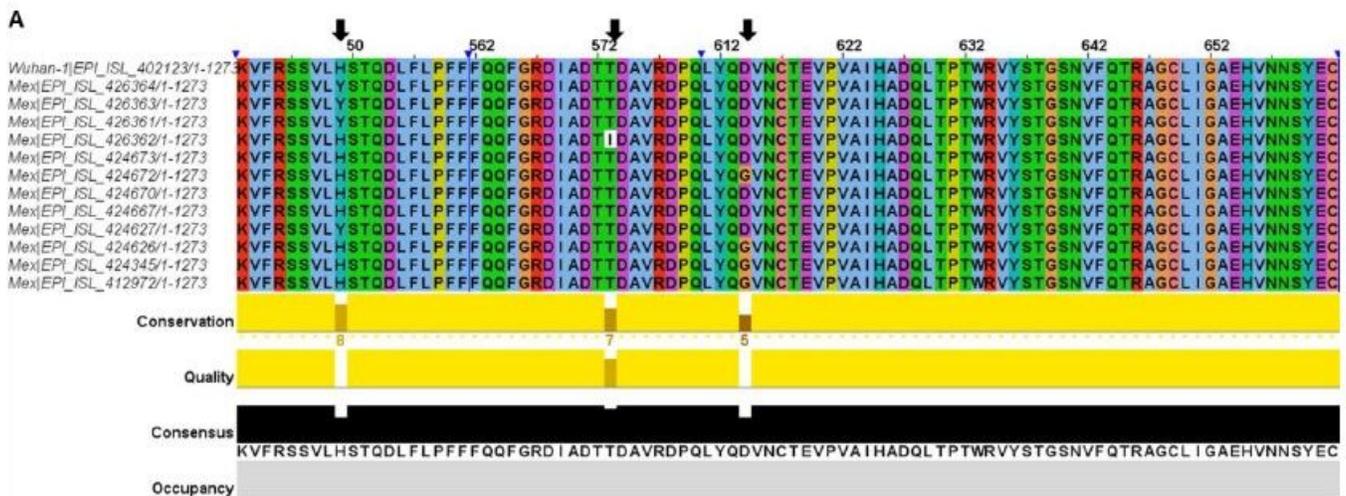


Figure 1

Multiple sequence alignment of Spike SARS-CoV-2 protein. The Wuhan Spike sequence was taken as wild type, the rest of the sequences correspond to Mexican Population. Arrows (black) indicate the row where the mutations were observed. The head of the arrows (blue) shows the hidden sequences.

