



## Accepted Article

**Title:** Binding of SARS-CoV-2 to cell receptors: a tale of molecular evolution

**Authors:** Santiago A. Gomez, Natalia Rojas-Valencia, Sara Gomez, Franco Egidi, Chiara Cappelli, and Albeiro Restrepo

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

**To be cited as:** *ChemBioChem* 10.1002/cbic.202000618

**Link to VoR:** <https://doi.org/10.1002/cbic.202000618>

# Binding of SARS–CoV–2 to cell receptors: a tale of molecular evolution

Santiago A. Gómez,<sup>†</sup> Natalia Rojas–Valencia,<sup>†,‡</sup> Sara Gómez,<sup>¶</sup> Franco Egidi,<sup>¶</sup>  
Chiara Cappelli,<sup>¶</sup> and Albeiro Restrepo<sup>\*,†</sup>

<sup>†</sup>*Instituto de Química, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín,  
Colombia*

<sup>‡</sup>*Escuela de Ciencias y Humanidades, Departamento de Ciencias Básicas, Universidad  
Eafit, AA 3300, Medellín, Colombia*

<sup>¶</sup>*Scuola Normale Superiore, Classe di Scienze, Piazza dei Cavalieri 7, 56126, Pisa, Italy*

E-mail: [albeiro.restrepo@udea.edu.co](mailto:albeiro.restrepo@udea.edu.co)

Accepted Manuscript

## Abstract

The magnified infectious power of the SARS-CoV-2 virus compared to its precursor SARS-CoV is intimately linked to an enhanced ability in the mutated virus to find available hydrogen bond sites in the host cells. This characteristic is acquired during virus evolution because of the selective pressure exerted at the molecular level. We pinpoint the specific residue (in the virus) to residue (in the cell) contacts during the initial recognition and binding and show that the virus...cell interaction is mainly due to an extensive network of hydrogen bonds and to a large surface of non-covalent interactions. In addition to the formal quantum characterization of bonding interactions, computation of absorption spectra for the specific virus...cell interacting residues yields significant shifts of  $\Delta\lambda_{max} = 47$  and  $66$  nm in the wavelength for maximum absorption in the complex with respect to the isolated host and virus, respectively.

# 1 Introduction

At the time of the writing of this manuscript, the situation regarding the global pandemic produced by the spread of the SARS-CoV-2 virus (over 17.7 million confirmed cases, over 675000 deaths, with no end in sight), with dire consequences in all aspects of life, from social interactions, to the overwhelming of health and economic systems, is changing fast. Because this is a critical problem, just as the rate of virus transmission on the early stages of dissemination, the number of scientific papers on the subject (mostly preprints) increases exponentially.

SARS-CoV-2 is an enveloped virus of the *Coronaviridae* family with a single-stranded RNA genome.<sup>[1]</sup> Figure 1 highlights the most important structural features of the virus: besides the nucleocapsid (N) proteins, the only proteins in direct contact with the genetic material (the N-RNA core is embedded in a lipid environment), there are membrane (M), envelope (E), and spike (S) proteins. It is the spike proteins which lead to the now familiar external morphology of the virus, but more importantly, S proteins are responsible for the interactions with receptors in the host membrane (epithelial cells in humans). These S···Receptors contacts initiate the infectious cycle of the virus.<sup>[2]</sup>

Each spike consists of a trimer of S proteins. Individual S proteins have been divided into two clear S1, S2 sections,<sup>[3]</sup> with S1 containing the N-terminal domain (NTD), and the receptor binding domain (RBD), the domain ultimately responsible for the interactions with the coupling factors present in cell membranes.<sup>[4]</sup> Coupling factors include a variety of proteins, carbohydrates, or other types of biomolecules expressed on the surface of the cell membrane and in charge of signaling and transport, among other functions. Viruses take advantage of these molecules during the infection process. It seems well established that initial virus↔host recognition and binding is driven by S1, and that further changes in the conformation of the S2 section mediate the viral envelope fusion to the host cell membrane.

