



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 website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Comparative study of SARS-CoV-2 infection in different cell types: Biophysical-computational approach to the role of potential receptors

Lenin González-Paz^{a,b,*}, María José Alvarado^a, María Laura Hurtado-León^a, Carla Lossada^c, Joan Vera-Villalobos^d, Marcos Loroño^e, J.L. Paz^f, Laura N. Jeffreys^g, F. Javier Torres^{h,i}, Ysaías J. Alvarado^{c,**}

^a Universidad del Zulia (LUZ), Facultad Experimental de Ciencias (FEC), Departamento de Biología, Laboratorio de Genética y Biología Molecular (LGBM), 4001, Maracaibo, Venezuela

^b Instituto Venezolano de Investigaciones Científicas (IVIC), Centro de Estudios Botánicos y Agroforestales (CEBA), Laboratorio de Protección Vegetal (LPV), 4001, Maracaibo, Venezuela

^c Instituto Venezolano de Investigaciones Científicas (IVIC), Centro de Investigación y Tecnología de Materiales (CITeMA), Laboratorio de Caracterización Molecular y Biomolecular, 4001, Maracaibo, Venezuela

^d Facultad de Ciencias Naturales y Matemáticas, Departamento de Química y Ciencias Ambientales, Laboratorio de Análisis Químico Instrumental (LAQUINS), Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador

^e Departamento Académico de Química Analítica e Instrumental, Facultad de Química e Ingeniería Química, Universidad Nacional Mayor de San Marcos, Lima, Peru

^f Departamento Académico de Química Inorgánica, Facultad de Química e Ingeniería Química, Universidad Nacional Mayor de San Marcos, Lima, Peru

^g Centre for Drugs and Diagnostics, Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

^h Grupo de Química Computacional y Teórica (QCT-UR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia

ⁱ Grupo de Química Computacional y Teórica (QCT-USFQ), Instituto de Simulación Computacional (ISC-USFQ), Departamento de Ingeniería Química, Universidad San Francisco de Quito (USFQ), Quito, Ecuador

ARTICLE INFO

Keywords:

Docking
Cell lines
Macromolecular crowding
Viral spike
Cell susceptibility

ABSTRACT

Cellular susceptibility to SARS-CoV-2 infection in the respiratory tract has been associated with the ability of the virus to interact with potential receptors on the host membrane. We have modeled viral dynamics by simulating various cellular systems and artificial conditions, including macromolecular crowding, based on experimental and transcriptomic data to infer parameters associated with viral growth and predict cell susceptibility. We have accomplished this based on the type, number and level of expression of the angiotensin-converting enzyme 2 (ACE2), transmembrane serine 2 (TMPRSS2), basigin2 (CD147), FURIN protease, neuropilin 1 (NRP1) or other less studied candidate receptors such as heat shock protein A5 (HSPA5) and angiotensin II receptor type 2 (AGTR2). In parallel, we studied the effect of simulated artificial environments on the accessibility to said proposed receptors. In addition, viral kinetic behavior dependent on the degree of cellular susceptibility was predicted. The latter was observed to be more influenced by the type of proteins and expression level, than by the number of potential proteins associated with the SARS CoV-2 infection. We predict a greater theoretical propensity to susceptibility in cell lines such as NTERA-2, SCLC-21H, HepG2 and Vero6, and a lower theoretical propensity in lines such as Calu3, RT4, HEK293, A549 and U-251MG. An important relationship was observed between expression levels, protein diffusivity, and thermodynamically favorable interactions between host proteins and the viral spike, suggesting potential sites of early infection other than the lungs. This research is expected to stimulate future quantitative experiments and promote systematic investigation of the effect of crowding presented here.

* Corresponding author. LUZ-FEC, LGBM, IVIC- CEBA- LPV, Maracaibo, 4001, Zulia, Venezuela.

** Corresponding author. Instituto Venezolano de Investigaciones Científicas (IVIC), Centro de Investigación y Tecnología de Materiales (CITeMA), Laboratorio de Caracterización Molecular y Biomolecular, Maracaibo, Venezuela.

E-mail addresses: lgonzalezpaz@gmail.com (L. González-Paz), alvaradoysaias@gmail.com (Y.J. Alvarado).

<https://doi.org/10.1016/j.combiomed.2022.105245>

Received 14 December 2021; Received in revised form 7 January 2022; Accepted 16 January 2022

Available online 20 January 2022

0010-4825/© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

Since severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was reported, great efforts have been made to understand how the virus enters target cells [1]. The spike protein, a surface protein present in other coronaviruses such as SARS-CoV-1 and MERS-CoV, has been shown to be primarily responsible for the ability of SARS-CoV-2 to interact with receptors expressed on the membrane of host cells [1]. Several proteins have been described that act as facilitators of the interaction between the virus and the target cells, such as neuropilin 1 (NRP1) [1,2], FURIN protease [3], transmembrane serine 2 (TMPRSS2) [3] and angiotensin converting enzyme 2 (ACE2) [4]. CD147 (also known as basigin2 or EMMPRIN) has also been identified as a mediator of the SARS-CoV-2 viral entry [1].

However, recombinant forms of the SARS-CoV-2 spike protein do not interact with CD147, suggesting CD147 and the spike protein do not directly interact [5]. Interestingly, studies based on transcriptomics show different levels of expression of the main receptor ACE2 in human tissues, which could contribute to different levels of cellular susceptibility [6]. Interestingly, ACE2 has a higher expression in the intestinal tract, kidney, testis, gallbladder and heart. Tissues with the lowest expression include the thyroid gland, the liver and the lungs. Interestingly, human lung analysis suggested expression in a very small fraction of cells (<1%), and low expression was also observed in 2–3% and 7% of cells in bronchi and nasal mucosa, respectively [6].

It is well-recognized that SARS-CoV-2 virions enter the host cell by binding to the ACE2, or another suitable receptor. For this to occur, the spike protein must be cleaved by host proteases such as FURIN, TMPRSS or cathepsin L between the S1 and S2 regions causing a conformational change leading to endocytosis [2]. Interestingly, the TMPRSS2 protease presents an expression pattern similar to ACE2 [3]. In addition, it has been reported that the expression of ACE2 could be regulated at the post-translational level by CD147 [1] and that the presence of NRP1 can increase cellular susceptibility by improving the infectivity of SARS-CoV-2 mediating a greater viral entry instead of increased viral binding to the cell membrane. This cellular susceptibility to infection can be further increased when ACE2 and TMPRSS2 are present [2].

The aforementioned variety of possible receptors, as well as their different expression levels, could explain the discrepancies in the various results observed in terms of the susceptibility of the cell lines used in *in vitro* assays against SARS-CoV-2 [7–9] which is very important in such experiments especially when looking to study the infectivity of the virus in the presence and absence of drugs [10,11]. Therefore, for a better understanding of the biological cycle of viral particles, predictive mathematical models have been proposed that allow describing the dynamics of different viral infections and even making recommendations for treatments [12].

In this sense, we applied widely described mathematical models to study viral dynamics [11–13]. In this study, to simulate the susceptibility of various cell lines to infection by SARS-CoV-2, we limited ourselves to considering the expression values described as more favorable for the most commonly described host receptor proteins in literature including NRP1, CD147, FURIN, TMPRSS2 and ACE2 [2,14] as well as less frequently studied ones such as heat shock protein A5 (HSPA5) and angiotensin II receptor type 2 (AGTR2) [2,15]. After reviewing the *Cellosaurus* server to determine the most relevant *in vitro* cell lines we focused on Vero6, A549, HepG2, CaCo2, HEK293, HeLa, U-251 MG, CaLu3, NTERA-2, RPTEC TERT1, RT4, HBEC3-KT and SCLC-21H cell lines [1,5,10,11,16]. Therefore, the purpose of this study is to predict the susceptibility of various cell lines used in *in vitro* assays against SARS-CoV-2, based on various analyzes of the group of proteins described as the most important possible receptors [2].

Moreover, this study sought to guide, by applying a biophysical-computational approach, on appropriate models for *in vitro* replication tests of SARS-CoV-2. These results represent the first comparative analysis of parameters associated with the replication of SARS-CoV-2

considering cell susceptibility in terms of the type, number and level of expression of proteins of interest. The relative affinity of these host proteins and the spike protein, and their theoretical cytoplasmic diffusions, could therefore offer information on the theoretical dynamics of this virus according to cell type, and could be the basis for proposing recommendations for a wide range of *in vitro* experiments.

2. Materials and methods

2.1. Search for cell lines and proteins in databases

Based on what was reported by various studies [1,5,10,11], a total of 13 cell lines associated with the SARS-CoV-2 research were chosen, using the “browser by cell line group” of the *Cellosaurus* server (https://web.expasy.org/cgi-bin/cellosaurus/browse_by_group) developed by Amos Bairoch of the CALIPHO group at the SIB - Swiss Institute of Bioinformatics, Version 38 (using the update from May - 2021) (<https://web.expasy.org/cellosaurus/>). Specifically, the recommended lines in *Cellosaurus* for the culture of SARS-CoV-2 chosen were Vero6 (African green monkey kidney epithelial cells) and CaCo2 (human colorectal adenocarcinoma epithelial cells). Cell lines such as A549 (Human epithelial cells from lung carcinoma) and HEK293 (Human embryonic kidney) were used as they were classified in *Cellosaurus* as useful for SARS-CoV-2 research (<https://web.expasy.org/cellosaurus/sars-cov-2.html>) [16]. Also included were the cell lines HepG2 (Human epithelial cells from liver carcinoma), CaLu3 (Human epithelial cells from lung adenocarcinoma), U-251 MG (Human neuronal - Glioblastoma), NTERA-2 (Kidney & urinary bladder Cancer cell line), RPTEC-TERT1 (Kidney & urinary bladder Human kidney cortex telomerase immortalized cell line), RT4 (Kidney & urinary bladder Urinary bladder Cancer Cell line), HBEC3-KT (Lung, central lung bronchiole, Telomerase immortalized cell line), SCLC-21H (Lung, pleural effusion, Cancer cell line), because they are cell lines in which the expression of most of the receptors considered in this study have been reported according to *The Human Protein Atlas* database (<https://www.proteinatlas.org/>), or because they have been used experimentally in research associated with SARS-CoV-2 (<https://web.expasy.org/cellosaurus/sars-cov-2.html>) [16].

In addition, the HeLa (Human cervix - Cervical adenocarcinoma) cell line was included as a control as this cell line is not classified as suitable for culturing SARS-CoV-2 by the *Cellosaurus* server (<https://web.expasy.org/cellosaurus/sars-cov-2.html>) [16]. In order to search for the proteins described as possible receptors, we started with data from previous studies [1–4]. The presence and level of expression of a total of 7 proteins associated with SARS-CoV-2 infection was examined using the database *The Human Protein Atlas* (<https://www.proteinatlas.org/>). Specifically, we considered 5 proteins described as the most important (NRP1, CD147, FURIN, TMPRSS2 and ACE2) [2,14] as well as 2 proteins with weaker evidence of interaction (HSPA5 and AGTR2) [2]. In the case of cell lines for which no specific data was found in *The Human Protein Atlas* server, their taxonomic group of membership was searched in the NCBI/GenBank and protein-protein BLAST was performed among the possible receptors of the cell line of interest and the receptors expressed in the cell lines present in *The Human Protein Atlas* that were related at the genomic, proteomic and functional level (in relation to the type of tissue/region/cell).

2.2. Design of theoretical expression systems

A set of 25 types of systems (from 1 to 25) were studied which were divided into two groups. The first group consisted of theoretical systems of possible individual expression (from 1 to 13) further subdivided into 7 subtypes (from A-G) which were established in a totally random manner. The second group consisted of theoretical systems of possible multiple expression (from 14 to 25) which represented expression systems that simulated the simultaneous participation of ≥ 2 proteins (up to a

maximum of 7), and subdivided in turn into 13 types (from A-M). All this resulted in 247 possible expression systems that group the cell lines Vero6, A549, HepG2, CaCo2, HEK293, HeLa, U-251 MG, CaLu3, NTERA-2, RPTEC-TERT1, RT4, HBEC3-KT and SCLC -21H, as well as the possible receptors or facilitators NRP1, CD147, FURIN, TMPRSS2, ACE2, HSPA5 and AGTR2.

To delimit the number of simulations of the theoretical expression systems, it was proposed to study individually all the receptors in each cell line, and also due to their relevance, a set of possible “fixed” receptor proteins combinations were proposed assuming a simultaneous and constitutive expression. Specifically, the ACE2 and TMPRSS2 pair was assumed as a main and fixed set of proteins expressed simultaneously, as has been suggested [2].

Consequently, the ACE2-TMPRSS2 pair was studied alone and in conjunction with the second group of proteins that have been suggested as the most relevant (NRP1, CD147 and FURIN) [1–3]. The third group of less relevant proteins (HSPA5 and AGTR2) [2] was simulated only jointly and associated with the ACE2-TMPRSS2 pair. The expression of all the proteins was also simulated simultaneously in each cell line considered (a total of 7). Previous cell line studies have revealed considerable variation in the expression and co-expression of these proteins, which justifies simulating individual and group expression scenarios (systems) where the susceptibility of cells is mediated by the type, number and level of expression of proteins [17].

In view of the above, we assumed that a simulated expression system represents the expression of a specific protein or set of proteins in one of the cell lines considered. In the case of the individual expression systems, the cell lines were numbered, and the receptors were assigned letters, as indicated above, for example, the 1-A system, corresponded to the individual expression of the NRP1 (A) receptor in the Vero6 cell line (see Table 1).

While, in the expression systems of more than one possible receptor, the nomenclature was reversed, and a number was assigned to the proteins followed by a letter to identify the cell line, for example, the 14-

Table 1
Profile of the proposed theoretical expression systems.

System	Cell lines	Individual Expression ^a
1	Vero6	NRP1 (A), CD147 (B),
2	A549	FURIN (C), TMPRSS2
3	HepG2	(D),
4	CaCo2	ACE2 (E), HSPA5 (F),
5	HEK293	AGTR2 (G)
6	HeLa	
7	U251 MG	
8	CaLu3	
9	NTERA-2	
10	RPTEC TERT1	
11	RT4	
12	HBEC3-KT	
13	SCLC-21H	
System	Multiple Expression ^a	Cell lines
14	NRP1+CD147+FURIN + TMPRSS2+ACE2+HSPA5+AGTR2	Vero6 (A), A549 (B),
15	TMPRSS2+ACE2	HepG2 (C), CaCo2 (D),
16	NRP1+CD147+FURIN + TMPRSS2+ACE2	HEK293 (E), HeLa (F),
17	NRP1+FURIN + TMPRSS2+ACE2	U251 MG (G), CaLu3 (H),
18	CD147+FURIN + TMPRSS2+ACE2	NTERA-2 (I), RPTEC TERT1 (J),
19	NRP1+CD147+TMPRSS2+ACE2	RT4 (K), HBEC3- KT (L),
20	FURIN + TMPRSS2+ACE2	SCLC-21H (M)
21	NRP1+TMPRSS2+ACE2	
22	CD147+TMPRSS2+ACE2	
23	CD147+TMPRSS2	
24	HSPA5+AGTR2	
25	HSPA5+AGTR2+TMPRSS2+ACE2	

^a Angiotensin converting enzyme 2 (ACE2), transmembrane serine 2 (TMPRSS2), basigin2 (CD147), FURIN protease, neuropilin 1 (NRP1), heat shock proteinA5 (HSPA5) and angiotensin II receptor type 2 (AGTR2).

A system, corresponded to the co-expression of putative possible receptors NRP1, CD147, FURIN, TMPRSS2, ACE2, HSPA5 and AGTR2 in the same Vero6 (A) cell line. The numbers and letters assigned to the designed systems are completely random, as is the order of the proteins and cell lines considered (see Table 1).

It has been reported that the expression levels of cellular receptors can confer permissiveness for infections including those associated with SARS-CoV-2 [18]. In order to compare the susceptibility of cell lines to SARS-CoV-2 infection, we assumed as a reference the expression level of each possible putative receptor using the consensus transcriptomic data used to classify all genes according to their expression specific tissue, single cell, brain region, blood cell or cell line specific reported in *The Human Protein Atlas*. In this study, the consensus normalized expression (NX) value was used for each gene, which represents the maximum value of NX in the transcriptomics database in *The Human Protein Atlas*. An NX value ≥ 1 is indicative of the expression of the protein in at least one type of tissue/region/cell, while an NX < 1 represents the absence of expression.