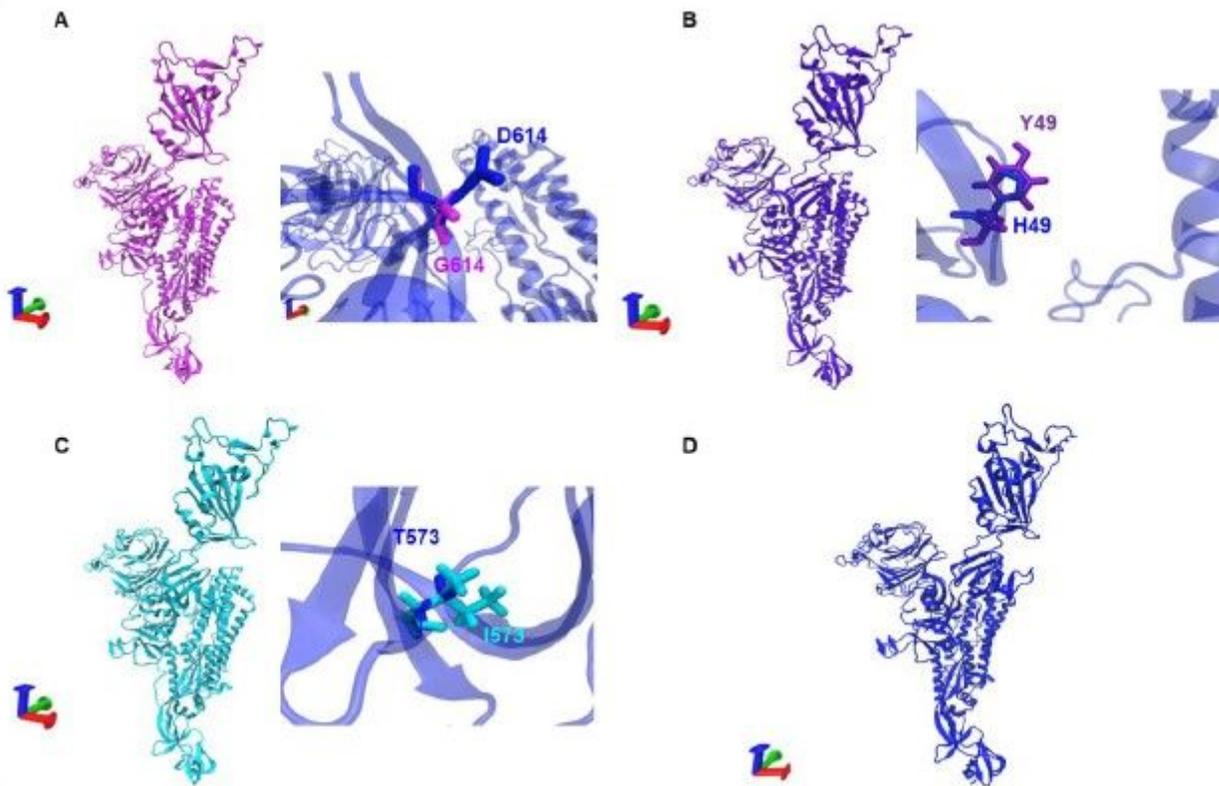


Figure 2

3D structure of Spike protein of SARS-COV2 and variants. A) 3D Structural conformation of D614G (pink) next, the zoom of the amino acid G614 is shown in pink while the WT residue is indicated in blue. B) 3D structure of H49Y (purple), the zoom of the amino acid variant Y49 is shown in purple, the WT residue is shown in the sticks in blue. C) 3D structure of T573I (cyan) the zoom of the residue variant I573 is shown cyan while the WT residue T573 is indicated in blue sticks. D) 3D structure of wild type S protein is shown in blue. Axes: X, red; Y, green; Z, blue.

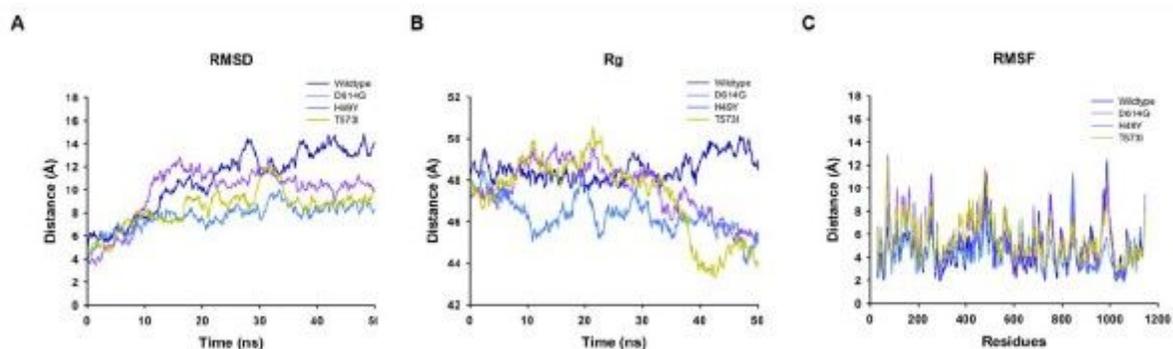


Figure 3

The trajectory of MD Simulations. A)RMSD B) Radius of gyration C) RMSF. The trajectories of the WT are shown in navy blue, D614G are shown in purple, H49Y are shown in blue and T573I are shown in yellow.