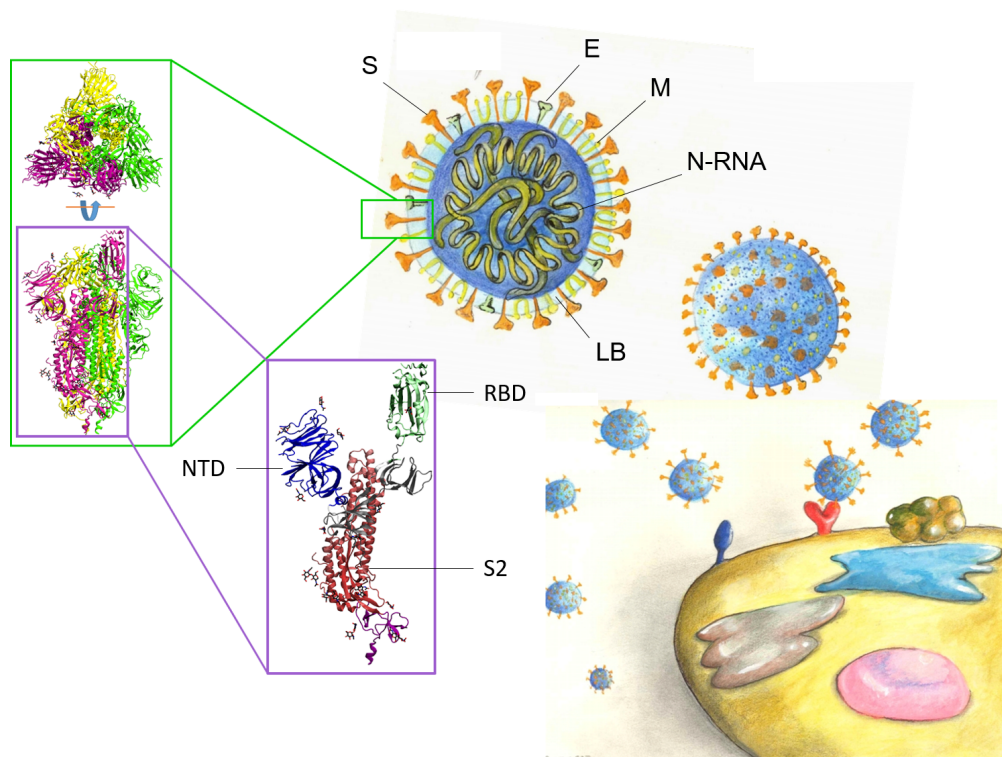


Figure 1: **Structural features of the SARS-CoV-2 virus and of the spike (S) protein.** The envelope (E) and membrane (M) proteins as well as the lipid bilayer (LB), and nucleocapsid-RNA core (N-RNA) are highlighted in the virus. The receptor binding domain (RBD), the N-terminal domain (NTD), and section 2 (S2) are highlighted in one protomer extracted from the trimer constituting the S protein. We also show a cell with internal organelles and with a few enzymes that act as a virus receptors.

The most commonly invoked culprit (with plenty of experimental evidence<sup>[5]</sup>) for the reception of SARS-CoV-2 is the Angiotensin Converting Enzyme 2 (ACE2). This receptor is the subject of intensive studies aiming at finding effective therapies, and is the central focus of the ongoing race to find a vaccine. In this work, we are interested in two crucial aspects of the initial virus···cell interaction problem: to pinpoint the specific residue to residue binding sites between the structurally known spike proteins of the virus<sup>[6]</sup> and the structurally known ACE2 receptor in cell membranes,<sup>[5]</sup> and to understand, from a fundamental, quantum perspective, the molecular factors driving the virus···cell binding. We expect this knowledge to considerably better our understanding of the problem and to hopefully contribute to a rational design of drugs and vaccines to fight the virus.

## 2 Results and discussion

See the Computational Methods section for details of our calculations. Our data shows that the RBD(S)···ACE2 complex reached well defined persistent equilibrium states long before the 600 ns of the molecular dynamics (MD) simulation time are consumed in each of the three replicas. This stability is especially encouraging in the interaction region as clearly shown in the highlighted areas of the bottom panels in Figure 2. We obtained an interaction energy  $\Delta G_{int} = -419.91$  kcal/mol, which is in excellent agreement with calculations reported in closely related systems.<sup>[7]</sup>

### 2.1 Virus···cell contacts

In all cases, only hydrogen bonds (HBs) were found as responsible for explicit virus···cell pair-wise interactions. Naturally, this does not mean that other weak, long range cumulative interactions are ruled out. The number of hydrogen bonds (Figure 3) fluctuates around 5 during the intermediate to late stages of the simulation of the RBD(S)···ACE2 complex. Fittingly, as is characteristic of equilibrated systems, the number of HBs is quite stabilized

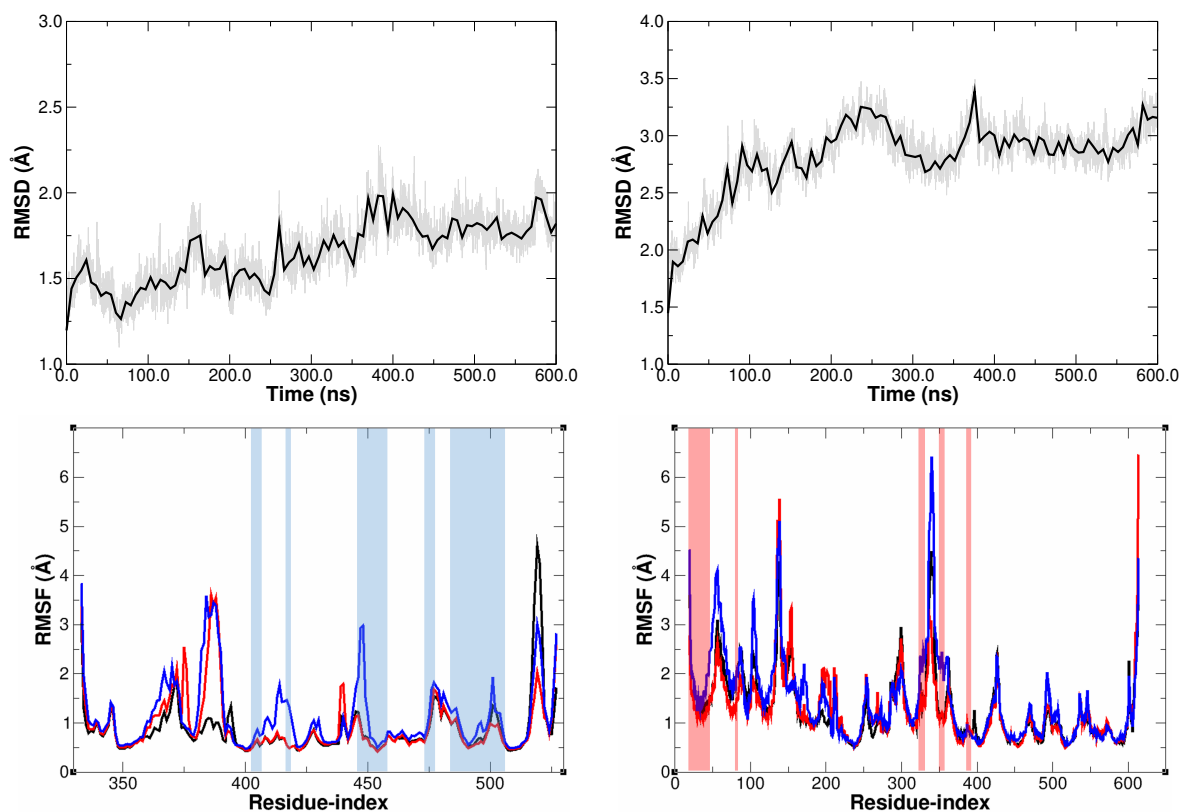


Figure 2: **Equilibration of the Molecular Dynamics simulations.** Structural fluctuations of the receptor binding domain of the spike protein (left) and of the ACE2 receptor (right) in the RBD(S)···ACE2 complex. Root mean square deviations of atomic positions (RMSD) plots, averaged over three independent MD replicas are shown at the top. Plots of the root mean square fluctuations (RMSF) of the residues are shown at the bottom. The RBD(S)···ACE2 interaction region is highlighted.