However, since the expression levels of some important receptors are very low in some cell lines, it has been suggested to assign minimum expression values as a reference [11]. Therefore, to reduce the error resulting from the limitations that both the databases and the expression level detection methods may present, we assumed that an NX = 0.0 would be equivalent to an NX = 0.01 (see Fig. 1, see Fig. 2). For more details on the normalization of transcriptomic data, as well as its classification and the source of the data, it is recommended to read the “Assays and Annotations section” in *The Human Protein Atlas* [16].

2.3. Viral dynamics data

For illustrative purposes, we considered an *in vitro* data set of SARS-CoV-2 virus infection in cell lines recommended for culturing SARS-CoV-2. In order to observe the theoretical infectivity of the virus as a function of the level of susceptibility mediated by the type, number and level of expression of the proteins associated with virus receptors, we assumed constant values of the viral titer with a multiplicity of infection (MOI) equal to 1 as suggested to ensure the same probability of infection [10, 19,20,21], and especially to consider the recommended viral load to induce cytopathological effects after infection (ρ) [20,21]. We also assumed an initial cell concentration of $\sim 1 \times 10^5$ cells/mL as suggested [1,10,11]. The time period for measuring the infection was set at 24 h post infection (hpi) corresponding to reported values [10]. An effective infection rate (β) equal to 2 was assumed, considering the reported number of basic reproduction value, R_0 [8].

In order to determine the hypothetical concentration of target or susceptible cells based on the type, number and level of expression of the

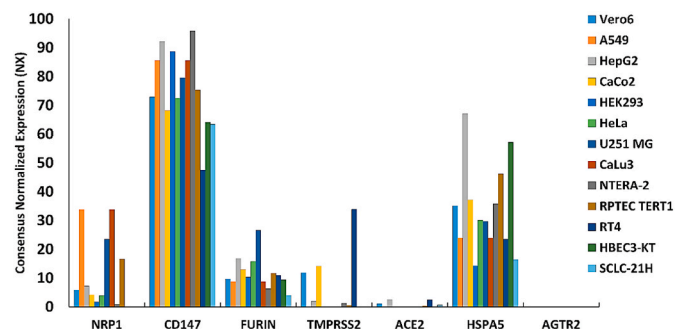


Fig. 1. Cell lines and the level of RNA expression of the potential receptors considered in this study in terms of normalized NX values. The expression levels of transmembrane serine 2 (TMPRSS2), basigin2 (CD147), FURIN protease, neuropilin 1 (NRP1), heat shock protein A5 (HSPA5) and angiotensin II receptor type 2 (AGTR2), are shown.

Cell lines		NRP1	CD147	FURIN	TMPRSS2	ACE2	HSPA5	AGTR2						
Vero6		5.8		72.9		9.7		11.9		1.0		35.2		0.01
A549		33.8		85.6		8.8		0.01		0.01		24.0		0.01
HepG2		7.2		92.2		16.8		1.9		2.5		67.1		0.01
CaCo2		4.2		68.2		12.9		14.1		0.01		37.2		0.01
HEK293		1.7		88.7		10.3		0.01		0.1		14.3		0.01
HeLa		3.9		72.4		15.8		0.01		0.01		30.1		0.01
U251 MG		23.5		79.6		26.6		0.01		0.01		29.7		0.01
CaLu3		33.8		85.6		8.8		0.01		0.01		24.0		0.01
NTERA-2		0.8		95.8		6.3		1.2		0.1		35.8		0.01
RPTEC TERT1		16.6		75.3		11.7		0.5		0.4		46.2		0.01
RT4		0.1		47.6		11.0		33.9		2.5		23.5		0.01
HBEC3-KT		0.01		64.1		9.4		0.01		0.1		57.2		0.01
SCLC-21H		0.01		63.4		4.0		0.01		0.6		16.4		0.01

Fig. 2. Summary of the RNA expression levels of the receptors studied in different cell lines analyzed in the Atlas of Human Proteins. The generated RNA sequencing results are reported as normalized NX values. Numerical values and a pie chart representation of the expression levels of angiotensin converting enzyme 2 (ACE2), transmembrane serine 2 (TMPRSS2), basigin2 (CD147), FURIN protease, neuropilin 1 (NRP1), heat shock protein A5 (HSPA5) and angiotensin II receptor type 2 (AGTR2).

proteins considered, the following was carried out. First, we assumed the same initial total number of cells/mL previously described for each cell line, assigned minimum values of percentage expression to calculate a hypothetical number of cells and the percentage of receptors expressed [11,22]. The percentage expressions of some of the receptors considered very close to the NX values have already been experimentally reported [1], in this sense, we assumed that each NX value would represent the percentage value of susceptible cells in the case of the receptors considered individually, and the mean of the expression systems with 2 or more proteins would equally represent the percentage of susceptible cells in the modeled systems.

2.4. Mechanistic model for the simulation of viral dynamics

The employed viral dynamics model was adjusted using a limited number of target cells. The mathematical expression depends on the available data and the hypothesis to be addressed based on the susceptibility of cells as a function of the type, number and level of expression of receptors, for which we sought to estimate rates of infection of susceptible cells and death of infected cells. Therefore, a widely used model for viral infection called the limited target cell model was used [11,13]. The model included three compartments: susceptible cells (U), infected cells (I) and viral titers (V). The applied model is represented in the following differential equations:

$$\frac{dU}{dt} = -\beta UV \quad (1)$$

$$\frac{dI}{dt} = \beta UV - \delta I \quad (2)$$

The term on the left of the equations represents the change of the variables with respect to time. Parameters β and δ represent the effective infection rates and dead infected cells, respectively. It is considered that the virus (V) infects susceptible cells (U) with a β rate, and that infected cells are eliminated with a δ rate. Based on this model, the basic reproduction number was calculated, R_0 , representing the average number of cells that can be infected from a single infected cell at the beginning of infection:

$$R_0 = \frac{\beta \beta U}{\delta(c + \beta U)} \quad (3)$$

Where virions are released from infected cells productively at a rate (p) per day and are eliminated from the circulation at a rate c or are lost when infecting a target cell, for which it was assumed that the rate of elimination of the virus, c , was equal to 10 days⁻¹ as recommended [11]. For SARS-CoV-2, it has been suggested that the rate p is equal to ≈ 22.7 copies/day per cell, and that since the rate p depends on U , it must be estimated by the product $p \times U$ [12].

2.5. Protein-protein docking and molecular dynamics (MD) of the SARS-CoV-2 spike protein and the proteins of interest

As it has been shown that the spike protein is responsible for the ability of SARS-CoV-2 to interact with receptors expressed on the membrane of host cells leading to virus entry [1]. We performed a comparative analysis using two popular molecular docking models for a rigorous prediction of the standard free energy (ΔG) of protein-protein complexes made up of the spike protein and each of the proteins that behave as possible facilitators of the interaction between the virus and the target cells. The protein structures used were neuropilin 1 (NRP1) (PDB: 7JJC), CD147 (PDB: 3I84), FURIN protease (PDB: 6HZB), serine transmembrane protease 2 (TMPRSS2) (PDB: 7MEQ) [1–3], as well as angiotensin converting enzyme 2 (ACE2) (PDB: 6VW1) which has been described as the main cellular receptor for interaction with the spike protein [4].

Additionally, the less cited HSPA5 (PDB: 6ZYH) and AGTR2 (PDB: 6JOD) were also investigated [2]. The spike protein has been crystallized in a number of forms including in the closed conformation (PDB: 6VXX), the open conformation (PDB: 6VYB) and in fragments including the S1 region (PDB: 7DEO) and the S2 region (PDB: 7COT). We sought to determine if there was the theoretical possibility of a thermodynamically favorable interaction between any region of the viral spike protein and any of the potential receptors considered in this study and so all the spike protein structures listed were used in these experiments. All structures were obtained in PDB format from the RCSB protein database (<https://www.rcsb.org/>), and the quality of the crystal structures was validated using MolProbity (<http://molprobity.biochem.duke.edu/>).

Complexes were constructed using HDock (<http://hdock.phys.hust.edu.cn/>) for hybrid protein-protein docking with template-based modeling, and the web server HawkDock (<http://cadd.zju.edu.cn/hawkdock/>) for the prediction and structural analysis of the protein-protein complex that combines the ATTRACT algorithm for global

macromolecular docking and the HawkRank algorithm for scoring. Additionally, the MM/GBSA method was used to predict free bond energy. As usual, for all the docking methods all the water molecules were removed and the PDB files were separated into two different files, one containing the spike protein and the other containing the structure of the potential receptor. In the sampling of the probabilistically most feasible and thermodynamically most favorable positions in the complexes, only the three runs with the most favorable berth were considered. This criterion was used to discriminate the complexes that would be subjected to further analysis, including molecular dynamics [23].

Additionally, to validate the results of the docking, the DockScore algorithm (<http://caps.ncbs.res.in/dockscore/>) was used to score the protein-protein complexes. This method identified the optimal interactions between the two associated proteins using several putative interface characteristics such as area, short contacts, conservation, spatial clustering, and the presence of hydrophobic and positively charged residues [24].

In order to determine the subset of residues at the interface that account for most of the free binding energy in the protein-protein complex, residues called “binding hot spots” were predicted. For which the KFC Server (Knowledge-based FADE and Contacts) (https://mitchell-web.ornl.gov/KFC_Server/index.php) was used, which provides a web tool to predict protein binding hot spots based on machine learning approaches. For each residue within the link interface, the KFC server characterizes its local framework and compares it to known experimentally determined hot spot environments. Specifically, the server analyzes various chemical and physical characteristics surrounding an interface residue and predicts the residue’s classification using a model trained on previous experimental data [25].

Molecular dynamics (MD) simulations for favorable docking were performed with two purposes: 1) to study the relative stability of the bond; and 2) obtain the minimum energy conformation of the complexes. For a protein-protein complex, the MD system was first relaxed through a series of minimization procedures that include three phases: relaxation, equilibrium, and sampling, as recommended [23]. The MD simulation of the crystalline structures was carried out in an explicit water system. Specifically, the solvation of the system was carried out in an 8.0 Å solvation box. Our MD system also consisted of a copy of the coupled chain region of each protein system. An Amber99SB-ILDN force field was applied to the complex, with TIP3P water model. The whole system was neutralized.

Water molecules were treated as rigid bodies in all models, allowing a simulation time interval of 2 fs. Periodic boundary conditions were applied and the Berendsen algorithm was adopted for the docking of temperature and pressure. After a descent to 5000 steps and a conjugate gradient to 5000 steps energy minimizations with positional constraints on the solute, an initial simulation of 100 ps was performed with the positions of the solute atoms constrained by a constant of force of 10 kcal/(mol·Å²) to allow water to diffuse around the molecule and for equilibrium. The PME method was used to calculate the electrostatic contribution to unbound interactions with a limit of 14.0 Å and a time interval of 1 fs. The cutoff distance of the van der Waals interaction was 14.0 Å. After this equilibration run, the NVT production run (number of particles, volume, temperature) at 300 K was performed with the cell size remaining the same. The SHAKE algorithm (algorithm used to satisfy link geometry constraints) was applied to the system and the time interval was set at 2 fs. Ten structures were obtained every 10 ns as target structures extracted from a total path of 100 ns. For calculations of mean squared deviations (RMSD), the equation,

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n \delta_i^2} \quad (4)$$

where δ_i is the distance between atom i and a reference structure or the average position of the n equivalent atoms. All MD simulations and additional adjustments were performed using COSGENE/myPresto [23,

26]. Additionally, to predict conformational changes in each protein-protein complexes after its interaction, the NMSim software (<https://cpclab.uni-duesseldorf.de/nmsim/main.php>) was used, which is a computational technique that uses a three-step including coarse-graining (CG), normal mode analysis (NMA), and elastic network model (ENM) to provide realistic conformations in reasonable computation times. For this, the minimum energy structure obtained with mypresto at 100 ns was used to also calculate the root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF), using the radius-guided movements (ROG) approach of the NMSim Server. The ROG-guided NMSim simulation is a method to search for a conformation of a given protein-protein complex, allowing to describe the compactness of a protein. In an ROG-guided NMSim simulation, the trajectory is tailored towards the bound structure by selecting the pathway that leads to a decrease in Radius of gyration (Rg), and conformations are generated by structure distortion along directions of random linear combinations of low-frequency normal modes [23].

2.6. Prediction of the theoretical diffusivity of the proteins of interest associated with cellular susceptibility in SARS-CoV-2 infection

It is known that the mobility of biological molecules depends not only on the concentration of the crowding macromolecular solutions, which is generally related to the viscosity η of the crowding solution, but also on the relative size between the crowded and the biological molecule of interest. The diffusion coefficients of the molecules of interest were measured as a function of the concentrations of binders (crowding solution), for which the viscosity of some of the model cell lines used in this study was used for which there are reports of cytoplasmic viscosity such as A549 (adenocarcinomic human alveolar basal epithelial cells with a viscosity $\sim 1.4 \times 10^3$ Pa/s) (designated as model 1) and the HeLa negative control model (with a cytoplasmic viscosity of $\sim 4.4 \times 10^{-2}$ Pa/s) (designated as model 2) [27]. Normal Swiss 3T3 cells (with a viscosity of $\sim 2.4 \times 10^{-2}$ Pa/s) (designated as model 3) were also considered in order to model the mobility of biological molecules considering solutions of high and normal macromolecular crowding [27,28]. Additionally, the cytoplasmic viscosity reported for other cell lines such as ASTC-a-1 (human lung adenocarcinoma with a viscosity of $\sim 1.6 \times 10^{-3}$ Pa/s) (designated as model 4) [29] and H1299 (lung carcinoma with a viscosity of $\sim 1.1 \times 10^2$ Pa/s) (designated as model 5) [30]. The translational diffusion coefficient (D_t) was determined according to the Stokes-Einstein equation (equation (5)) as recommended [27,28]:

$$D_t = \frac{k_B T}{6\pi\eta R} \quad (5)$$

In this model, the only source of dissipative effects (internal friction) is the η of the solvent (cytoplasm) [28]. Where k_B is Boltzmann’s constant, T is absolute temperature, η is viscosity and R is the radius of the particle. The results were compared with the relative viscosity of the water calculated considering the friction ratio (f/f_w) [31] of the minimum energy structures and the molecular weight, both parameters calculated with the HullRad server (<https://hullrad.wordpress.com/>) in each case, and using the module *Protein Research* for diffusion calculation from *Cytiva Life Sciences* (<https://www.cytivalifesciences.com/en/solutions/protein-research/products-and-technologies/diffusion-coefficient-calculator>). On the other hand, when taking the weighted average over the distances in the equation described by Phillips et al., 2009 [32], which describes how a group of molecules located at a point extend over time (t), and taking into account several dimensions (using the Pythagorean theorem), we obtain according to Schavemaker et al., 2018 [33]:

$$t = \frac{d^2}{6nD} \quad (6)$$

Here d is the distance, n is the number of dimensions considered and D , the diffusion coefficient with translational contribution.

Likewise, a preliminary study was carried out to determine the potential theoretical flux of the proteins considered in this study through the cell membrane according to Fick's law:

$$J = -D \left(\frac{\Delta C}{\Delta X} \right) \quad (7)$$

where J ($\text{mol m}^{-2} \text{s}^{-1}$) is the flow, D ($\text{m}^2 \text{s}^{-1}$) the diffusion coefficient, and $\Delta C/\Delta X$ (mol m^{-4}) is the concentration gradient. According to Fick's law [34,35], the flux (J) is directly proportional to the difference in concentration (ΔC) to the diffusion coefficient (D) and inversely proportional to the thickness of the membrane (ΔX). For illustrative purposes only, a ΔC with a diffusion of one third of the hypothetical initial concentration of $\approx 1 \text{ mol m}^{-4}$ was assumed for transmembrane flux, and dimensions reported for the plasma membrane ($\sim 5 \text{ nm}$) were considered for ΔX [36].

3. Results and discussion

3.1. Theoretical propensity for cellular susceptibility depending on the type, diversity and level of expression of proteins associated with SARS-CoV-2 infection

No significant correlation was found between the consensus normalized expression (NX) levels between any of the proteins associated with possible receptors to mediate SARS-CoV-2 infection regardless of the type of cell line, except for a moderate positive correlation between the ACE2 receptor and TMPRSS2 protease ($r = 0.60$) as expected, and a moderate negative correlation between CD147 protein and TMPRSS2 ($r = 0.70$) (see Fig. 3).