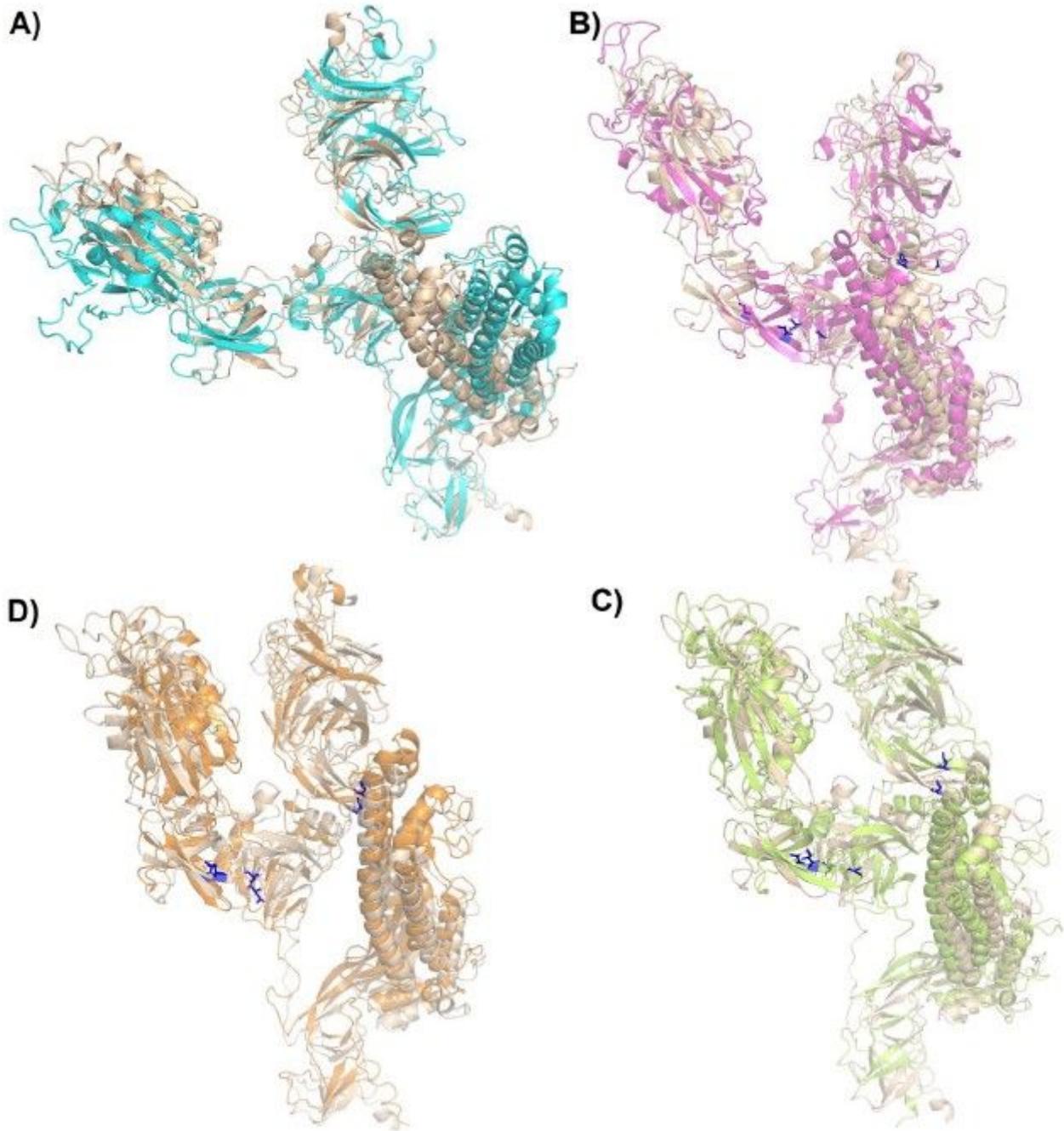


Figure 4

Structural changes between the starting conformation and the most populated cluster on the trajectory of MDS.