for the later steps of the simulations. Table 1 lists all individual binary contacts in the form of hydrogen bonds between residues in the RBD(S)···Receptor complex found in our MD simulations with an arbitrary threshold average of 15% occupancy on the triplicate runs. Notice that this procedure intends to extract a representative sample from the simulations, thus, there are considerably more contacts not explicitly shown because they have lower occupancies or, because while having high occupancies, are not persistent in the 3 replicas. These HB contacts, whose bonding interactions are dissected below, are responsible for the attachment of the virus to epithelial cells in humans, initiating the infection process.

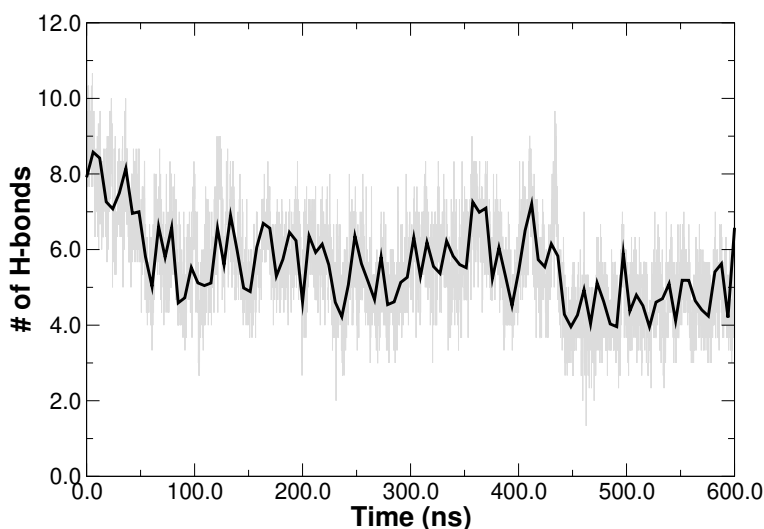


Figure 3: **Hydrogen bonding during the MD simulations.** Number of fragment to fragment hydrogen bonds in the RBD(S)···ACE2 complex averaged over three independent replicas as the MD progresses.

## 2.2 Quantum interactions

A summary of the quantum descriptors for the virus···cell interactions is listed in Table 1, the corresponding pictures are shown in Figures 4, 5. Without exception, despite being weak organic acids, residue to residue hydrogen bonds are stronger than the archetypal HB in the water dimer, this is seen in the larger binding energies, smaller distances, orbital interaction energies,  $E_{d \rightarrow a}^{(2)}$  of comparable magnitudes, larger bond indices, and in the properties of the bond critical points. The LYS417→ASP30 is an exceptional case because it corresponds to



Table 1: **Properties of virus...cell hydrogen bonds.** Persistent hydrogen bonds in the RBD(S)...Receptor complex exceeding the 15% occupancy threshold during the entire 600 ns MD simulation of each replica. The arrows state the directionality of the donor  $\rightarrow$  acceptor interaction in the corresponding hydrogen bond according to the classical electrostatic  $X^{\delta-} - H^{\delta+} \rightarrow Y^{\delta-}$  description. All occupancies averaged over the three MD replicas. WBI are the Wiberg bond indices.<sup>[8]</sup> The archetypal hydrogen bond in the water dimer is included for comparison purposes. See the specialized literature<sup>[9-14]</sup> for the formalism on how NBO and QTAIM descriptors are related to bonding.

RBD(S)	ACE2	% Occ	$d$ Å	$-E_{d \rightarrow a}^{(2)}$ kcal/mol	WBI	$10^2 \rho(\mathbf{r}_c)$ a. u.	$\frac{ \mathcal{V}(\mathbf{r}_c) }{\mathcal{G}(\mathbf{r}_c)}$	$10^2 \frac{\mathcal{H}(\mathbf{r}_c)}{\rho(\mathbf{r}_c)}$ a. u.
GLY502	$\rightarrow$ LYS353	38.71	1.94	6.06	0.04	2.51	0.92	6.8
LYS417	$\rightarrow$ ASP30	35.88	1.55	23.20	0.13	6.46	1.16	-14.0
ASN487	$\leftarrow$ TYR83	34.65	2.11	3.28	0.02	1.54	0.91	7.4
THR500	$\rightarrow$ ASP355	30.32	1.87	8.08	0.04	2.85	0.90	8.0
GLN493	$\rightarrow$ GLU35	28.44	2.21	3.68	0.02	1.46	0.98	1.1
TYR505	$\rightarrow$ GLU37	16.92	1.90	8.30	0.04	2.72	0.91	6.7
H-O-H	$\rightarrow$ OH <sub>2</sub>		1.98	7.09	0.01	2.30	0.89	10.0

an unusually strong, highly ionic  $[H_2N - H]^+ \cdots^- [O_2C]$  interaction, characterized as a salt bridge in a previous work.<sup>[15]</sup>

All hydrogen bonds are well characterized long range interactions. This is clearly seen in (i) There is never a formal  $\sigma$  orbital between the fragments, on the contrary, orbital interactions are always of the  $n_O \rightarrow \sigma_{N-H}^*$  or  $n_O \rightarrow \sigma_{O-H}^*$  form. In other words, in the NBO picture (Figure 4), all explicit virus...cell contacts are stabilized by charge transfer from one lone pair in an oxygen atom in the donor residue to an antibonding orbital in the acceptor residue. (ii) All properties of the bond critical points support the same picture: small  $\rho(\mathbf{r}_c)$ , small bond orders, virial ratios smaller than 1, and positive bond degree parameters. Again, the  $n_O \rightarrow \sigma_{N-H}^*$  in LYS417 $\rightarrow$ ASP30 is the exception, with all the calculated descriptors indicating a highly ionic contact. The non covalent interactions (NCI) calculation for the interaction region uncovers a large discontinuous non-covalent wall separating RBD(S) from the receptor. Therefore, we characterize the virus...cell binding as due to a large number of non-covalent contacts between the two proteins, enhanced by the water molecules, acting in conjunction with the specific residue to residue hydrogen bonds.