In the individually expressed protein systems, it was predicted that 85.42% (211/247) of the theorized expression systems could exhibit a susceptibility (U) of up to 95.8% cells/mL, with a range between 8×10^2 - 1×10^5 cells/mL potentially susceptible. While in multiple expression systems for 2 or more proteins, a susceptibility of up to 48.5% cells/mL was predicted, with a range of 6×10^2 - 5×10^4 cells/mL. In contrast, 14.57% (36/247) of the simulated expression systems were predicted to exhibit a theoretical susceptibility $\leq 2 \times 10^2$ cells/mL. This susceptibility was represented by 11.33% (28/247) of the theoretical individual expression systems, and 3.23% (8/247) in the systems for the expression of 2 or more proteins. These results show that when considering the NX expression values for each protein as a factor of cellular susceptibility, the individual expression systems would present 49.4% more susceptibility compared to the expression systems of ≥ 2 proteins (Tables 1 and 2, Figs. 1–2, Fig. 4).

In the theoretical individual expression systems, CD147 was the protein with the highest expression in all the cell lines considered (13/

13), promoting theoretical susceptibility by an average of 76.26% cells/mL. The proteins with the lowest expression in the individual systems were AGTR2, ACE2 and TMPRSS2 with a frequency of 100% (13/13), 61.54% (8/13) and 53.85% (7/13) in the cell lines studied, respectively, mediating a low theoretical susceptibility of between 0.01% and 0.04% cells/mL. The cell line in which the highest mean individual protein expression was predicted, and therefore, the highest theoretical susceptibility was HepG2, with a mean NX = 26.82, mediating a mean theoretical susceptibility of $\sim 3 \times 10^4$ cells/mL. The cell line in which the lowest expression level was observed was SCLC-21H, with a mean NX = 12.1 and a theoretical susceptibility of $\sim 1 \times 10^4$ cells/mL (see Tables 1 and 2, see Figs. 1 and 2, see Fig. 4).

Specifically, in terms of the simulated models of individual expression, it was predicted that the 9-B system, represented by the individual expression of the CD147 protein in the NTERA-2 line, was the system with the highest theoretical susceptibility, mediating a susceptibility of up to 1×10^5 cells/mL Individual systems 1, 3, 10 and 11 (all of type G); 4 and 9 (types E and G), 2, 5–8 and 12 (D, E and G); and 13 (D and G) presented the lowest theoretical susceptibility with a range between 10 and 100 cells/mL (see Tables 1 and 2, see Figs. 1 and 2, see Fig. 4). All systems designated with the letters D, E and G, were represented by individual expression systems of the TMPRSS2, ACE2 and AGTR2 proteins, respectively. It is important to note that in all theoretical individual expression systems at least two cell lines predicted a theoretical mean susceptibility $\geq 1 \times 10^4$ cells/mL.

Regarding the HeLa cell line used as a susceptibility control, it was predicted as one of the lines with the lowest susceptibility propensity, generally located and with respect to the rest of 12 cell lines below the first 5 cell lines with the best propensity to susceptibility as the case may be, being located in some cases between positions 7–10 (Tables 1 and 2, Figs. 1–2, Fig. 4).

In the specific case of the theoretical models of multiple expression, it was observed that one of the systems with the highest theoretical expression of proteins simultaneously (7 proteins) was represented by the 14-C system, in which the highest propensity was simulated to cellular susceptibility with $\sim 3 \times 10^4$ cells/mL. This system was represented by the HepG2 cell line, hypothetically expressing all the proteins simultaneously (NRP1, CD147, FURIN, TMPRSS2, ACE2, HSPA5 and AGTR2). However, at the global level, the simultaneous expression system that the highest cellular susceptibility according to the protein arrangement was type 23, represented by the theoretical expression of only 2 proteins (CD147 and TMPRSS2) with a global mean of 4×10^4 cells/mL and a susceptibility range between 3×10^4 - 5×10^4 cells/mL depending on the cell type. Specifically, the 23-I system (NTERA-2 cell line) exhibited the highest propensity to cell susceptibility, and the 23-M system (SCLC-21H cell line) exhibited the lowest susceptibility within

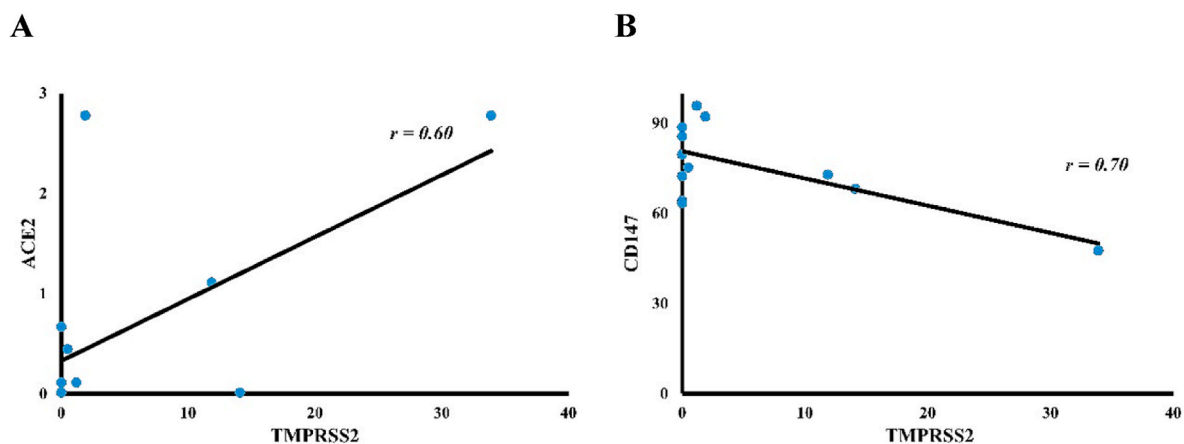


Fig. 3. Proteins associated with SARS-CoV-2 infection with correlation between their levels of consensus normalized expression (NX). A) moderate positive correlation between the receptor ACE2 and TMPRSS2 protease ($r = 0.60$), and B) moderate negative correlation between CD147 protein and TMPRSS2 ($r = 0.70$).

Table 2

Estimates of the number of susceptible cells (U), dead infected cells (δ), productive release rate of virions from cells (p) and basic reproduction number (R_0).

System (s)	U (cell/mL) (mean [min-max])	p (copies/cell*day) (mean [min-max])	δ (cells/mL) (mean [min-max])	R_0 (mean \pm SD)*
1 (A-G)	1.9×10^4 [1.0 x10 ¹ - 7.3 x10 ⁴]	4.4×10^5 [2.3 x10 ² - 1.7 x10 ⁶]	2.1×10^4 [1.4 x10 ¹ - 7.6 x10 ⁴]	19.1 ± 3.8
2 (A-G)	2.2×10^4 [1.0 x10 ¹ - 8.6 x10 ⁴]	4.9×10^5 [2.3 x10 ² - 1.9 x10 ⁶]	2.3×10^4 [1.4 x10 ¹ - 8.9 x10 ⁴]	16.7 ± 5.5
3 (A-G)	2.7×10^4 [1.0 x10 ¹ - 9.2 x10 ⁴]	6.1×10^5 [2.3 x10 ² - 2.1 x10 ⁶]	2.8×10^4 [1.4 x10 ¹ - 9.5 x10 ⁴]	19.2 ± 3.8
4 (A-G)	2.0×10^4 [1.0 x10 ¹ - 6.8 x10 ⁴]	4.4×10^5 [2.3 x10 ² - 1.5 x10 ⁶]	2.1×10^4 [1.4 x10 ¹ - 7.1 x10 ⁴]	18.0 ± 4.9
5 (A-G)	1.7×10^4 [1.0 x10 ¹ - 8.9 x10 ⁴]	3.8×10^5 [2.3 x10 ² - 2.0 x10 ⁶]	1.7×10^4 [1.4 x10 ¹ - 9.2 x10 ⁴]	17.6 ± 4.7
6 (A-G)	1.7×10^4 [1.0 x10 ¹ - 7.2 x10 ⁴]	4.0×10^5 [2.3 x10 ² - 1.6 x10 ⁶]	1.8×10^4 [1.4 x10 ¹ - 7.5 x10 ⁴]	16.6 ± 5.4
7 (A-G)	2.3×10^4 [1.0 x10 ¹ - 8.0 x10 ⁴]	5.2×10^5 [2.3 x10 ² - 1.8 x10 ⁶]	2.4×10^4 [1.4 x10 ¹ - 8.3 x10 ⁴]	16.8 ± 5.6
8 (A-G)	2.2×10^4 [1.0 x10 ¹ - 8.6 x10 ⁴]	4.9×10^5 [2.3 x10 ² - 1.9 x10 ⁶]	2.3×10^4 [1.4 x10 ¹ - 8.9 x10 ⁴]	16.7 ± 5.5
9 (A-G)	2.0×10^4 [1.0 x10 ¹ - 9.6 x10 ⁴]	4.6×10^5 [2.3 x10 ² - 2.2 x10 ⁶]	2.1×10^4 [1.4 x10 ¹ - 9.9 x10 ⁴]	18.7 ± 3.7
10 (A-G)	2.2×10^4 [1.0 x10 ¹ - 7.5 x10 ⁴]	4.9×10^5 [2.3 x10 ² - 1.7 x10 ⁶]	2.3×10^4 [1.4 x10 ¹ - 7.8 x10 ⁴]	18.9 ± 3.8
11 (A-G)	1.7×10^4 [1.0 x10 ¹ - 4.8 x10 ⁴]	3.9×10^5 [2.3 x10 ² - 1.1 x10 ⁶]	1.8×10^4 [1.4 x10 ¹ - 5.0 x10 ⁴]	19.1 ± 3.8
12 (A-G)	1.9×10^4 [1.0 x10 ¹ - 6.4 x10 ⁴]	4.3×10^5 [2.3 x10 ² - 1.5 x10 ⁶]	2.0×10^4 [1.4 x10 ¹ - 6.7 x10 ⁴]	18.6 ± 5.7
13 (A-G)	1.2×10^4 [1.0 x10 ¹ - 6.3 x10 ⁴]	2.7×10^5 [2.3 x10 ² - 1.4 x10 ⁶]	1.3×10^4 [1.4 x10 ¹ - 6.6 x10 ⁴]	18.3 ± 5.6
14 (A-M)	2.0×10^4 [1.2 x10 ¹ - 2.7 x10 ⁴]	4.5×10^5 [2.7 x10 ⁵ - 6.1 x10 ⁵]	2.1×10^4 [1.3 x10 ⁴ - 2.9 x10 ⁴]	20.9 ± 0.1
15 (A-M)	2.9×10^3 [1.0 x10 ¹ - 1.8 x10 ⁴]	6.5×10^4 [2.3 x10 ² - 4.1 x10 ⁵]	3.2×10^3 [1.4 x10 ¹ - 2.0 x10 ⁴]	16.7 ± 4.1
16 (A-M)	2.1×10^4 [1.4 x10 ¹ - 2.6 x10 ⁴]	4.7×10^5 [3.1 x10 ⁵ - 5.9 x10 ⁵]	2.3×10^4 [1.5 x10 ⁴ - 2.8 x10 ⁴]	20.9 ± 0.1
17 (A-M)	6.8×10^3 [1.2 x10 ³ - 1.3 x10 ⁴]	1.5×10^5 [2.4 x10 ⁴ - 2.8 x10 ⁵]	7.6×10^3 [1.4 x10 ³ - 1.4 x10 ⁴]	20.0 ± 0.5
18 (A-M)	2.3×10^4 [1.7 x10 ⁴ - 2.8 x10 ⁴]	5.3×10^5 [3.9 x10 ⁵ - 6.4 x10 ⁵]	2.5×10^4 [1.9 x10 ⁴ - 3.0 x10 ⁴]	21.0 ± 0.1
19 (A-M)	2.3×10^4 [1.6 x10 ⁴ - 3.0 x10 ⁴]	5.2×10^5 [3.6 x10 ⁵ - 6.8 x10 ⁵]	2.5×10^4 [1.9 x10 ⁴ - 3.0 x10 ⁴]	20.6 ± 2.8
20 (A-M)	5.7×10^3 [1.5 x10 ³ - 1.6 x10 ⁴]	1.3×10^5 [3.5 x10 ⁴ - 3.6 x10 ⁵]	3.3×10^3 [9.2 x10 ² - 8.8 x10 ³]	19.0 ± 0.8
21 (A-M)	5.2×10^3 [4.0 x10 ¹ - 1.2 x10 ⁴]	1.2×10^5 [9.1 x10 ² - 2.8 x10 ⁵]	3.0×10^3 [2.7 x10 ¹ - 6.9 x10 ³]	17.9 ± 2.9
22 (A-M)	2.7×10^4 [2.1 x10 ⁴ - 3.2 x10 ⁴]	6.2×10^5 [4.8 x10 ⁵ - 7.3 x10 ⁵]	1.5×10^4 [1.2 x10 ⁴ - 1.8 x10 ⁴]	21.2 ± 0.2
23 (A-M)	4.1×10^4 [3.2 x10 ⁴ - 4.9 x10 ⁴]	9.2×10^5 [7.2 x10 ⁵ - 1.1 x10 ⁶]	2.2×10^4 [1.7 x10 ⁴ - 2.6 x10 ⁴]	21.4 ± 1.5
24 (A-M)		3.8×10^5 [1.6 x10 ⁵ - 7.6 x10 ⁵]		20.5 ± 0.6

Table 2 (continued)

System (s)	U (cell/mL) (mean [min-max])	p (copies/cell*day) (mean [min-max])	δ (cells/mL) (mean [min-max])	R_0 (mean \pm SD)*
	1.7×10^4 [7.2 x10 ³ - 3.4 x10 ⁴]		9.4×10^3 [4.1 x10 ³ - 1.8 x10 ⁴]	
25 (A-M)	9.8×10^3 [3.6 x10 ³ - 1.8 x10 ⁴]	2.2×10^4 [8.2 x10 ⁴ - 4.1 x10 ⁵]	5.6×10^4 [2.1 x10 ³ - 1.0 x10 ⁴]	19.8 ± 0.6

The mean, minimum and maximum values of the simulated expression systems are shown, *, The mean and standard deviation values are displayed.

the group (Tables 1 and 2, Figs. 1–2, Fig. 4).

The systems in which the lowest propensity for cell susceptibility was predicted were type 15 (joint expression of TMPRSS2 and ACE2) with a global mean of 3×10^3 cells/mL, and a susceptibility range of between 10 – 2×10^4 cells/mL depending on the type of cell line. Specifically, in system 15, the 15-K subtype (cell line RT4) exhibited the highest propensity to cell susceptibility, unlike systems 15-B (cell line A549), 15-F (cell line HeLa), 15-G (U-251 MG cell line) and 15-H (CaLu3 cell line) in which the lowest cell susceptibility was predicted (~ 10 cells/mL) (Tables 1 and 2, Figs. 1–2, Fig. 4).

The susceptibility profiles predicted for all cell lines of the systems type 23 (CD147 + TMPRSS2), 19 (NRP1 + CD147 + TMPRSS2 + ACE2), 18 (CD147 + FURIN + TMPRSS2 + ACE2), 16 (NRP1 + CD147 + FURIN + TMPRSS2 + ACE2) and 14 (NRP1 + CD147 + FURIN + TMPRSS2 + ACE2 + HSPA5 + AGTR2) a theoretical mean susceptibility $\geq 1 \times 10^4$ cells/mL in 100% of the cell lines considered. Whilst in system 24 (HSPA5 + AGTR2) it was 85%, in system 22 (CD147 + TMPRSS2 + ACE2) it was 77%, in system 17 (NRP1 + FURIN + TMPRSS2 + ACE2) it was 31%, and in systems 21 (NRP1 + TMPRSS2 + ACE2) was 23%. These results show that $\approx 66\%$ (6/9) of the most prevalent expression profiles had the presence of CD147 + TMPRSS2 proteins (Tables 1 and 2, Figs. 1 and 2).