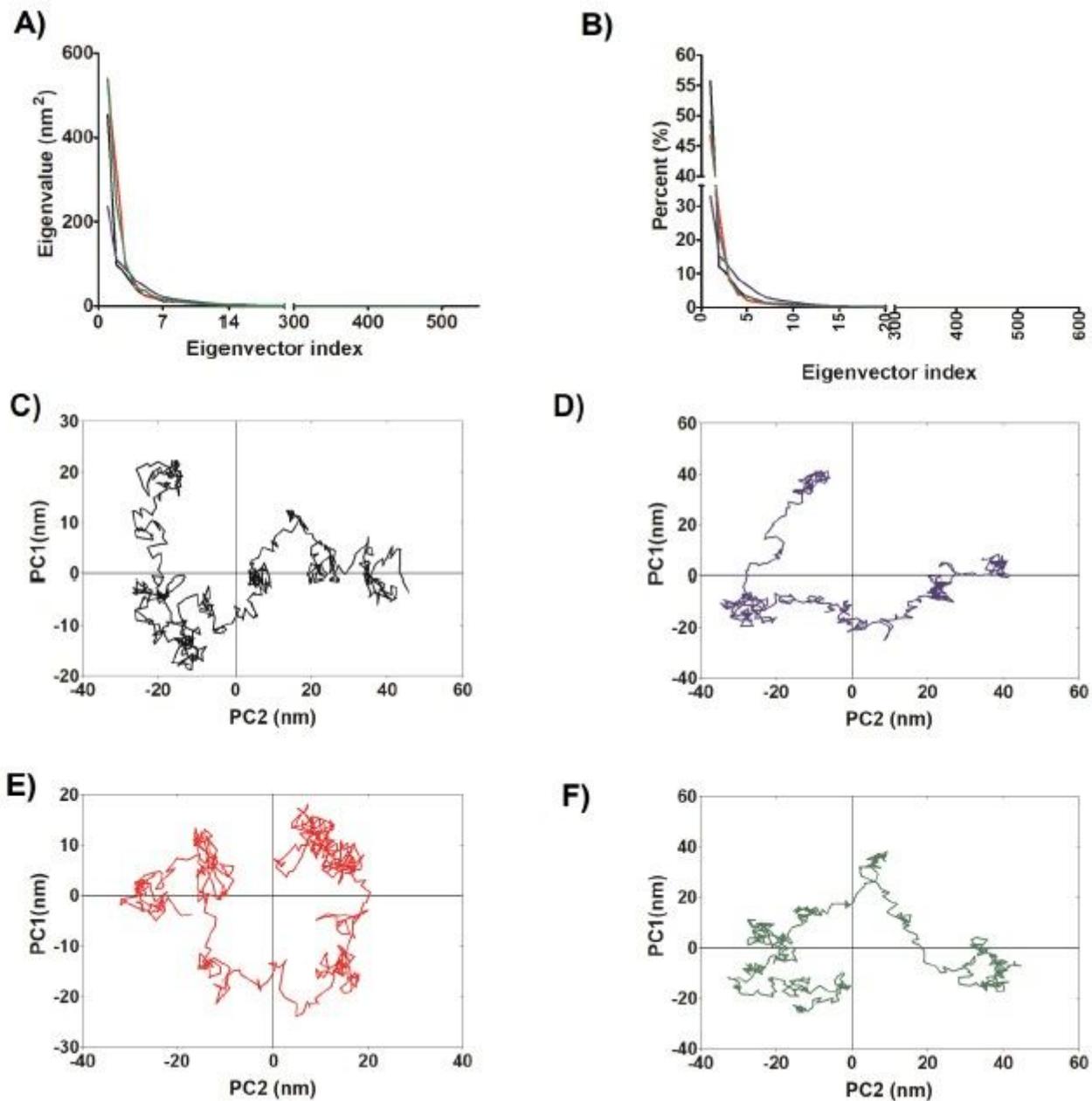


Figure 5

Principal component analysis of WT and mutated spike proteins. A) eigenvalues of the covariance matrix, B) Percentage of each eigenvector vs eigenvalues, C) Projection of the motion in the phase space along the first and second eigenvectors (PC2 vs. PC1) of the WT, D) Projection of the motion in the phase space along the first and second eigenvectors (PC2 vs. PC1) of the D614G, E) Projection of the motion in the phase space along the first and second eigenvectors (PC2 vs. PC1) of the H49Y, F) Projection of the motion in the phase space along the first and second eigenvectors (PC2 vs. PC1) of T573I. Wild type spike protein is depicted in black, D614G in red, H49Y in green, and T573I in purple.