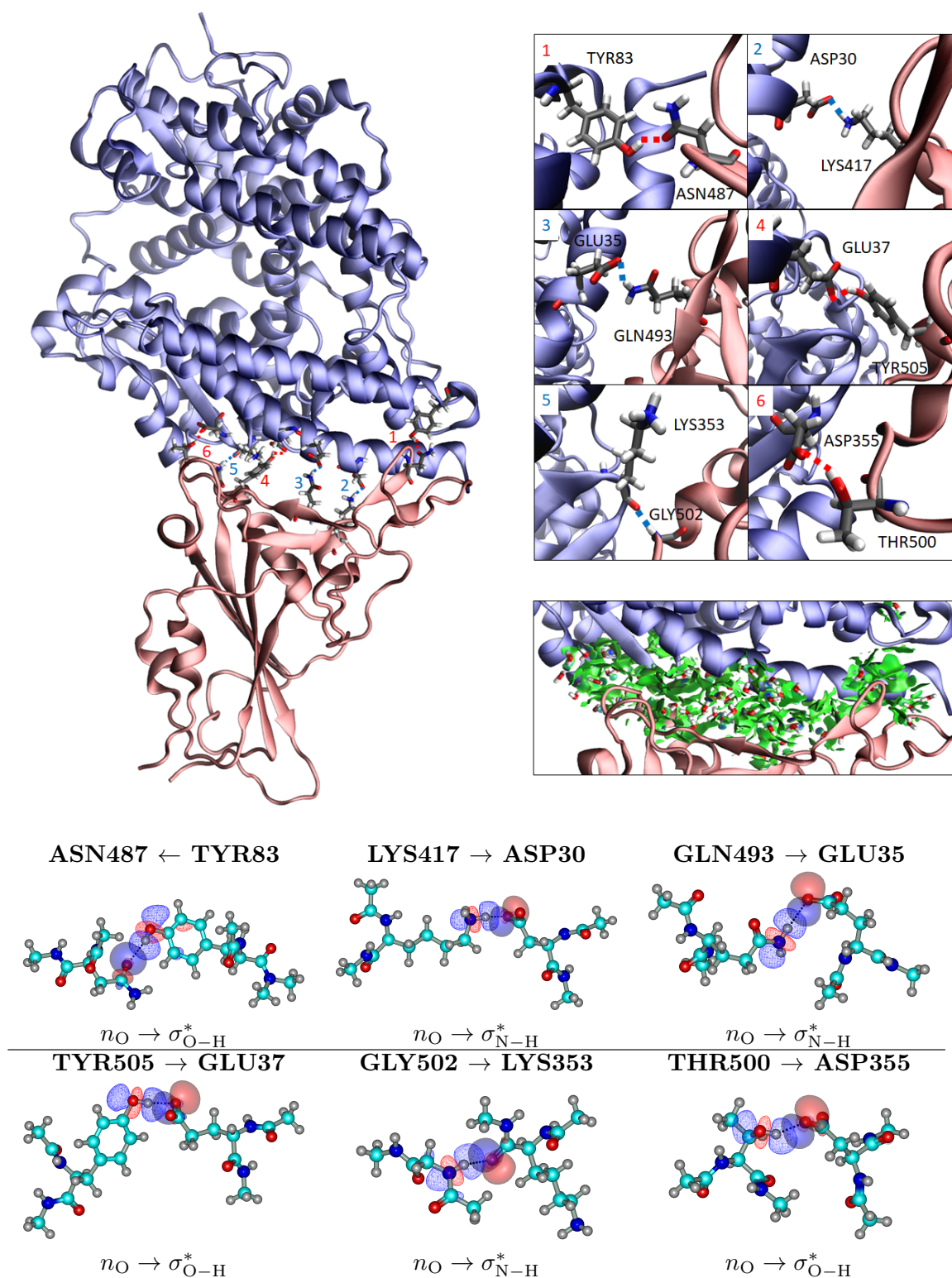


Figure 4: **Explicit virus···cell interactions.** Top: interaction between the SARS-CoV-2 RBD(S) and the ACE2 receptor (in blue). The snapshot was randomly extracted from the late stages of one of the three MD replicas.<sup>[16]</sup> All persistent virus···cell hydrogen bonds listed in Table 1 are explicitly highlighted on the right frames. The non covalent<sup>[17,18]</sup> virus···cell interaction surface is explicitly shown in green, including the water molecules. Notice that both fragments contain glycosylated glycoproteins and that all fine structure is accounted for during the calculations, however, the glycans are not shown in this picture for clarity. Bottom: NBO donor→acceptor interactions responsible for the persistent hydrogen bonds.

## 2.3 Electronic absorption spectroscopy

We concentrate on simulating the spectra of the aminoacids involved to investigate whether measurable changes in their spectral response occur upon binding, while residues that are not involved in the interaction can safely be viewed as changing very little as the two structures connect. For this purpose we propose a QM/MM approach where the aminoacid pairs involved in the binding are treated quantum mechanically by means of Density Functional Theory (DFT), while the rest of the protein environment is modelled classically, through the use of the AMBER force field.<sup>[19]</sup> In this way, the electronic structure of the QM portion is influenced by its environment by means of an Electrostatic Embedding paradigm,<sup>[20]</sup> where fixed charges are assigned to the MM atoms and directly affect the QM density and computed electronic excitations. Figure 5 shows that in all six cases, binding events set off drastic changes in the spectral response of the system, explicitly seen in red-shifted absorptions. The red-shift of the absorption bands is clearly visible in the convoluted spectrum and therefore provides unequivocal evidence of virus...cell bonding. The results listed here are quite encouraging and constitute an initial step that will hopefully motivate the design of experimental protocols to detect virus infection. However, it is clear that a number of details need to be worked out before practical applications can be devised. In particular, the potential interference of signals arising from functional groups in the same region of  $\lambda_{max}$ , and the ability of the dimer model to accurately mimic physiological environments, should be addressed.