Regarding the amount of virions that could theoretically be released from susceptible, productively infected cells, the type 23 system was predicted to have a rate (p) with a mean of 9.2×10^5 copies/day per cell, with minimum values of 7.2×10^5 and maximum values of 1.1×10^6 copies/day per cell, represented by the cell lines NTERA-2 (system 23-I) and SCLC-21H (system 23-M) respectively. Globally, it was predicted that systems 14, 16, 18, 19, 22–24 could theoretically release the largest number of virions from infected susceptible cells (classified from A-M) at a rate (p) of between 1.6×10^5 - 7.6×10^5 copies/day per cell. While in system 21 the lowest p rate was predicted with a mean of 1.2×10^5 copies/day per cell, and a rate of 9.1×10^2 to 2.8×10^5 copies/day per cell corresponding to the RT4 cell lines (system 21- L) and HBEC3-KT (21-K system) respectively (Tables 1–2, Figs. 1 and 2).

The theoretical number of dead infected cells (δ) was calculated suggesting systems 18 and 19 could generate the highest number of dead cells with a mean of 2.5×10^4 cells/mL with a range of values from 1.9×10^4 to 3.0×10^4 cells/mL, in SCLC-21H cell lines (18-M system) and (19-M system) respectively. In general terms, systems 14, 16–19 and 22–24 (sub-classified from A-M) could generate the highest number of dead cells, specifically between 1.4×10^4 - 3.9×10^4 cells/mL. In comparison, system 21 had the lowest number of dead infected cells predicted with a mean of 3.0×10^3 cells/mL, and minimum and maximum values of 2.7×10^1 and 6.9×10^3 cells/mL corresponding in the same way to the RT4 cell lines (21-L system) and HBEC3-KT (21-K system) respectively (Tables 1 and 2, Figs. 1 and 2).

In order to validate the theoretical susceptibility of the cell lines considered under the conditions of this study, the basic reproduction number (R_0) which represents the average number of susceptible cells infected from a single cell infected at the beginning of the infection, was also calculated predicting in systems 18, 22 and 23 giving a value of $R_0 \geq 21$. Specifically, in system 18 all cell lines, except in systems 18-L

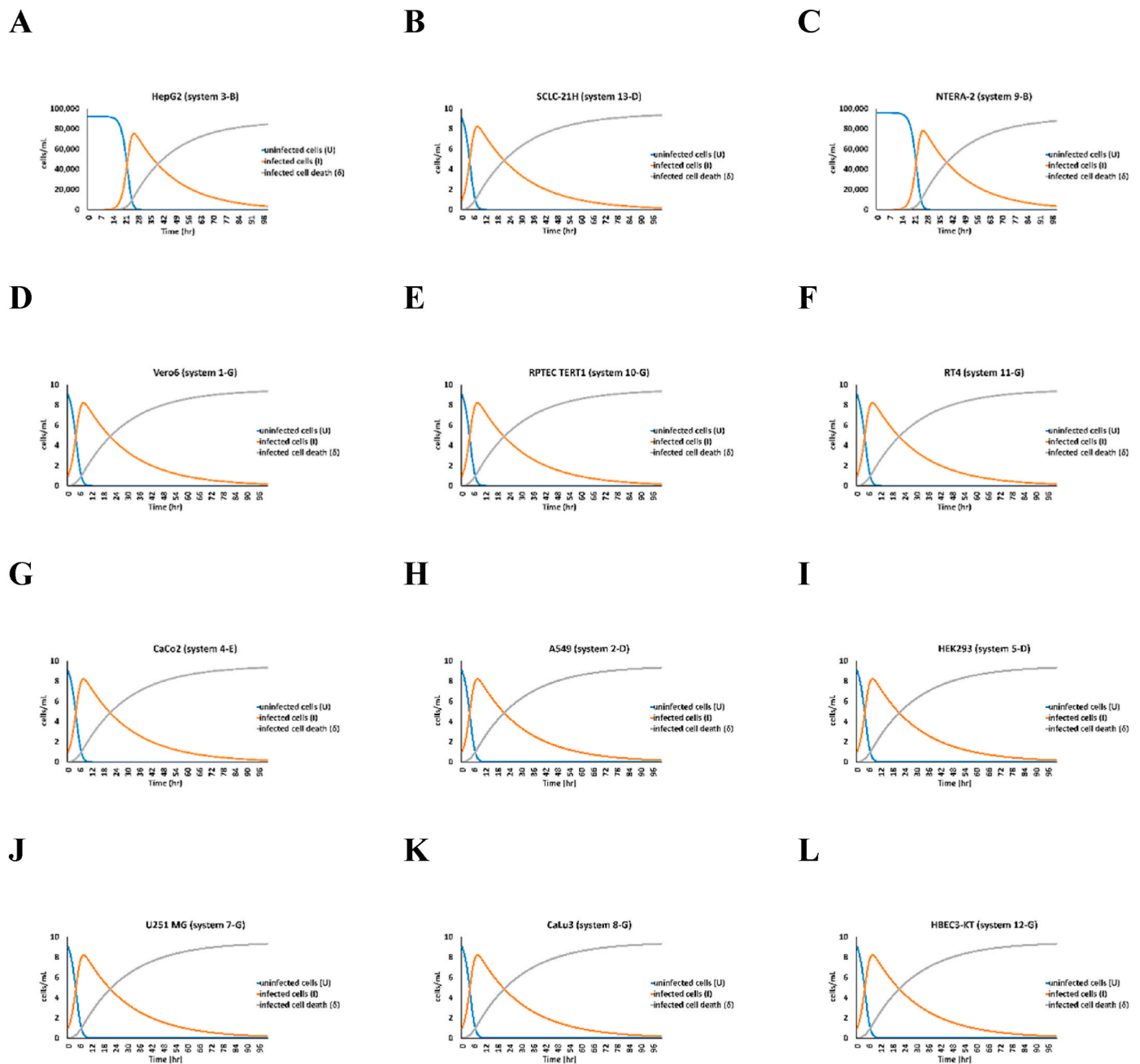


Fig. 4. Mechanistic model of the estimates of the rates of uninfected cells (U), infected cells (I) and death of infected cells (δ). For illustrative purposes, only a representative graph of the individual expression systems cited in the text is shown for each cell line studied using the limited target cell model. Additional graphics are shown in supplementary material. A) Individual expression of CD147 in the HepG2 cell line (3-B system), B) individual expression of FURIN in the SCLC-21H cell line (13-D system), C) individual expression of CD147 in the NTERA-2 cell line (9-B system), D) individual expression of AGTR2 in the Vero6 cell line (1-G system), E) Individual expression of AGTR2 in the RPTEC TERT1 cell line (10-G system), F) individual expression of AGTR2 in cell line RT4 (11-G system), G) Individual expression of ACE2 in the CaCo2 cell line (4-E system), H) individual expression of TMPRSS2 in the A549 cell line (2-D system), I) individual expression of TMPRSS2 in the HEK293 cell line (5-D system), J) individual expression of AGTR2 in the U251 MG cell line (7-G system), K) individual expression of AGTR2 in the CaLu3 cell line (8-G system), L) individual expression of AGTR2 in the HBEC3-KT cell line (12-G system).

(HBEC3-KT) and 18-M (SCLC-21H), an $R_0 \geq 21$ was predicted, as in 22, with the exception of system 22-M (SCLC-21H), while in system 23 the 23-C (HepG2) and 23-I (NTERA-2) models presented the highest basic reproduction value with an $R_0 \geq 22$. Unlike systems 2, 6, 7, 8 and 15 in which the lowest basic reproduction was predicted with an $R_0 \leq 16$, specifically the 2-D, 2-E and 2-G systems, as in the case of systems 6, 7 and 8 D, E and G, which are represented by the cell lines CaCo2, HEK293 and U251 MG, respectively (see [Tables 1 and 2](#), see [Figs. 1 and 2](#)).

It is important to note that the type 2, 6, 7 and 8 systems subdivided

into D, E and G corresponded to the simulation of individual expression of the receptors/mediators ACE2, TMPRSS2 and AGTR2 respectively, as do the multiple expression systems of type 15-B, 15-F, 15-G and 15-H, constituted by the cell lines A549, HeLa, U251 MG and CaLu3 under each hypothetical condition of joint expression of TMPRSS2 and ACE2. The difference between the predicted R_0 values was statistically significant ($p < 0.01$) under the conditions of this study and suggests a differential viral kinetic behavior dependent on the degree of cellular susceptibility. This behavior was influenced to a great extent by the type

and expression level of potentially receptor/mediator proteins associated with SARS-CoV-2 infection, and to a lesser extent by the number of potential receptor/mediator proteins expressed (Tables 1 and 2, Figs. 1 and 2). All estimates of U , δ , p and R_0 are shown in detail in the supplementary material (Table S1).

The correlation found between the level of expression of the ACE2 protein and TMPRSS2 was expected because a positive correlation between ACE2 and TMPRSS2 has already been reported in most organs [37]. Indeed, other coronaviruses also utilize ACE2 and TMPRSS2 for viral entry, for example SARS-CoV-1 [37]. TMPRSS2 facilitates viral entry by cleaving the SARS-CoV-2 spike protein between the S1 and S2 subunits causing a dramatic structural change leading to endocytosis [38–41]. However, it is important to note that, although a moderate positive correlation was found between the expression of ACE2 and TMPRSS2 under the conditions of this study, an absence of correlation has been reported between the expressions of the ACE2, TMPRSS2 and the FURIN protein (also considered in this study) in healthy volunteers with various clinical characteristics [38].

Although we found a negative correlation between TMPRSS2 and CD147, it has been reported that some cell types are more likely to be infected with SARS-CoV-2 through the CD147 receptor and the TMPRSS2 protease than through ACE2 [42]. This alternative infection method is likely due to the low expression levels observed for ACE2 whilst CD147 has a higher level of expression in various cell lines and tissues as observed in this study [42]. Of note, it was found that receptors such as ACE2, TMPRSS2 and AGTR2 have the lowest levels of expression according to the consensus transcriptomics data of *The Human Protein Atlas*, which corresponds to what has been reported in other studies [42,43].

Our observations show that the higher probability of an infection could be mediated by the individual expression of certain candidate receptors than by the simultaneous expression of several of them. This corresponds to what has been reported by some authors who point out that the expression of several proteins simultaneously is an event with considerable variation from one cell to another, and heterogeneous co-expression patterns have been observed including cells that do not express all genes even individually [17].

It has been reported that the expression of ACE2 in some tissues can be approximately 1/3 of the level of expression of TMPRSS2 in the same tissues, [44,45]. Furthermore, it has been reported that ACE2 and TMPRSS2 expression may be associated with the respiratory tract, for example, ACE2 may become downregulated in nasal epithelial cells whilst TMPRSS2 may be upregulated in bronchial cells. Additionally, it has been described that some respiratory epithelia can be negative for the simultaneous expression of the ACE2 receptor and TMPRSS2 but positive for the simultaneous expression of CD147 and FURIN [46].

Although more attention has been paid to the ACE2 receptor and TMPRSS2 as well as to cells associated with the respiratory tract which showed important basic reproduction numbers in this study. Our predictions show very low levels of expression in all the types of cells studied indicating a possible greater theoretical propensity to susceptibility in cell lines such as NTERA-2, SCLC-21H, HepG2 and Vero6, and a lower propensity to infectivity in lines such as CaLu3, RT4, HEK293, A549 and U-251 MG. This corresponds to the RNA sequence profiles for our proteins of interest that showed high levels of expression in epithelial tissues outside the respiratory tract, indicative that cells of another nature may be particularly susceptible to infection by SARS-CoV-2 [44,45]. In the case of the CD147 protein, a higher level of expression has been observed in brain cell lines and tissues, as was observed in this study [42].

It has also been shown that ACE2 and TMPRSS2, in addition to being expressed in cells associated with the lung and esophagus, are expressed at very high levels in human colorectal, stomach and liver-like cell, indicating that the gastrointestinal system could also be a possible route of infection by SARS-CoV-2 [44,45,47]. Interestingly, the ACE2 gene has been shown to be negatively regulated in some cells during SARS-CoV-1

infection [45]. These variations in the susceptibility of cells associated with the respiratory system have already been described in other studies that have demonstrated topographic differences in the replication of SARS-CoV-2 in the respiratory tract and almost undetectable levels of subgenomic viral material at the bronchoalveolar level [48].

The susceptibility profiles predicted as the most prevalent and capable of theoretically mediating the greater propensity to infection by SARS-CoV-2 based on the type, number and level of expression of the potential receptors coincided in their entirety with the presence of the CD147 and TMPRSS2 proteins, with or without the presence of ACE2 and the rest of the receptor candidates. This corresponds to what has been previously described in brain cell lines where it has been proposed that there may be a greater probability of infection with SARS. The viral spike protein has also been proposed to bind to the extracellular protease CD147 to mediate cell invasion by SARS-Cov-2 [49]. However, this is debated as purified recombinant spike protein is unable to interact with CD147 [5]. Of note, it has been experimentally confirmed that single nucleotide polymorphisms (SNPs) affect expression with a level of alteration that is often enough to induce phenotypic changes [49,50]. Therefore, although our results are based on experimental and transcriptomic kinetic data, they must be confirmed experimentally due to the inherent genetic variability.

3.2. Theoretical affinity of the spike protein and the proteins of interest associated with cellular susceptibility in SARS-CoV-2 infection

Thermodynamically favorable dockings were predicted between the spike protein (for all its conformations and regions) and the target proteins considered and described as potential facilitators of SARS-CoV-2 infection. As expected, the HDock algorithm predicted a thermodynamically favorable docking between the spike protein and the ACE2 and TMPRSS2 proteins described with a mean of $\Delta G = -294.21$ kcal/mol and $\Delta G = -300.79$ kcal/mol, respectively. The most favorable docking was predicted between the AGTR2 protein and the spike in all its conformations and tested regions, with a thermodynamic mean $\Delta G = -380.42$ kcal/mol, being only surpassed by ≈ -1 kcal/mol by the docking between ACE2 and the S1 spike region. Specifically, the most favorable docking of the AGTR2 protein with the spike regions was predicted for the S2 portion, followed by the binding with the TMPRSS2 (-262.47 kcal/mol) for the same S2 region (see Table 3).

It is important to highlight that according to the HDock algorithm, the S1 portion was the most favored spike region to establish thermodynamically feasible interactions with all the potential receptors considered, compared to the S2 region with a statistically significant difference ($p < 0.05$) at the energy level. These results correspond to the MM/GBSA measurements using the HawkDock method where it was predicted that for the S1 region a favorable docking was observed in a similar way with TMPRSS2 (-49.94 kcal/mol) and ACE2 (-49.53 kcal/mol). While for the S2 region the most thermodynamically favorable docking with the MM/GBSA approach was predicted with AGTR2 (-68.29 kcal/mol) followed by TMPRSS2 (-47.19 kcal/mol). In relation to the open and closed conformation of the spike protein, the most favorable interaction at the thermodynamic level was predicted with the AGTR2 protein with MM/GBSA values of -63.71 kcal/mol and -60.49 kcal/mol, respectively. It is important to note that the AGTR2 protein presented the most favorable docking with all the conformations and tested regions of the spike, except for the affinity for the S1 region, and in comparison with the rest of the potential receptors, exhibiting a thermodynamic mean $\Delta G = -57.94$ kcal/mol (Table 3).

In order to study the contribution of the residues in the predicted interactions, the subset of residues designated as binding hot spots was determined, which constitute a small fraction of the interface residues in the protein-protein complexes that are determinants in the binding affinity calculated. In this sense, after calculating the binding hot spots considering characteristics such as interface solvation, atomic density and plasticity, the highest number of hot spots between the spike protein

Table 3

Comparative analysis of the relative macromolecular binding affinity of the spike and each of its domains (S1 and S2) with each of the proteins of interest using different docking methods and scoring functions.