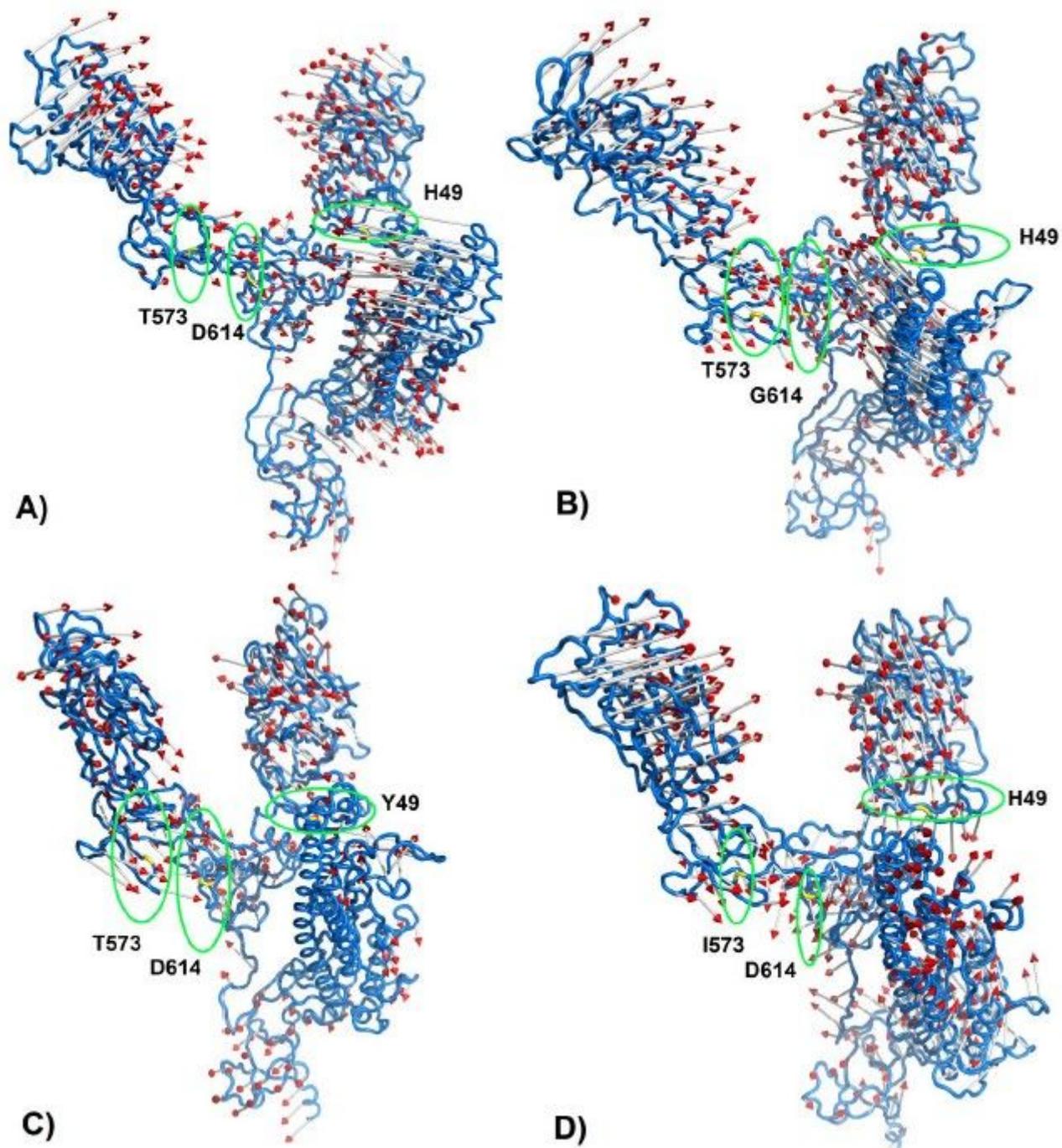


Figure 6

Graphical representation of the two extreme projections. Representation of the two extreme projections along the first eigenvector of MD simulation of A) WT spike protein, B) D614G, C) H49Y and D) T573I.