## 3 Virus mutation as a molecular evolution problem

The basics of molecular and viral evolution are well established and are worth summarizing in the context of this manuscript. Both processes are number games, that is, although the probabilities of individual specific mutations occurring are quite small (rates of mutation for single nucleotide sites in viruses are typically  $10^{-6}$ , literally one in a million), they oc-

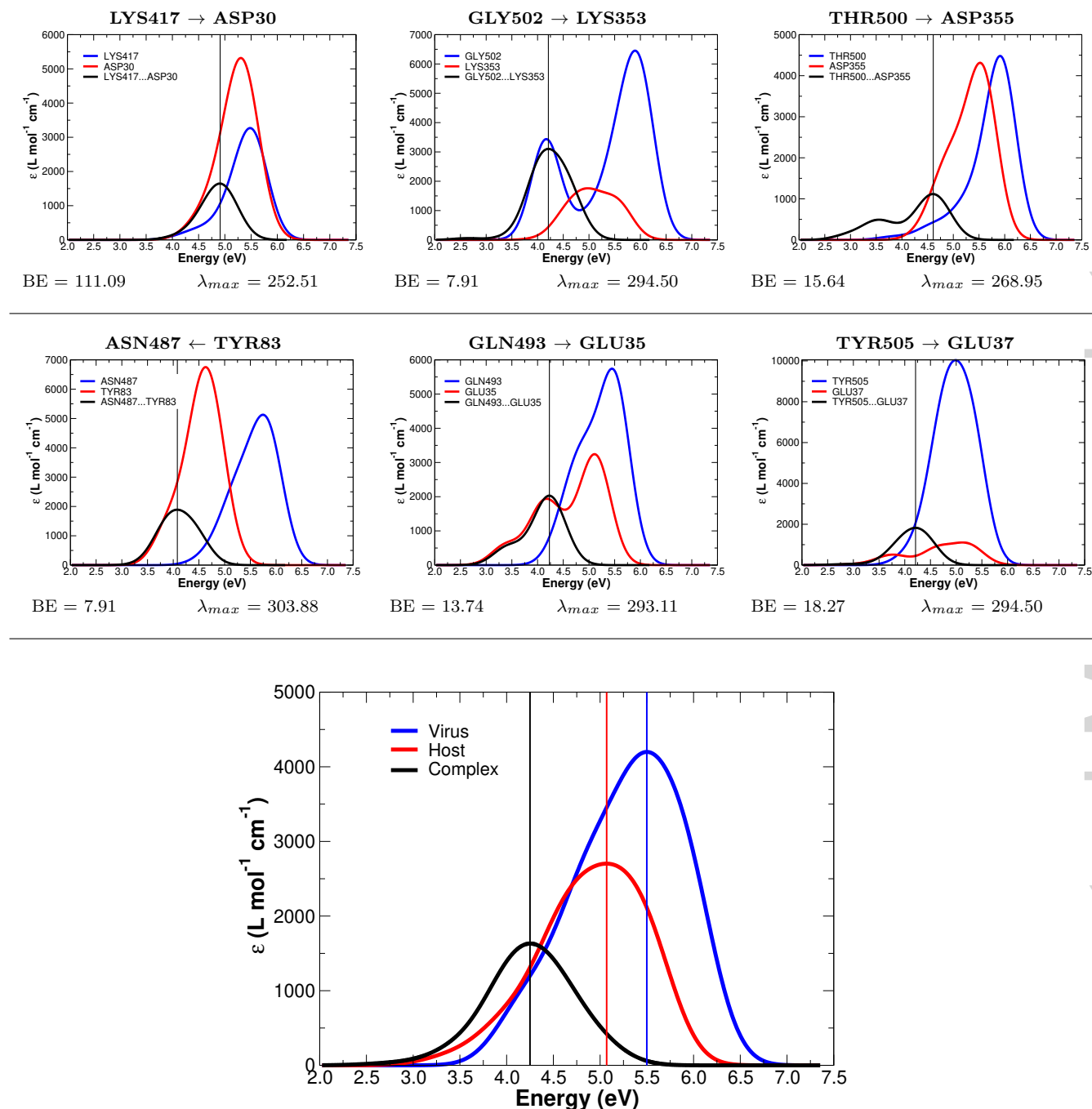


Figure 5: **Calculated QM/MM absorption spectra of the dimers and of the virus·cell complex.** Calculated QM/MM absorption spectra for the monomers (blue spectra refer to aminoacids belonging to the virus, red spectra refer to aminoacids belonging to the receptor) and for the bound structures (in black).  $\lambda_{max}$ , the wavelength (nm) of maximum absorption for the bound structure is explicitly included. All binding energies in kcal/mol. The convoluted absorption spectrum for the complex at the bottom shows  $\Delta\lambda_{max} = 47$  and  $66$  nm shifts from the isolated host and virus, respectively. Level: B3LYP/aug-cc-pVDZ/AMBER.