Protein	PDB	HDOCK (kcal/mol)				MM/GBSA (kcal/mol)*			
		Spike (Open)	Spike (Close)	S1	S2	Spike (Open)	Spike (Close)	S1	S2
NRP1	7JJC	-274.84	-277.27	-269.17	-244.32	-61.81	-37.79	-53.60	-35.93
CD147	3I84	-218.99	-244.20	-210.55	-183.22	-46.47	-45.17	-46.57	-27.40
FURIN	6HZZ	-290.38	-289.39	-292.35	-251.20	-49.54	-39.84	-44.43	-37.58
TMPRSS2	7MEQ	-341.26	-331.70	-267.71	-262.47	-31.83	-36.36	-49.94	-47.19
ACE2	6VW1	-304.94	-269.98	-352.45	-249.45	-21.90	-24.18	-49.53	-28.02
HSPA5	6ZYH	-266.58	-270.19	-233.86	-208.61	-40.04	-35.80	-29.75	-25.32
AGTR2	6JOD	-446.61	-366.70	-351.16	-357.21	-63.71	-60.49	-39.26	-68.29

For the spike protein, the open (PDB: 6VYB) and closed (PDB: 6VXX) conformations of the structure are considered; *. For the energy calculation with the molecular mechanics approach combined with generalized Born calculations and continuous surface area solvation (MM/GBSA) calculations to predict the binding free energy of the protein-protein complex per residue, the HawkDock web server was used.

and the ACE2 receptor was predicted, which corresponds to the energy values of binding previously calculated. Specifically, more than 20% of the residues present in the interface correspond to residues located at a distance $<4 \text{ \AA}$ (Table 4, Figs. 5 and 6).

The highest number of residues with important interactions were predicted between the ACE2 receptor and the S1 region of the spike, contributing to the interface with $\approx 28\%$ (10/35) and $\approx 20\%$ (6/30) of residues with the most interaction probable, respectively. In contrast, the predicted interaction between the ACE2 receptor and the S2 region were $\approx 21\%$ (6/28) and $\approx 5\%$ (1/18) respectively. The largest contribution in the interaction between the potential NRP1 receptor and the S1 region was predicted with $\approx 25\%$ (6/24) and $\approx 10\%$ (3/30) of binding hot spots, respectively. It is important to consider that according to the Hawkdock algorithm the interaction between the receptor potential NRP1 and the S1 region was the most favorable energy level ($\Delta G = -292.35 \text{ kcal/mol}$) according to the MM/GBSA approach. 71% (5/7) of the dockings with the highest number of residues with important interactions were predicted in the S1 region, compared to $\approx 29\%$ (2/7) in the S2 region (Table 4; Figs. 5 and 6).

The only interfaces in which a larger subset of residues designated as binding hot spots was predicted in the S2 region were between the potential CD147 receptor and the S2 region with $\approx 8\%$ (2/24) and $\approx 19\%$ (3/16) of binding hot spots respectively and between HSPA5 and the S2 region with $\approx 21\%$ (5/24) and $\approx 10\%$ (2/21) of binding hot spots respectively. The interfaces with the least favorable interactions were predicted between the FURIN protease and the S2 region, and between AGTR2 and the S1 region in which, under the conditions of this study, it

was not possible to determine more than 6 relevant binding hot spots (Table 4; Figs. 5 and 6).

The calculation of RMSD and RMSF allowed observations that the complexes constructed were stable during the simulation period, exhibiting fluctuations $\leq 2.5 \text{ \AA}$ in all cases and with respect to the proteins proposed as possible free receptors of the S1 or S2 portions of the viral spike, except for the RMSD values of the CD147 + S1, FURIN + S1 and HSPA5 + S2 complexes. In particular, the systems that presented the lowest fluctuations were the complexes formed by the S1 region of the spike and the TMPRSS2, NRP1 and ACE2 proteins with an RMSD and RMSF $\leq 0.5 \text{ \AA}$, as well as the predicted RMSD and RMSF values for the complex made up of the S2 region and the ACE2 protein (RMSD and RMSF $\leq -0.6 \text{ \AA}$). The complexes of the two regions of the spike with the proteins AGTR2, TMPRSS2, CD147, and of the S2 region with NRP1 were the systems in which the lowest fluctuations were predicted, with RMSD and RMSF values lower than those calculated from the free proteins (Fig. 7; Fig. 8).

In the case of the CD147 + S1 and CD147 + S2 complexes, the lowest fluctuations were predicted compared to those observed for their corresponding free protein with a difference of RMSD $\approx 3.5 \text{ \AA}$ and RMSD $\approx 1.5 \text{ \AA}$ respectively, and a difference in terms of RMSF $\approx 1 \text{ \AA}$ and RMSF $\approx -0.01 \text{ \AA}$ respectively. These results show that the docking with the portions of the viral spike protein studied can stabilize the trajectories of the complexes formed with the CD147 protein, unlike the free protein, which shows high flexibility and potential conformational deformability. On the contrary, the complexes formed by FURIN and HSPA5 with each of the regions of the viral spike showed a greater fluctuation

Table 4

Comparative analysis of the residues called hot spots in the relative macromolecular junction of the spike domains (S1 and S2) with each of the proteins of interest.

PDB	Spike	Residues determined as Binding Hot Spots ^a							
		NRP1 (n = 30/24)	CD147 (n = 27/18)	FURIN (n = 21/27)	TMPRSS2 (n = 28/31)	ACE2 (n = 30/35)	HSPA5 (n = 25/28)	AGTR2 (n = 18/23)	
7DEO	S1	Glu16(A), Glu21(A), Glu24(A), Glu43(A), Glu51(A), Asp128(A)/Glu340(B), Glu471(B), Glu484(B)	Glu340(A), Glu516(A)/Glu49(B), Glu64(B), Glu84(B), Glu92(B)	Asp23(A), Asp284(A), Asp322(A)/Asp405(B), Glu406(B)	Glu289(B), Glu329(B), Asp359(B), Asp491(B)/Asp405(A), Glu406(A), Asp420(A), Glu484(A)	Asp118(A), Glu122(A), Glu132(A), Asp139(A), Glu142(A), Asp274(A), Asp277(A), Asp281(A), Asp285(A), Asp317(A)/Glu340(B), Asp442(B), Glu465(B), Asp467(B), Glu471(B), GLU484(B)	Glu44(A), Glu46(A), Glu116(A), Glu117(A), Glu128(A), Glu131(A), Asp159(A)/Glu471(B), Glu484(B)	Glu340(B)	
		NRP1 (n = 24/16)	CD147 (n = 16/24)	FURIN (n = 19/17)	TMPRSS2 (n = 21/31)	ACE2 (n = 18/28)	HSPA5 (n = 21/24)	AGTR2 (n = 34/36)	
7COT	S2	Glu58(A), Asp59(A), Glu113(A), Glu151(A)/Glu106(B)	Glu119(A), Asp123(A)/Asp32(B), Glu84(B), Glu92(B)	Asp430(B)	Glu406(B), Asp482(B)/Asp27(A), Glu106(A)	Asp195(A), Glu417(A), Glu509(A), Glu518(A), Glu553(A), Glu571(A)/Asp70(B)	Glu89(B), Glu121(B), Asp131(B), Asp281(B), Glu314(B)/Asp70(A), Asp92(A)	Glu257(A)/Asp92(B), Glu106(B)	

^a Only the residues present at the interface are shown that represent most of the binding free energy (*hot spots*) in the protein-protein complex according to the KFC (Knowledge-based FADE and Contacts) model. *n* (total number of residues at the interface (number of S domain residues/number of receptor type residues), including "no hot spots" residues).

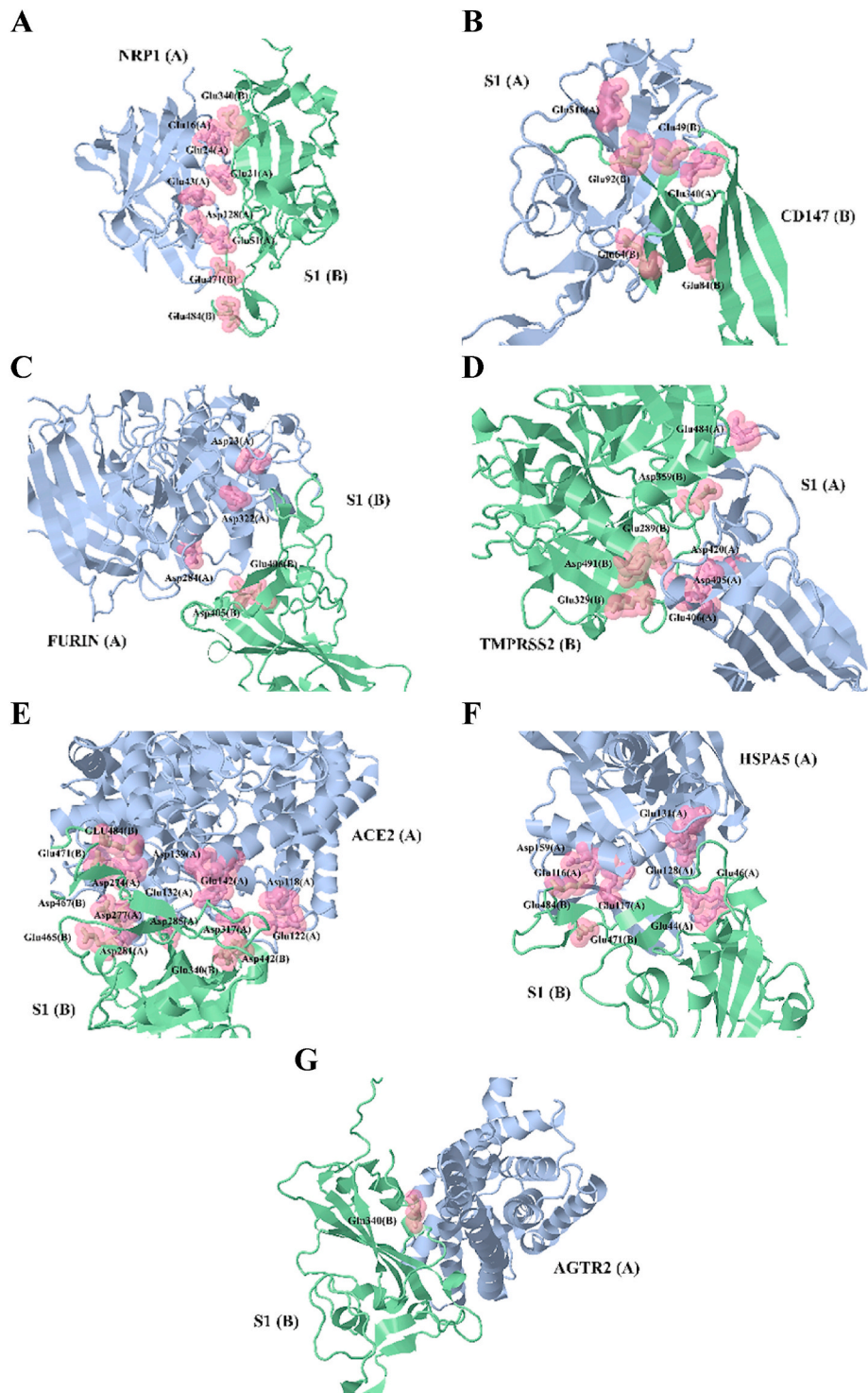


Fig. 5. Graphic representation of the interface between the residues called hot spots in the relative macromolecular junction of the spike S1 domain with each of the proteins of interest. A) Neuropilin 1 (NRP1), B) Basigin2 (CD147), C) Furin protease, D) Transmembrane serine 2 (TMPRSS2), E) Angiotensin converting enzyme 2 (ACE2), F) Heat shock protein A5 (HSPA5) and G) angiotensin II receptor type 2 (AGTR2). Only the residues present at the interface are shown that represent most of the binding free energy in the protein-protein complex according to the KFC (Knowledge-based FADE and Contacts) model.

with respect to the free proteins both in terms of RMSD and RMSF, specifically a difference was observed between the Furin + S1 complex and the Furin + S2 complex with respect to the control represented by values of $\text{RMSD} \approx -4 \text{ \AA}$ and $\text{RMSD} \approx -1 \text{ \AA}$ respectively and values of $\text{RMSF} \approx 2 \text{ \AA}$ and $\text{RMSF} \approx -0.1 \text{ \AA}$ respectively. On the other hand, the complex HSPA5 + S1 and HSPA5 + S2 with respect to the control showed an $\text{RMSD} \approx -2 \text{ \AA}$ and $\text{RMSD} \approx -5 \text{ \AA}$ respectively, as well as values of $\text{RMSF} \approx -1 \text{ \AA}$ and $\text{RMSF} \approx -2 \text{ \AA}$ respectively. Although the trajectories calculated for the complexes built between the spike and the Furin or HSPA5 proteins showed greater fluctuation and, therefore, less

stability, the RMSD and RMSF values showed slightly favorable dockings throughout the simulation, especially between Furin and the S2 region, and HSPA5 with the S1 region (Fig. 7; Fig. 8).

It is important to note that the ACE2 protein with which a stable and favorable binding energy was predicted, as well as the greater number of determining residues in the binding affinity in the complexes formed with any of the regions of the viral spike protein, was possibly expressed under the conditions of this study in a concentration of between 1 and 3% of the cells and only in $\approx 23\%$ (3/13) of the cell models studied, specifically in the HepG2, RT4 and Vero6 lines. Thus, the theoretical

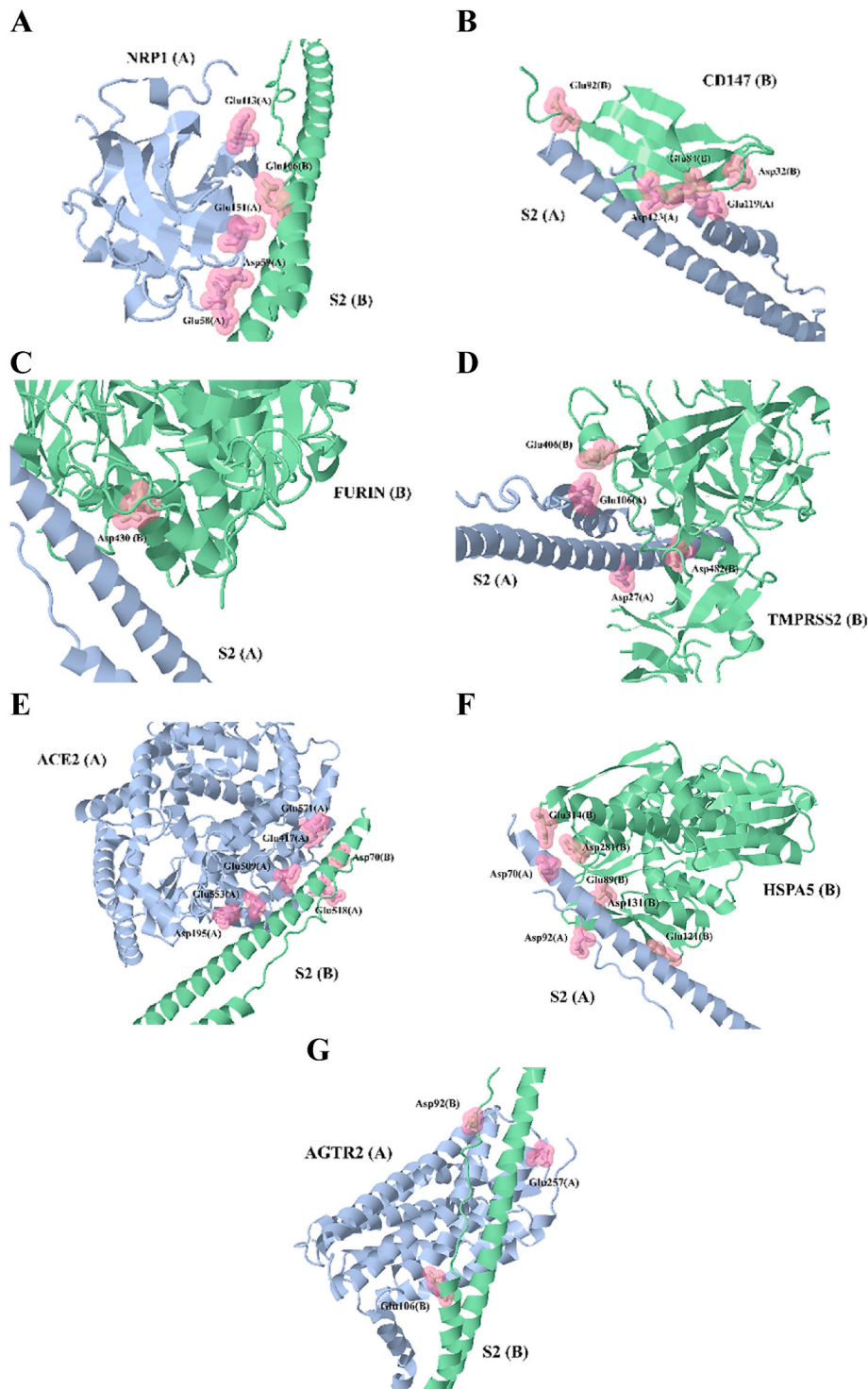


Fig. 6. Graphic representation of the interface between the residues called hot spots in the relative macromolecular junction of the spike S2 domain with each of the proteins of interest. A) Neuropilin 1 (NRP1), B) Basigin2 (CD147), C) FURIN protease, D) Transmembrane serine 2 (TMPRSS2), E) Angiotensin converting enzyme 2 (ACE2), F) Heat shock protein A5 (HSPA5) and G) angiotensin II receptor type 2 (AGTR2). Only the residues present at the interface are shown that represent most of the binding free energy in the protein-protein complex according to the KFC (Knowledge-based FADE and Contacts) model.

affinity of ACE2 and its level of expression could favor the infection of the cell types associated with the urinary and hepatic systems.