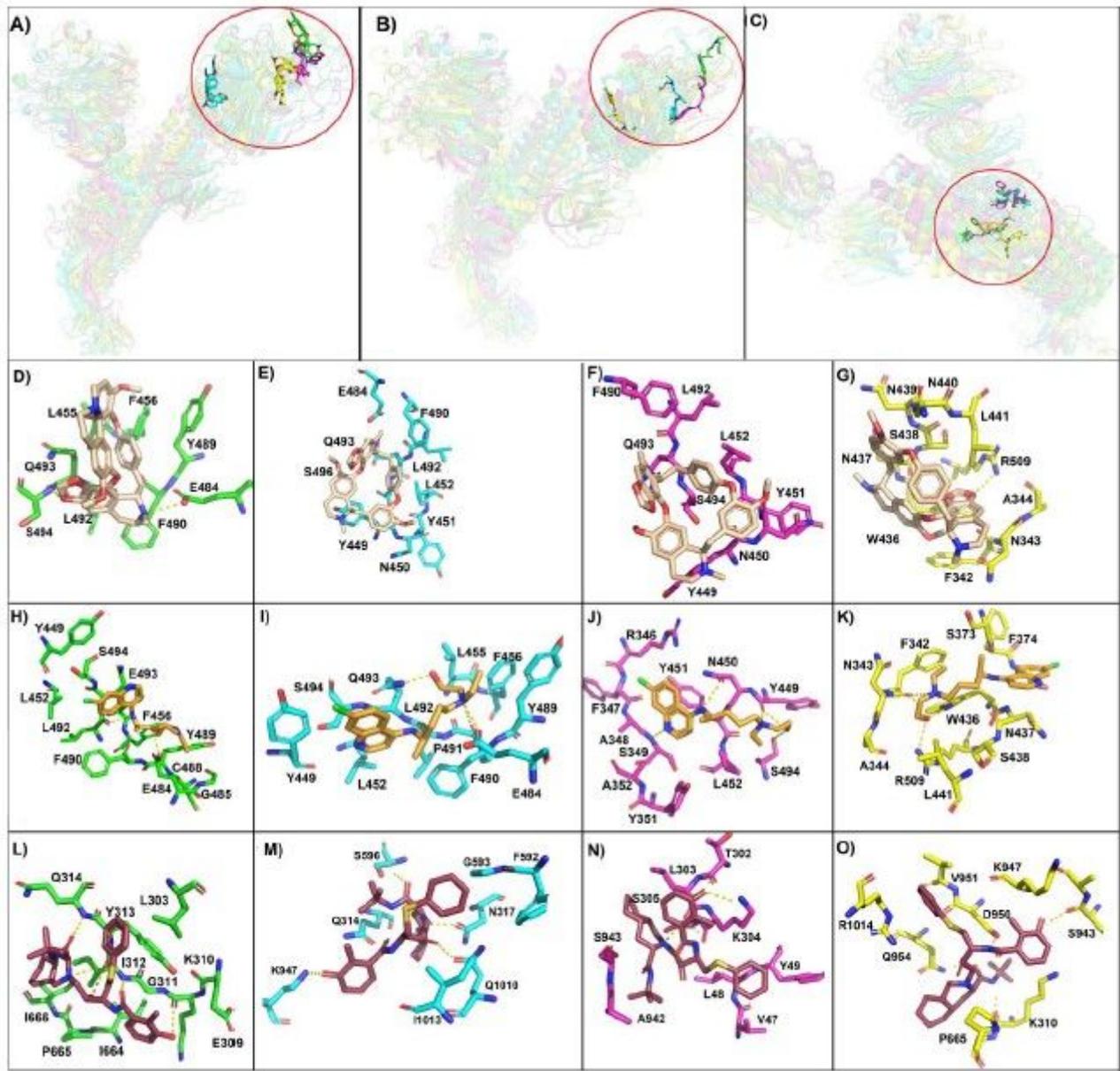


Figure 7

Molecular docking of most populated cluster conformation of spike proteins, D614G, H49Y with experimentally proven compounds. General view of the binding pose obtained by docking of A) cepharanthine, B) hydroxychloroquine, C) Nelfinavir with wild type spike protein (green ribbon), D614G (cyan ribbon), H49Y (magenta ribbon) and T573I (yellow ribbon). Also, the interactions are between Cepharanthine (wheat sticks) with D) wild type spike protein, E) D614G, F) H49Y and G) T573I; and Hydroxychloroquine (orange sticks) with H) wild type spike protein, I) D614G, J) H49Y and K) T573I; and Nelfinavir (raspberry sticks) with L) wild type spike protein, M) D614G, N) H49Y and O) T573I.