cur nonetheless because of the random nature of aminoacid sequencing errors proper of the replication of genetic material and because of the large number of reproduction events. In the absence of artificial or sexual selection, out of the large number of mutations that do occur, some are neutral, some are harmful, and some are beneficial to the evolving entity as a function of the environment demands. Due to the evolution pressure imposed by the environment, those individuals with beneficial mutations, which start with very small populations, have a larger probability of surviving their generation and of producing descendants into the next generation after yet another round of replication, thus progressively increasing their population until eventually, after several generations, the traits brought about by those changes become the dominant characteristic. There are usually diverse ways to survive the evolution pressure, leading to speciation. Because of the very fast reproduction rates, environment driven virus mutations help them quickly develop resistance against external agents designed to fight their infections. This ability of viruses to quickly mutate is among the most serious problems faced when developing vaccines and therapies, and is particularly true for SARS-CoV-2.<sup>[21]</sup>

We argue that this view of evolution as driven by environment induced molecular responses at the virus/biomolecules scales helps explain many aspects of evolution that are difficult to rationalize otherwise, namely, *(i)* in most cases evolution is a highly localized process *(ii)* because of the large number of mutation possibilities, which occur no matter how small the individual probabilities, an increase in entropy of the universe is the ultimate factor driving evolution *(iii)* evolution is a deterministic process driven by cumulative random changes *(iv)* in that sense, life itself is a deterministic process that only requires a large increase in the entropy of the universe (in other words, a long time), such that it will emerge in local environments capable of sustaining it. See for example early arguments by Schrodinger<sup>[22]</sup> stating the the apparent macroscopic stability is due to the microscopic chaos resulting from random events, and invoking a net entropy gain by the universe due to the continual energy transformation despite the heavy entropy investment in maintaining highly

organized living organisms. More recently, England has discussed the statistical physics of self-replication.<sup>[23]</sup>

In the context of this work, the previous discussion of molecular and virus evolution leads us to hypothesize that one of the key factors in the molecular evolution problem faced by the precursor SARS-CoV virus on its way to mutating into SARS-CoV-2 was solved by favoring those changes in RBD(S) that lead to an improved ability to locate available sites for hydrogen bonding in the host cell, ability that is further enhanced by the slightly basic pH  $\approx$  7.4 found in physiological environments. This improved hydrogen bonding capabilities may be achieved in a number of ways, for example, incorporating aminoacids with more acidic protons, or incorporating larger aminoacids whose hydrogen bonding regions are simply closer to the receptor, among others. We support the need for improved hydrogen bonding as the selective pressure in virus mutation hypothesis in the following evidence:

1. Table 1 shows that fragment to fragment contacts are all in the form of hydrogen bonds. For the specific case of the SARS-CoV-2 virus, in five of the six identified contacts, including the most persistent HBs, the residues in RBD(S) act as donors to the corresponding hydrogen bond
2. Besides a small sheet and a small helix (Figure 4), there is no secondary structure in RBD(S), thus, the receptor binding domain of the spike protein has a high structural flexibility which allows the virus to probe for available hydrogen bonding sites in the receptor, which in contrast has well defined secondary and tertiary structures in the interaction region
3. We obtained from the GenBank the sequences of aminoacids for the precursor SARS-CoV (ID AFR58742.1) and for the mutated SARS-CoV-2 (ID QHD43416.1) viruses.<sup>[24]</sup> We compare below only the 196 aminoacids in the RBD(S) and highlight in red the receptor binding motif (RBM). We also underline aminoacid substitution in the mutated virus

<b>SARS--CoV</b>	NITNLCPFGE	VFNATKFPSV	YAWERKKISN	CVADYSVLYN	STFFSTFKCY	<b>367</b>
<b>SARS--CoV--2</b>	NITNLCPFGE	VFNAT <u>RF</u> ASV	YAW <u>NRKR</u> ISN	CVADYSVLYN	<u>S</u> ASFSTFKCY	<b>380</b>
<b>SARS--CoV</b>	GVSATKLNDL	CFSNVYADSF	VVKGDDVRQI	APGQTGVIAD	YNYKLPDDFM	<b>417</b>
<b>SARS--CoV--2</b>	GVS <u>P</u> TKLNDL	C <u>F</u> TNVYADSF	<u>V</u> IRGDEV <u>R</u> QI	APGQT <u>G</u> KIAD	YNYKLPDD <u>F</u> T	<b>430</b>
<b>SARS--CoV</b>	GCVLAWNTRN	<u>I</u> DATSTGNYN	<u>Y</u> KYR <u>Y</u> LRHGK	<u>L</u> RP <u>F</u> ERDISN	<u>V</u> PFSPDGKPC	<b>467</b>
<b>SARS--CoV--2</b>	GCV <u>I</u> AWNS <u>N</u> N	<u>L</u> DSK <u>V</u> GGNYN	<u>Y</u> LY <u>R</u> L <u>F</u> ERKSN	<u>L</u> K <u>P</u> FERD <u>I</u> ST	<u>E</u> I <u>Y</u> QAGST <u>P</u> C	<b>480</b>
<b>SARS--CoV</b>	<u>T</u> P- <u>P</u> ALNCYW	<u>P</u> LNDYGFY <u>T</u> T	<u>T</u> GIGYQ <u>P</u> YRV	VVLSFELLNA	PATVCGPKLS	<b>516</b>
<b>SARS--CoV--2</b>	<u>N</u> G <u>V</u> EG <u>F</u> NC <u>Y</u> F	<u>P</u> L <u>Q</u> SYGF <u>Q</u> P <u>T</u>	<u>N</u> GVGYQ <u>P</u> YRV	VVLSFELL <u>H</u> A	PATVCGPK <u>K</u> S	<b>530</b>

From this sequences, we point out that

- (a) There are a total of 49 substitutions ( $\approx 25\%$ ) in the RBD(S) of SARS-CoV-2
- (b) Most of the mutations occur in the actual interaction region: the rate of substitution in the binding motif is quite higher: 36 substitutions that amount to  $\approx 50\%$  in RBM
- (c) Estimating the overall acidity of large biomolecules is an open problem, there are in fact many available algorithms to calculate isoelectric points of peptides and proteins.<sup>[25–29]</sup> Here, we take a pragmatic approach to determine relative acidities between the precursor SARS-CoV and the mutated SARS-CoV2 viruses. We took averages of the experimental isoelectric points<sup>[30]</sup> (IEP) for the 49 aminoacids involved in the mutation, that is, we calculated the average IPE in the replaced aminoacids in RBD(S) and found 6.50 and 6.44 for the precursor and for the mutated viruses, respectively. We also calculated the same averages for the interaction motifs only and obtained 6.44 and 5.98 over the 34 mutations. Thus the mutated virus is collectively considerably more acidic in the interaction region, which improves its ability to donate protons to hydrogen bonds