Due to its thermodynamically stable and favorable affinity, as well as due to a significant number of determining residues in the binding affinity for spike, the TMPRSS2 protein could be expressed depending on the cell type between 1 and 33% of the cells, and in the $\approx 38\%$ (5/13) of the cell models considered, specifically in the CaCo2, RT4, Vero6, HepG2 and NTERA-2 lines. This would possibly also contribute to a better infection of the cell types associated with the urinary and liver systems, including to the gastrointestinal system. The NRP1 protein expressed in $\approx 31\%$ (4/13) of the cell lines tested in a concentration of

between 1 and 34% of the cells, depending on the cell type, presented a significant number of key residues for binding, as well as a stable and thermodynamically favorable interaction with the S1 region, which could facilitate infection in all cell lines considered except for NTERA-2, RPTEC-TERT1, RT4, HBEC3-KT and SCLC-21H. In this sense, these predictions would translate into a susceptibility mediated by the NRP1 receptor in only 50% and 25% of the cell types studied associated with the pulmonary and urinary systems respectively.

Although HSPA5 was one of the candidates with the highest level of theoretical expression in all cell models considered (between 14 and 67% of cells) and presented one of the largest subsets of key residues for

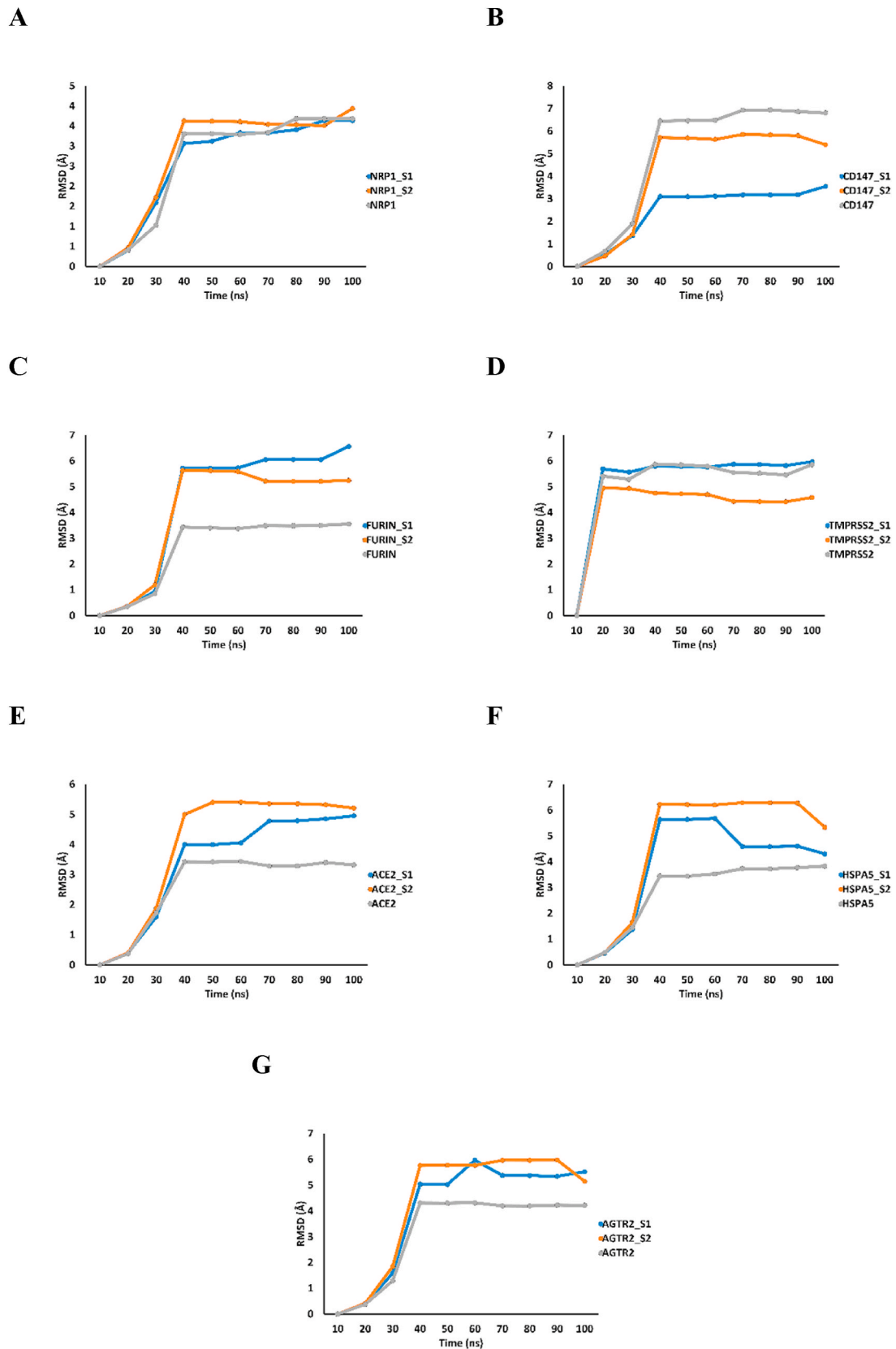


Fig. 7. The mean squared deviations (RMSD) of C α during each 10ns of MD simulation for each complex (total time 100ns). A) NRP1, B) CD147, C) FURIN, D) TMPRSS2, E) ACE2, F) HSPA5 and G) AGTR2. The gray lines represent the free receptor proteins, and the blue and orange lines the complexes with the S1 and S2 region, respectively.

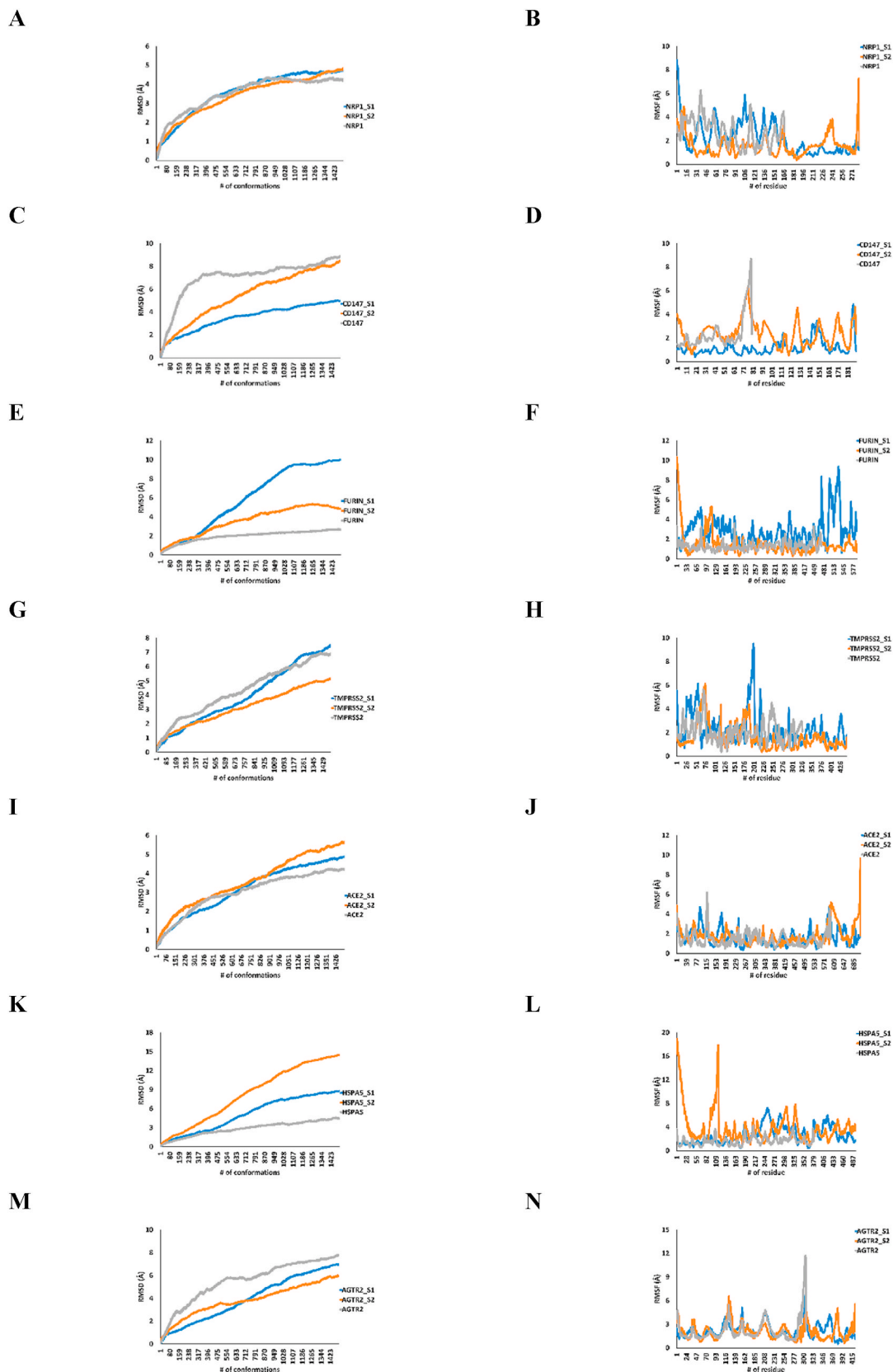


Fig. 8. The root mean square deviations (RMSD) of C α and root mean square fluctuations (RMSF) of individual amino acid residues relative to the starting frame during 100ns MD simulation of each complex using NMSim. NRP1 (A and B), CD147 (C and D), FURIN (E and F), TMPRSS2 (G and H), ACE2 (I and J), HSPAS5 (K and L), AGTR2 (M and N). The gray lines represent the free receptor proteins, and the blue and orange lines the complexes with the S1 and S2 region, respectively.

binding with a thermodynamically favorable affinity, the molecular dynamics was represented by junctions with many fluctuations in time for the S2 region, which could compromise the stability of docking complex. Nonetheless, conditions for a stable affinity with the S1 portion were predicted. Although the FURIN protease presented a low number of residues considered as keys for the interaction under the conditions of this study, which corresponded to one of the least stable predicted junctions, it presented thermodynamically favorable junctions and important levels in all cell lines (between 4 and 27% depending on the cell type). Interestingly, although the AGTR2 protein presented the best values in terms of the relative affinity energy mean, as well as a stable binding throughout the simulation time, it established complexes with both the spike regions with the least number of hot spots of relevant binding, and presented the lowest level of detectable expression reported in all cell lines considered in this study.

The results of the predicted interactions corresponded to those reported by other investigations that have shown that the S1/S2 complex of the viral spike protein adheres to the host cell membrane through the ACE2 enzyme prior to proteolytic cleavage by the transmembrane protease TMPRSS2 exposing the viral fusion peptide that allows entry of the SARS-CoV2 virus into the cell [49,51]. An affinity of ACE2 for the S1 domain and of TMPRSS2 for the S2 region were predicted in energy terms in this study. Simulations of the protein-protein interaction based on the 3D structure have also revealed that AGTR2 shows a higher binding affinity with the spike protein of SARS-CoV-2 than ACE2, as observed in this study [52]. Furthermore, experimental and clinical findings also suggest that CD147 could act as a receptor that mediates the entry of SARS-CoV-2 through its interaction with the protein S of the viral spike [42,49]. Additionally, the predicted thermodynamically favorable dockings between the spike protein and FURIN have been associated with the presence of a unique motif in SARS-CoV-2-S that is easily recognized and hydrolyzed by FURIN, thus, FURIN has been proposed as a therapeutic target [53].

In the case of the controversial protein CD147 for which its participation in the infection by SARS-CoV-2 is still debated, the interaction between the candidate receptor CD147 and the peak protein of SARS-CoV-2 has been reported by surface plasmon resonance (SPR), enzyme-linked immunosorbent assays (ELISA), co-immunoprecipitation (Co-IP) and optimized negative staining electron microscopy (OpNS-EM). It has been noted that loss or blockade of CD147 in cell lines (such as those considered in this study) can inhibit the amplification of SARS-CoV-2, suggesting it as a new route of entry of virus independent of ACE2 [54]. These observations correspond to our results in which we predicted a thermodynamically favorable interaction between spike protein and CD147. A later study reported that it could not find evidence to support the role of CD147 as a spike-binding receptor, using specialized assays designed to detect even weak interactions for the proposed CD147-spike binding using various conformations of the CD147 protein [5]. However, a recent study that used cell lines similar to those considered in this study, demonstrated that the silencing of CD147 could reduce viral entry into cells, directly or indirectly through the reduction of ACE2 expression levels and therefore affected even spike protein levels, although at the post-translational level [1]. And as with ACE2, it was demonstrated that at the RNA level, SARS-CoV-2 viral infection reduces both CD147 and ACE2, a phenomenon of negative regulation that, although not considered in this study, would further affect susceptibility cell mediated by this type of receptors as has been reported [1,45].

Similarly, the favorable energetic docking between the spike protein and the HSPA5 protein has been reported, another possible route for the adhesion and entry of SARS-CoV-2. Indeed, it has been described that the binding site in the SARS-CoV-2 spike protein is found in the C-terminal domain of the S1 region, as predicted in this study as its most favorable docking. Other authors have suggested this receptor could mediate both the infectivity of SARS and a certain degree of cross-immunity against SARS-CoV-2 as it is present in other members of the

coronavirus family [55].

Therefore, as observed in this study and as suggested by other authors, binding free energy may support that the conformation and domain type of the SARS-CoV-2 spike protein leads to a stronger or weaker binding. However, a stronger receptor binding still cannot fully explain the high infectivity of this particular type of coronavirus. Finally, it has been proposed that part of this infectivity may be mediated by the binding of the virus to other predominant receptors in other organs and that the lung may not be the earliest site of infection, as observed previously [53].

3.3. Theoretical diffusivity of the proteins of interest associated with cellular susceptibility in SARS-CoV-2 infection

The ability of the proteins considered to move in artificial cellular environments was studied simulating various types of cytoplasmic congestion, with the interest of observing the impact of macromolecular congestion on the diffusion of proteins and the theoretical relationship between their levels of expression and relative affinity as described above for the spike protein. In this sense, and after studying 5 conditions of cell-type cytoplasmic congestion and an aqueous model, a minimum translational diffusion coefficient (D_t) of $7.9 \times 10^{-27} \text{ cm}^2 \text{ s}^{-1}$ and a maximum of $6.7 \times 10^{-21} \text{ cm}^2 \text{ s}^{-1}$ was predicted for the cellular models, in contrast to a minimum D_t of $4.6 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$ and a maximum of $6.5 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$ for the aqueous model. Although all the diffusion values between the proteins studied were very close, it was predicted that under conditions of cytoplasmic congestion the NRP1 protein presented the best diffusion coefficient ($D_t = 1.9 \times 10^{-21} \text{ cm}^2 \text{ s}^{-1}$) followed by CD147 ($D_t = 1.8 \times 10^{-21} \text{ cm}^2 \text{ s}^{-1}$). Interestingly, the ACE2 protein ($D_t = 1.2 \times 10^{-21} \text{ cm}^2 \text{ s}^{-1}$) presented the lowest diffusion coefficient in the cell models tested. In cell models, a mean diffusivity time of $\approx 3 \times 10^{-10} \text{ s}$ was predicted, with a minimum time of $\approx 2 \times 10^{-10} \text{ s}$ for TMPRSS2 and a maximum time of $\approx 5 \times 10^{-10} \text{ s}$ for ACE2 (see Table 5).