## 4 Summary and conclusions

The most important contributions from this work may be summarized as follows:

1. We pinpoint the specific residue (in the virus) to residue (in the cell) interactions during the initial virus···cell binding
2. We characterize the virus···cell molecular attachment as the result of a large number of non-covalent contacts between the receptor binding domain in the spike protein of the virus and the angiotensin converting enzyme 2 in the receptor expressed in epithelial human cells. This large surface of non-covalent interactions is enhanced by water molecules located in the interfragment region and acts in conjunction with an extended network of hydrogen bonds
3. The need for improved hydrogen bonding is the selective pressure in virus mutation. Thus, one of the key factors in the molecular evolution problem faced by the precursor SARS-CoV virus on its way to mutating into SARS-CoV-2 was solved by favoring those changes in RBD(S) that lead to an improved ability to locate available sites for hydrogen bonding in the host cell
4. Mutated SARS-CoV-2 is more contagious than the precursor SARS-CoV because it is better at finding hydrogen bonding sites in the receptor cell. More specifically, virus mutations have produced SARS-CoV-2, a virus that is more acidic in the interaction region than SARS-CoV
5. The calculated QM/MM absorption spectra show that as a result of virus···cell binding, absorption bands are shifted and spectral intensities are quenched, thus suggesting that significant changes occur at the electronic level as a consequence of binding.



## 5 Computational Methods

The starting point of our calculations was the complex between the receptor binding domain of the S protein, RBD(S), and the ACE2 receptor. Cartesian coordinates for the RBD(S)···ACE2 complex were taken from the protein data bank (PDB ID 6LZG<sup>[31]</sup>) and then treated with CHARMM-GUI, the graphical user interface of CHARMM<sup>[32,33]</sup> to include missing hydrogen atoms at pH = 7.4, to ensure that all glycans are included, and to construct the force field. The entire system was enclosed by a truncated octahedral box such that the smallest atom···wall distance was set to 9 Å, then the available volume in the box was filled with TIP3P<sup>[34]</sup> water molecules. NaCl molecules were added until a physiological 0.154 M concentration was attained and, finally, counterions were added to restore charge neutrality. This procedure led to a system comprising a total of 171161 atoms, with 52735 water molecules, 9585 atoms in the receptor, and 3034 atoms in RBD(S).

The system was subjected to a steepest descent energy minimization in order to correct for potential inconsistencies in atom coordinates that may arise during the procedure of randomly filling the available space with water, NaCl, and counterions. Once minimized, we ran triplicate all-atom MD simulations under the conditions summarized in Table 2 and described next. First, there were three equilibration steps lasting a total of 0.625 nanoseconds (ns) with 1 femtosecond (fs) time intervals, during which the structural constraints were progressively relaxed until finally being totally lifted. These structural constraints were imposed by harmonic constants that prevent deformation of the backbone ( $k_{bb}$ ), side chain ( $k_{sc}$ ), and dihedrals ( $k_d$ ). Then, the system underwent a production step lasting 600 ns with time intervals of 2 fs. For all MD runs, the Lennard-Jones potential was softened starting at 0.8 nm until eventually vanishing at 1.0 nm. Also, the cutoff radius for electrostatic interactions was set to 1.0 nm. All these simulations were conducted using the CHARMM36m force field<sup>[35,36]</sup> as implemented in GROMACS 2019.4<sup>[37]</sup> at 310.15 K and 1 bar.

The RBD(S)···ACE2 interaction energy ( $\Delta G_{int}$ ) was estimated via the MM/PBSA

Table 2: **Conditions for the MD simulations of the RBD(S)···ACE2 complex.** All times intervals (0 – 600.625) in nanoseconds and all constants in  $\text{kJ mol}^{-1} \text{nm}^{-2}$ . Harmonic  $k_{bb}$ ,  $k_{sc}$ ,  $k_d$  constants prevent deformation of the backbone, sidechains, and dihedrals, respectively.

Parameter	0 – 0.125	0.125 – 0.325	0.325 – 0.625	0.625 – 600.625
$k_{bb}$	400	400	0	0
$k_{sc}$	40	40	0	0
$k_d$	4	4	0	0
Ensemble	NVT	NPT	NPT	NPT
Thermostat	v-rescale	v-rescale	v-rescale	Parrinello–Raman
Barostat	–	Berendsen	Berendsen	Nose–Hoover

method<sup>[38]</sup> as implemented in GROMACS.<sup>[37]</sup> Dielectric constants were set to  $\epsilon_{solute} = 1$ ,  $\epsilon_{solvent} = 80$  at the simulation temperature. In short,  $\Delta G_{int} = \Delta E_{virus\cdots cell} + \Delta G_p + \Delta G_{np} - T\Delta S$ . Here,  $\Delta E_{virus\cdots cell}$  is the gas phase energy of the RBD(S)···ACE2 complex,  $\Delta G_p$ ,  $\Delta G_{np}$  are the solvation energies due to the polar and non-polar interactions, respectively, and  $T\Delta S$  is the entropy contribution. More precisely,  $\Delta G_p$  was computed under the Poisson–Boltzmann model,  $\Delta G_{np}$  was estimated using the solvent accessible surface area, and the entropy term was obtained from the model of Duan and coworkers.<sup>[39]</sup> To finally estimate  $\Delta G_{int}$ , we took 300 points in the 140–170 ns interval of one of the MD replicas.