In relation to the diffusion model in aqueous medium, the results showed CD147 as one of the proteins with the best diffusion coefficient ($D_t = 8.3 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$) followed by TMPRSS2 ($D_t = 7.5 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$) and FURIN ($D_t = 7.1 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$), keeping ACE2 as the protein with the lowest diffusivity ($D_t = 4.6 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$). Diffusivity in aqueous medium allowed prediction of a mean diffusion time of $\approx 13 \text{ s}$, with a minimum and maximum time of $\approx 10 \text{ s}$ and $\approx 18 \text{ s}$ respectively, diffusion rates that correspond to the CD147 and ACE2 proteins, respectively. The cell lines considered with experimental viscosity data in which the lowest and highest diffusion coefficients were predicted were Model 1 ($D_t = 7.9 \times 10^{-27} \text{ cm}^2 \text{ s}^{-1}$) and Model 4 ($D_t = 6.7 \times 10^{-21} \text{ cm}^2 \text{ s}^{-1}$) (the latter not considered in this study for the susceptibility analyzes) respectively. The HeLa cell line used in this study as a negative control for SARS-CoV-2 infectivity assays presented a better diffusion coefficient ($D_t = 2.5 \times 10^{-22} \text{ cm}^2 \text{ s}^{-1}$) than that predicted in Model 1 (line also considered in this study) (Table 5).

After making a preliminary approximation of the transmembrane flux (J) of these proteins under the various types of cytoplasmic congestion tested, similar results were predicted, with a mean of $J = 9.9 \times 10^{-16} \text{ mol cm}^2 \text{ s}^{-1}$ where the value minimum was predicted for ACE2 ($J = 8.0 \times 10^{-16} \text{ mol cm}^2 \text{ s}^{-1}$) and the maximum for NRP1 ($J = 1.3 \times 10^{-15} \text{ mol cm}^2 \text{ s}^{-1}$) followed by the predicted values for CD147 ($J = 1.2 \times 10^{-15} \text{ mol cm}^2 \text{ s}^{-1}$). The cell lines considered in which the lowest and highest transmembrane flux was predicted were therefore Model 1 ($J = 5.2 \times 10^{-21} \text{ mol cm}^2 \text{ s}^{-1}$) and Model 4 ($J = 4.5 \times 10^{-15} \text{ mol cm}^2 \text{ s}^{-1}$), respectively. The HeLa control presented a better transmembrane flux ($J = 1.6 \times 10^{-16} \text{ mol cm}^2 \text{ s}^{-1}$) than that calculated for Model 1 (see Table 6). Together with these results, a high correlation was found between the proteins studied with the lowest expression levels that presented the thermodynamically more favorable binding energy towards the viral spike protein with the lowest diffusion times and coefficients, as well as with the lowest hypothetical flux transmembrane ($r = 0.80$). Alternatively, the proteins that presented the least thermodynamically

Table 5

Comparative analysis of the theoretical translational diffusion coefficient and the diffusion time of the proteins of interest associated with cellular susceptibility in SARS-CoV-2 infection.

Protein	MW (kDa)	r (Å)	D_t (cm ² s ⁻¹)						t (seg)*
			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	
NRP1	79.20	15.33	1.0x10 ⁻²⁶	3.2x10 ⁻²²	5.8x10 ⁻²²	8.6x10 ⁻²¹	1.3x10 ⁻²⁵	6.3x10 ⁻¹¹	13.2
CD147	21.15	16.06	9.6x10 ⁻²⁷	3.0x10 ⁻²²	5.6x10 ⁻²²	8.2x10 ⁻²¹	1.2x10 ⁻²⁵	8.3x10 ⁻¹¹	10.0
FURIN	53.70	21.78	7.1x10 ⁻²⁷	2.2x10 ⁻²²	4.1x10 ⁻²²	6.0x10 ⁻²¹	8.9x10 ⁻²⁶	7.1x10 ⁻¹¹	11.7
TMPRSS2	44.50	20.28	7.6x10 ⁻²⁷	2.4x10 ⁻²²	4.4x10 ⁻²²	6.5x10 ⁻²¹	9.6x10 ⁻²⁶	7.5x10 ⁻¹¹	11.2
ACE2	192.75	24.29	6.4x10 ⁻²⁷	2.0x10 ⁻²²	3.7x10 ⁻²²	5.4x10 ⁻²¹	8.0x10 ⁻²⁶	4.6x10 ⁻¹¹	18.0
HSPA5	84.75	21.38	7.2x10 ⁻²⁷	2.3x10 ⁻²²	4.2x10 ⁻²²	6.2x10 ⁻²¹	9.1x10 ⁻²⁶	6.0x10 ⁻¹¹	13.8
AGTR2	92.57	21.94	7.0x10 ⁻²⁷	2.2x10 ⁻²²	4.1x10 ⁻²²	6.0x10 ⁻²¹	8.9x10 ⁻²⁶	5.6x10 ⁻¹¹	14.8

D_t , translational diffusion coefficient. Neuropilin 1 (NRP1), Basigin2 (CD147), Furin protease, Transmembrane Serine 2 (TMPRSS2), Angiotensin Converting Enzyme 2 (ACE2), Protein of Thermal shock A5 (HSPA5), Angiotensin II receptor type 2 (AGTR2). Model 1, A549 (~4.4x10⁻² Pa/s) [27]; Model 2, HeLa negative control (~4.4x10⁻² Pa/s) [27]; Model 3, Normal Swiss 3T3 cells (~2.4x10⁻² Pa/s) [27,28]; Model 4, ASTC-a-1 (~1.6x10⁻³ Pa/s) [29]; Model 5, H1299 (~1.1x10² Pa/s) [30]; Model 6, standard viscosity of water (~1.0x10⁻³ Pa/s) [56]. *, time calculated in model 6.

Table 6

Comparative analysis of the theoretical transmembrane flux of the proteins of interest associated with cellular susceptibility in SARS-CoV-2 infection.

Protein	J (mol m ⁻² s ⁻¹)				
	Model 1	Model 2	Model 3	Model 4	Model 5
NRP1	6.7x10 ⁻²¹	2.1x10 ⁻¹⁶	3.9x10 ⁻¹⁶	5.7x10 ⁻¹⁵	8.5x10 ⁻²⁰
CD147	6.4x10 ⁻²¹	2.0x10 ⁻¹⁶	3.7x10 ⁻¹⁶	5.5x10 ⁻¹⁵	8.1x10 ⁻²⁰
FURIN	4.7x10 ⁻²¹	1.5x10 ⁻¹⁶	2.7x10 ⁻¹⁶	4.0x10 ⁻¹⁵	6.0x10 ⁻²⁰
TMPRSS2	5.1x10 ⁻²¹	1.6x10 ⁻¹⁶	2.9x10 ⁻¹⁶	4.3x10 ⁻¹⁵	6.4x10 ⁻²⁰
ACE2	4.2x10 ⁻²¹	1.3x10 ⁻¹⁶	2.5x10 ⁻¹⁶	3.6x10 ⁻¹⁵	5.3x10 ⁻²⁰
HSPA5	4.8x10 ⁻²¹	1.5x10 ⁻¹⁶	2.8x10 ⁻¹⁶	4.1x10 ⁻¹⁵	6.1x10 ⁻²⁰
AGTR2	4.7x10 ⁻²¹	1.5x10 ⁻¹⁶	2.7x10 ⁻¹⁶	4.0x10 ⁻¹⁵	5.9x10 ⁻²⁰

J , diffusion flow. Neuropilin 1 (NRP1), Basigin2 (CD147), Protease Furin, Transmembrane Serine 2 (TMPRSS2), Angiotensin Converting Enzyme 2 (ACE2), Protein of Thermal shock A5 (HSPA5), Angiotensin II receptor type 2 (AGTR2). Model 1, A549 (~4.4x10⁻² Pa/s) [27]; Model 2, HeLa negative control (~4.4x10⁻² Pa/s) [27]; Model 3, Normal Swiss 3T3 cells (~2.4x10⁻² Pa/s) [27,28]; Model 4, ASTC-a-1 (~1.6x10⁻³ Pa/s) [29]; Model 5, H1299 (~1.1x10² Pa/s) [30].

favorable affinity and that are described with the highest expression levels had a high correlation with the best times and diffusion coefficients.

Analyses conducted to establish diffusivity parameters such as the diffusion coefficient of structures associated with SARS-CoV-2 infection have focused on isolated regions of the spike protein [57], as well as of the viral particle [58]. Interestingly, there are few reports that consider the comparison between diffusivity parameters such as diffusion coefficient and speed, as well as the transmembrane flux of candidate receptors for SARS-CoV-2 infection, despite the fact that it has been described that membrane proteins such as the receptors studied can be affected by crowding effects similar to those of the solution phase [59, 60]. Although the predicted diffusivity reports for these viral structures allow us to infer in general terms that all the simulated structures in this study in congested environments show a comparatively slow diffusion as a result of confined environments [61], it is important to note that transcription tends to be significantly improved by macromolecular crowding as a result of increased effective concentrations of enzymes and biomolecular reagents associated with transcription. In this sense, there may be a trade-off between the diffusion coefficient of the transcript and the transcription efficiency in congested environments [62, 63]. However, this phenomenon is complex, because it has also been reported that increasing the viscosity of solutions can drastically reduce the diffusion coefficients of protein-type biomolecules by factors of up to 10 times, as predicted in this study by comparing the dynamics in aqueous medium and cytoplasmic medium [62].

Therefore, it is recommended to experimentally measure the effects of macromolecular crowding on the performance of transcription and diffusivity of the transcripts described as candidate receptors for SARS-

CoV-2 infection, as suggested for other protein systems. For example, using strategies such as the construction of *de novo* designed *in vitro* protein expression systems, in which the macromolecular concentration could be varied by adding compatible solutes, or inert polymers. Alternatively, one could take a diluted cytoplasmic cell extract and study the importance of crowding under physiological conditions by adding congestive agents as has been proposed to study the effect of crowding on protein signaling. The challenge of *in vivo* assays is represented in part by controlling the intracellular crowding concentration by manipulating extracellular conditions such as osmotic pressure with compatible solutes [60,64]. Additionally, it is recommended to validate the predictions made in this study with available scRNA-seq data from healthy humans, including a greater number of cell types and using other alternative databases such as *Gene Expression Omnibus* (GEO) and the *Tissue Stability Cell Atlas* [22], as well as considering other factors or external stimuli [22,65]. It is also recommended to study other types of receptor proteins not considered in this study, such as the receptor for advanced glycation end products (RAGE), heparan sulfate, sialic acids, cathepsins, among others less cited [2], which are already being incorporated in subsequent analyzes by our group.

4. Conclusions

Our predictions based on experimental and transcriptomic data indicate differential viral kinetic behavior dependent on the degree of cellular susceptibility, which is influenced to a greater extent by the type and level of expression of potentially receptor/mediator proteins associated with SARS-CoV-2 infection, and to a lesser extent by the number of potential receptor/mediator proteins expressed. We predict under the conditions of this study a possible greater theoretical propensity to susceptibility in cell lines such as NTERA-2, SCLC-21H, HepG2 and Vero6, and a lower propensity to infectivity in lines such as CaLu3, RT4, HEK293, A549 and U-251 MG.

In addition, a relationship was evidenced between lower expression levels and less favorable diffusivity parameters (diffusion coefficient, diffusion speed and transmembrane flux) with the thermodynamically more favorable interactions observed between the studied proteins and the viral spike, and vice versa. Therefore, we can conclude that part of the explanation for cellular susceptibility to infections caused by SARS-CoV-2 could be mediated by the ability of virus to stably bind to low-expression receptors in lung, which would suggest other potential sites for the earliest infection events.

Since this comparative research used previously published experimental data to be able to design expression systems and calculations of the effect of molecular crowding, it is expected that it will stimulate quantitative analysis in future experiments and promote systematic investigation of the effect of phenomena such as crowding presented here. Especially as cancer cell lines naturally present a high degree of cytoplasmic congestion, which could affect the expression levels of the

receptors considered. Additionally, the cells would affect the diffusivity of said receptors and the infection mediated by them, due to the increase in cytoplasmic viscosity, as already mentioned. To these limitations, we need to consider other factors such as healthy cells, polymorphisms or external stimuli (ethnic groups, age, sex, pathologies, exposure to pollutants, among others) in order to incorporate more factors that could alter the expression of the receptors studied and thus impact on susceptibility to SARS-CoV-2 infections, a viral mechanism that should continue to be further explored for the optimization of therapeutic approaches.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

CRedit authorship contribution statement

Lenin González-Paz: Conceptualization, Methodology, Investigation, Writing-Reviewing and Editing. María José Alvarado: Investigation. Carla Lossada: Investigation. Maria Hurtado: Investigation, Joan Vera-Villalobos: Reviewing. JLPaz: Writing-Reviewing, Original draft preparation. Marcos Loroño: Writing-Reviewing, F. J. Torres: Writing-Reviewing, Laura N. Jeffeys: Writing-Reviewing, Original draft preparation, investigation, Ysaías J. Alvarado: Writing-Reviewing and Editing, Conceptualization, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2022.105245>.

References

- C. Fenizia, S. Galbiati, C. Vanetti, R. Vago, M. Clerici, C. Tacchetti, T. Daniele, SARS-CoV-2 entry: at the crossroads of CD147 and ACE2, *Cells* 10 (1434) (2021), <https://doi.org/10.3390/cells10061434>.
- I. Kyrou, H.S. Randeva, D.A. Spandidos, et al., Not only ACE2—the quest for additional host cell mediators of SARS-CoV-2 infection: neuropilin-1 (NRP1) as a novel SARS-CoV-2 host cell entry mediator implicated in COVID-19, *Sig Transduct Target Ther* 6 (21) (2021), <https://doi.org/10.1038/s41392-020-00460-9>.
- K. Dobrindt, D.A. Hoagland, C. Seah, B. Kassim, C.P. O’Shea, A. Murphy, K. J. Brennand, Common genetic variation in humans impacts in vitro susceptibility to SARS-CoV-2 infection, *Stem Cell Rep.* 16 (3) (2021) 505–518, <https://doi.org/10.1016/j.stemcr.2021.02.010>.
- F. Jafari, S. Jafari, M.R. Ganjalikhany, In silico investigation of critical binding pattern in SARS-CoV-2 spike protein with angiotensin-converting enzyme 2, *Sci. Rep.* 11 (6927) (2021), <https://doi.org/10.1038/s41598-021-86380-2>.
- J. Shilts, T.W. Crozier, E.J. Greenwood, P.J. Lehner, G.J. Wright, No evidence for basigin/CD147 as a direct SARS-CoV-2 spike binding receptor, *Sci. Rep.* 11 (1) (2021) 1–10, <https://doi.org/10.1038/s41598-020-80464-1>.
- F. Hikmet, L. Méar, Å. Edvinsson, P. Micke, M. Uhlén, C. Lindskog, The protein expression profile of ACE2 in human tissues, *Mol. Syst. Biol.* 16 (7) (2020), e9610, <https://doi.org/10.15252/msb.20209610>.
- E. Keyaerts, L. Vijgen, P. Maes, J. Neyts, M. Van Ranst, Growth kinetics of SARS-coronavirus in Vero E6 cells, *Biochem. Biophys. Res. Commun.* 329 (3) (2005) 1147–1151, <https://doi.org/10.1016/j.bbrc.2005.02.085>.
- Y.M. Bar-On, A. Flamholz, R. Phillips, R. Milo, Science forum: SARS-CoV-2 (COVID-19) by the numbers, *Elife* 9 (2020), e57309, <https://doi.org/10.7554/eLife.57309>.
- F. Liu, K. Han, R. Blair, K. Kenst, Z. Qin, B. Upcin, X. Qin, SARS-CoV-2 infects endothelial cells in vivo and in vitro, *Front. Cell. Infect. Microbiol.* 11 (2021), <https://doi.org/10.3389/fcimb.2021.701278>.
- N.D. Kumar, B. Ter Ellen, E.M. Bouma, B. Troost, D.P. van de Pol, H.H. van der Ende-Metselaar, J. Smit, Moxidectin and ivermectin inhibit SARS-CoV-2 replication in Vero E6 cells but not in human primary airway epithelium cells, *Antimicrob. Agents Chemother.* (2021), <https://doi.org/10.1128/AAC.01543-21>. AAC-01543.
- A. Gonçalves, J. Bertrand, R. Ke, E. Comets, X. De Lamballerie, D. Malvy, J. Guedj, Timing of antiviral treatment initiation is critical to reduce SARS-CoV-2 viral load, *CPT Pharmacometrics Syst. Pharmacol.* 9 (9) (2020) 509–514, <https://doi.org/10.1002/psp4.12543>.
- V.K. Nguyen, S.C. Binder, A. Boianelli, M. Meyer-Hermann, E.A. Hernandez-Vargas, Ebola virus infection modeling and identifiability problems, *Front. Microbiol.* 6 (2015) 257, <https://doi.org/10.3389/fmicb.2015.00257>.
- E.A. Hernandez-Vargas, Parameter estimation in mathematical models of viral infections using R, *Influenza Virus 1836* (2018) 531–549, https://doi.org/10.1007/978-1-4939-8678-1_25.
- A. Digre, C. Lindskog, The human protein atlas—spatial localization of the human proteome in health and disease, *Protein Sci.* 30 (1) (2021) 218–233, <https://doi.org/10.1002/pro.3987>.
- N.Z. Cuervo, N. Grandvaux, ACE2: evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities, *Elife* 9 (2020), e61390, <https://doi.org/10.7554/eLife.61390>.
- A. Bairoch, The cellosaurus, a cell-line knowledge resource, *J. Biomol. Tech.: J. Biochem. (Tokyo)* 29 (2) (2018) 25–38, <https://doi.org/10.7171/jbt.18-2902-002>.
- A. Peixoto, M. Monteiro, B. Rocha, H. Veiga-Fernandes, Quantification of multiple gene expression in individual cells, *Genome Res.* 14 (10a) (2004) 1938–1947, <https://doi.org/10.1101/gr.2890204>.
- Y. Lecarpentier, A. Vallée, The key role of the level of ACE2 gene expression in SARS-CoV-2 infection, *Aging* 13 (11) (2021) 14552–14556, <https://doi.org/10.18632/aging.203181>.
- S. Matsuyama, M. Kawase, N. Nao, K. Shirato, M. Ujiie, W. Kamitani, S. Fukushi, The inhaled steroid ciclesonide blocks SARS-CoV-2 RNA replication by targeting the viral replication-transcription complex in cultured cells, *J. Virol.* 95 (1) (2020), <https://doi.org/10.1128/JVI.01648-20> e01648-20.
- J. Modrof, A. Kerschbaum, M.R. Farcet, D. Niemeyer, V.M. Corman, T.R. Kreil, SARS-CoV-2 and the safety margins of cell-based biological medicinal products, *Biologicals* 68 (2020) 122–124, <https://doi.org/10.1016/j.biologicals.2020.08.010>.
- B.Z. Zhang, H. Chu, S. Han, H. Shuai, J. Deng, Y.F. Hu, J.D. Huang, SARS-CoV-2 infects human neural progenitor cells and brain organoids, *Cell Res.* 30 (10) (2020) 928–931, <https://doi.org/10.1038/s41422-020-0390-x>.
- J. Qi, Y. Zhou, J. Hua, L. Zhang, J. Bian, B. Liu, S. Jin, The scRNA-seq expression profiling of the receptor ACE2 and the cellular protease TMPRSS2 reveals human organs susceptible to SARS-CoV-2 infection, *Int. J. Environ. Res. Publ. Health* 18 (1) (2021) 284, <https://doi.org/10.3390/ijerph18010284>.
- L. González-Paz, M.L. Hurtado-León, C. Lossada, F.V. Fernández-Materán, J. Vera-Villalobos, M. Loroño, Y.J. Alvarado, Comparative study of the interaction of ivermectin with proteins of interest associated with SARS-CoV-2: a computational and biophysical approach, *Biophys. Chem.* 278 (2021) 106677, <https://doi.org/10.1016/j.bpc.2021.106677>.
- S. Malhotra, O. Mathew, R. Sowdhagini, DOCKSCORE: a webserver for ranking protein-protein docked poses, *BMC Bioinf.* 16 (127) (2015), <https://doi.org/10.1186/s12859-015-0572-6>.
- X. Zhu, J.C. Mitchell, KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features, *Proteins: Struct. Funct. Bioinform.* 79 (9) (2011) 2671–2683, <https://doi.org/10.1002/prot.23094>.
- K. Kasahara, H. Terazawa, H. Itaya, S. Goto, H. Nakamura, T. Takahashi, J. Higo, myPresto/omegagene 2020: a molecular dynamics simulation engine for virtual-system coupled sampling, *Biophys. Physicobiol.* 17 (2020) 140–146, <https://doi.org/10.2142/biophysico.BSJ-2020013>.
- T. Kalwarczyk, N. Ziebac, A. Bielejewska, E. Zaboklicka, K. Koynov, J. Szymanski, R. Holyst, Comparative analysis of viscosity of complex liquids and cytoplasm of mammalian cells at the nanoscale, *Nano Lett.* 11 (5) (2011) 2157–2163, <https://doi.org/10.1021/nl2008218>.
- K. Regan, R. Dotterweich, S. Ricketts, R.M. Robertson-Anderson, Diffusion and conformational dynamics of single DNA molecules crowded by cytoskeletal proteins, *J. Undergrad. Rep. Phys.* 28 (1) (2018) 100005, <https://doi.org/10.1063/1.5109559>.
- L. Liang, X. Wang, X. Da, T. Chen, W.R. Chen, Noninvasive determination of cell nucleoplasmic viscosity by fluorescence correlation spectroscopy, *J. Biomed. Opt.* 14 (2) (2009), 024013, <https://doi.org/10.1117/1.3088141>.
- K. Wang, X.H. Sun, Y. Zhang, T. Zhang, Y. Zheng, Y.C. Wei, J. Chen, Characterization of cytoplasmic viscosity of hundreds of single tumour cells based on micropipette aspiration, *R. Soc. Open Sci.* 6 (3) (2019) 181707, <https://doi.org/10.1098/rsos.181707>.
- N.C. Garbett, C.S. Mekmaysy, J.B. Chaires, Sedimentation Velocity Ultracentrifugation Analysis for Hydrodynamic Characterization of G-Quadruplex Structures. G-Quadruplex DNA, Humana Press, 2010, pp. 97–120, https://doi.org/10.1007/978-1-59745-363-9_7.
- R. Phillips, J. Kondev, J. Theriot, *Physical Biology of the Cell*, Garland Science, 2009, 9780815344506.
- P.E. Schavemaker, A.J. Boersma, B. Poolman, How important is protein diffusion in prokaryotes? *Front. Mol. Biosci.* 5 (2018) 93, <https://doi.org/10.3389/fmolb.2018.00093>.
- J. Brand, K. Voigt, B. Zochowski, U. Kulozik, Lysozyme fractionation from egg white at pilot scale by means of tangential flow membrane adsorbers: investigation of the flow conditions, *J. Chromatogr. A* 1438 (2016) 143–149, <https://doi.org/10.1016/j.chroma.2016.02.023>.
- Z. Xu, L. Gao, P. Chen, L.T. Yan, Diffusive transport of nanoscale objects through cell membranes: a computational perspective, *Soft Matter* 16 (16) (2020) 3869–3881, <https://doi.org/10.1039/C9SM02338K>.

- [36] A.N. Leonard, E. Wang, V. Monje-Galvan, J.B. Klauda, Developing and testing of lipid force fields with applications to modeling cellular membranes, *Chem. Rev.* 119 (9) (2019) 6227–6269, <https://doi.org/10.1021/acs.chemrev.8b00384>.
- [37] J. Song, Y. Li, X. Huang, Z. Chen, Y. Li, C. Liu, X. Duan, Systematic analysis of ACE2 and TMPRSS2 expression in salivary glands reveals underlying transmission mechanism caused by SARS-CoV-2, *J. Med. Virol.* 92 (11) (2020) 2556–2566, <https://doi.org/10.1002/jmv.26045>.
- [38] P. Bradding, M. Richardson, T.S. Hinks, P.H. Howarth, D.F. Choy, J.R. Arron, S. Siddiqui, ACE2, TMPRSS2, and furin gene expression in the airways of people with asthma—implications for COVID-19, *J. Allergy Clin. Immunol.* 146 (1) (2020) 208–211, <https://doi.org/10.1016/j.jaci.2020.05.013>.
- [39] J. Shang, Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach, F. Li, Cell entry mechanisms of SARS-CoV-2, *Proc. Natl. Acad. Sci. Unit. States Am.* 117 (21) (2020) 11727–11734, <https://doi.org/10.1073/pnas.2003138117>.
- [40] D. Zipeto, J.D.F. Palmeira, G.A. Argañaraz, E.R. Argañaraz, ACE2/ADAM17/TMPRSS2 interplay may be the main risk factor for COVID-19, *Front. Immunol.* 11 (2020) 2642, <https://doi.org/10.3389/fimmu.2020.576745>.
- [41] C.B. Jackson, M. Farzan, B. Chen, H. Choe, Mechanisms of SARS-CoV-2 entry into cells, *Nat. Rev. Mol. Cell Biol.* (2021) 1–18, <https://doi.org/10.1038/s41580-021-00418-x>.
- [42] J. Qiao, W. Li, J. Bao, Q. Peng, D. Wen, J. Wang, B. Sun, The expression of SARS-CoV-2 receptor ACE2 and CD147, and protease TMPRSS2 in human and mouse brain cells and mouse brain tissues, *Biochem. Biophys. Res. Commun.* 533 (4) (2020) 867–871, <https://doi.org/10.1016/j.bbrc.2020.09.042>.
- [43] A. Leonardi, U. Rosani, P. Brun, Ocular surface expression of SARS-CoV-2 receptors, *Ocul. Immunol. Inflamm.* 28 (5) (2020) 735–738, <https://doi.org/10.1080/09273948.2020.1772314>.
- [44] C. Liu, K. Wang, M. Zhang, X. Hu, T. Hu, Y. Liu, J. Yue, High expression of ACE2 and TMPRSS2 and clinical characteristics of COVID-19 in colorectal cancer patients, *NPJ Precision Oncol.* 5 (1) (2021) 1–7, <https://doi.org/10.1038/s41698-020-00139-y>.
- [45] S. Nersisyan, M. Shkurnikov, A. Turchinovich, E. Knyazev, A. Tonevitsky, Integrative analysis of miRNA and mRNA sequencing data reveals potential regulatory mechanisms of ACE2 and TMPRSS2, *PLoS One* 15 (7) (2020), e0235987, <https://doi.org/10.1371/journal.pone.0235987>.
- [46] M. Matusiak, C.M. Schürch, Expression of SARS-CoV-2 entry receptors in the respiratory tract of healthy individuals, smokers and asthmatics, *Respir. Res.* 21 (1) (2020) 1–6, <https://doi.org/10.1186/s12931-020-01521-x>.
- [47] L. Lin, X. Jiang, Z. Zhang, S. Huang, Z. Zhang, Z. Fang, H. Shan, Gastrointestinal symptoms of 95 cases with SARS-CoV-2 infection, *Gut* 69 (6) (2020) 997–1001, <https://doi.org/10.1136/gutjnl-2020-321013>.
- [48] I. Sulaiman, M. Chung, L. Angel, S. Koralov, B. Wu, S. Yeung, L. Segal, Microbial signatures in the lower airways of mechanically ventilated COVID19 patients associated with poor clinical outcome. <https://doi.org/10.21203/rs.3.rs-266050/v1>, 2021.
- [49] F. Chiappelli, CoVID-19 susceptibility, *Bioinformatics* 16 (7) (2020) 501–504, <https://doi.org/10.6026/97320630016501>.
- [50] F. Piva, B. Sabanovic, M. Cecati, M. Giuliotti, Expression and co-expression analyses of TMPRSS2, a key element in COVID-19, *Eur. J. Clin. Microbiol. Infect.* Dis. 40 (2) (2021) 451–455, <https://doi.org/10.1007/s10096-020-04089-y>.
- [51] M. Hussain, N. Jabeen, A. Amanullah, A.A. Baig, B. Aziz, S. Shabbir, N. Uddin, Molecular docking between human TMPRSS2 and SARS-CoV-2 spike protein: conformation and intermolecular interactions, *AIMS Microbiol.* 6 (3) (2020) 350–360, <https://doi.org/10.3934/microbiol.2020021>.
- [52] C. Cui, C. Huang, W. Zhou, X. Ji, F. Zhang, L. Wang, Cui, Q. Agr2, One possible novel key gene for the entry of 2019-nCoV into human cells, *Preprints* (2020) 2020020194, <https://doi.org/10.20944/preprints202002.0194.v1>.
- [53] C. Wu, M. Zheng, Y. Yang, X. Gu, K. Yang, M. Li, H. Li, Furin: a potential therapeutic target for COVID-19, *iScience* 23 (10) (2020) 101642, <https://doi.org/10.1016/j.isci.2020.101642>, 2020.
- [54] K. Wang, W. Chen, Z. Zhang, Y. Deng, J.Q. Lian, P. Du, Z.N. Chen, CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells, *Signal Transd. Targeted Ther.* 5 (1) (2020) 1–10, <https://doi.org/10.1038/s41392-020-00426-x>.
- [55] A.A. Elfiky, SARS-CoV-2 spike-heat shock protein A5 (GRP78) recognition may be related to the immersed human coronaviruses, *Front. Pharmacol.* 11 (2020), <https://doi.org/10.3389/fphar.2020.577467>, 1997.
- [56] A. Das, V. Vishvakarma, A. Dey, S. Dey, A. Gupta, M. Das, S. Maiti, Biophysical properties of the isolated spike protein binding helix of human ACE2, *Biophys. J.* 120 (14) (2021) 2785–2792, <https://doi.org/10.1016/j.bpj.2021.06.017>.
- [57] Z. Liu, M. Xia, Z. Chai, D. Wang, Tracing the driving forces responsible for the remarkable infectivity of 2019-nCoV: 1. Receptor binding domain in its bound and unbound states, *Phys. Chem. Chem. Phys.* 22 (48) (2020) 28277–28285, <https://doi.org/10.1039/D0CP04435K>.
- [58] Z. Shang, S.Y. Chan, W.J. Liu, P. Li, W. Huang, Recent insights into emerging coronavirus: SARS-CoV-2, *ACS Infect. Dis.* 7 (6) (2021) 1369–1388, <https://doi.org/10.1021/acsinfectdis.0c00646>.
- [59] H.X. Zhou, Crowding effects of membrane proteins, *J. Phys. Chem. B* 113 (23) (2009) 7995–8005, <https://doi.org/10.1021/jp8107446>.
- [60] M. Schlame, Protein crowding in the inner mitochondrial membrane, *Biochim. Biophys. Acta Bioenerg.* 1862 (1) (2021) 148305, <https://doi.org/10.1016/j.bbabi.2020.148305>.
- [61] K. Gorshkov, K. Susumu, J. Chen, M. Xu, M. Pradhan, W. Zhu, E. Oh, Quantum dot-conjugated SARS-CoV-2 spike pseudo-virions enable tracking of angiotensin converting enzyme 2 binding and endocytosis, *ACS Nano* 14 (9) (2020) 12234–12247, <https://doi.org/10.1021/acsnano.0c05975>.
- [62] X. Ge, D. Luo, J. Xu, Cell-free protein expression under macromolecular crowding conditions, *PLoS One* 6 (12) (2011), e28707, <https://doi.org/10.1371/journal.pone.0028707>.
- [63] M.J. Morelli, R.J. Allen, P.R. Ten Wolde, Effects of macromolecular crowding on genetic networks, *Biophys. J.* 101 (12) (2011) 2882–2891, <https://doi.org/10.1016/j.bpj.2011.10.053>.
- [64] N.F. Dupuis, E.D. Holmstrom, D.J. Nesbitt, Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics, *Proc. Natl. Acad. Sci. Unit. States Am.* 111 (23) (2014) 8464–8469, <https://doi.org/10.1073/pnas.1316039111>.
- [65] J. Chen, Q. Jiang, X. Xia, K. Liu, Z. Yu, W. Tao, J.D.J. Han, Individual variation of the SARS-CoV-2 receptor ACE2 gene expression and regulation, *Aging Cell* 19 (7) (2020), e13168, <https://doi.org/10.1111/acel.13168>.