It has been recently shown<sup>[16]</sup> that randomly chosen configurations from late stages of MD simulations are adequate sources to obtain deep insight into interfragment bonding. Accordingly, aiming at understanding the fundamental forces driving the attachment of RBD(S) to host cells, virus···cell bonding interactions were dissected following these steps:

1. Persistent residue (in the virus) to residue (in the host cell) contacts during the 600 ns of the MD simulations were identified using the VMD program<sup>[40]</sup> with a cutoff radius of 3.5 Å
2. One frame was randomly chosen from the late stages of one MD run
3. We extracted all extended interacting pairs in the chosen frame, kept them in the configurations they had in the interacting system (this is more accurate to understand

- the virus···cell bonding interactions than reoptimizing the isolated pairs), and
- (a) Computed accurate interaction energies using highly correlated domain based local pair–natural orbital coupled–cluster (DLPNO–CCSD(T)/aug–cc–pVDZ) single point energy calculations<sup>[41,42]</sup> on the dimers and in the monomers. The ORCA suite of programs, version 4.0.1.2, was used to this end<sup>[43]</sup>
  - (b) Dissected the intermolecular interactions using the tools provided by the natural bond orbitals (NBO<sup>[12–14]</sup> as implemented in NBO7.0<sup>[44]</sup>) and by the quantum theory of atoms in molecules (QTAIM<sup>[9–11]</sup> as implemented in AIMall<sup>[45]</sup>)
  - (c) Calculated QM/MM absorption spectra for the monomers and for the dimers. All TD–DFT calculations were carried out using the B3LYP/aug–cc–pVDZ model chemistry<sup>[46–49]</sup> (tests using the dispersion corrected B3LYP–D3,  $\omega$ B97xD, functionals yielded essentially identical results). QM/MM Electrostatic Embedding was exploited,<sup>[20]</sup> in which only the extended dimers were considered as the quantum region, and the rest of the system as the MM region, which was modelled by the Amber force field<sup>[19]</sup> and by assigning to atom types the same charges used in the MD runs. A large number of excited states are needed to guarantee that both the intensities and shapes of the absorption spectra are accurately reproduced. Therefore, in this work the first 20 excited states were computed at the TD–DFT/QM–MM level in each case. Vertical excitations were shifted by -0.7 eV to account for the systematic error due to the choice of functional. This value was chosen in order to match the experimental absorption maximum for Tyrosine.<sup>[50,51]</sup> All Spectra were then convoluted with Gaussian lineshapes with full width half maximum (FWHM) of 0.6 eV. All QM/MM calculations were carried out with Gaussian16<sup>[52]</sup>
4. We isolated the interaction region by including everything within a 3.0 Å radius from the last atom at the end of each aminoacid (1325 atoms in total) and calculated the

interfragment non covalent interaction (NCI as implemented in NCIPLOT<sup>[53]</sup>) surface using the promolecular densities approximation.<sup>[17,18]</sup>

## Acknowledgement

Internal support from Universidad de Antioquia via “Estrategia para la sostenibilidad” is acknowledged. Partial funding for this project from H2020–MSCA–ITN–2017 European Training Network “Computational Spectroscopy In Natural sciences and Engineering” (CO-SINE), grant number 765739 is also acknowledged. N.R. thanks Colciencias for her doctoral scholarship.

## Supporting Information Available

Cartesian coordinates for all dimer pairs listed in Table 1 and for the fragment taken for the NCI surfaces. A figure with the minimum interacting distances used to determine the interaction regions is included as well. A video of one of the trajectories is also provided.

## References

- [1] D. Schoeman,; B. C. Fielding, *Virology journal* **2019**, *16*, 69.
- [2] Y. Yuan,; D. Cao,; Y. Zhang,; J. Ma,; J. Qi,; Q. Wang,; G. Lu,; Y. Wu,; J. Yan,; Y. Shi,; et al. *Nature communications* **2017**, *8*, 15092.
- [3] S. Liu,; G. Xiao,; Y. Chen,; Y. He,; J. Niu,; C. R. Escalante,; H. Xiong,; J. Farmar,; A. K. Debnath,; P. Tien,; et al. *The Lancet* **2004**, *363*, 938–947.
- [4] W. Tai,; L. He,; X. Zhang,; J. Pu,; D. Voronin,; S. Jiang,; Y. Zhou,; L. Du, *Cellular & Molecular Immunology* **2020**, 1–8, <https://doi.org/10.1038/s41423-020-0400-4>.

- [5] M. Hoffmann,; H. Kleine-Weber,; S. Schroeder,; N. Krüger,; T. Herrler,; S. Erichsen,; T. S. Schiergens,; G. Herrler,; N.-H. Wu,; A. Nitsche, et al. *Cell* **2020**, *181*, 271–280.e8.
- [6] D. Wrapp,; N. Wang,; K. S. Corbett,; J. A. Goldsmith,; C.-L. Hsieh,; O. Abiona,; B. S. Graham,; J. S. McLellan, *Science* **2020**, *367*, 1260–1263.
- [7] Y. Han,; P. Král, *ACS Nano* **2020**, *14*, 5143–5147.
- [8] K. Wiberg, *Tetrahedron* **1968**, *24*, 1083 – 1096.
- [9] R. Bader, *Atoms in Molecules: A Quantum Theory*; Oxford Univ. press Oxford, 1990.
- [10] E. Espinosa,; I. Alkorta,; J. Elguero,; E. Molins, *The Journal of Chemical Physics* **2002**, *117*, 5529–5542.
- [11] S. J. Grabowski, *Chem. Rev.* **2011**, *111*, 2597–2625.
- [12] A. E. Reed,; F. Weinhold, *J. Chem. Phys.* **1983**, *78*, 4066–4073.
- [13] F. Weinhold,; C. R. Landis, *Discovering Chemistry with Natural Bond Orbitals*; Wiley-VCH, Hoboken NJ, 319pp, 2012.
- [14] E. D. Glendening,; C. R. Landis,; F. Weinhold, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 1–42.
- [15] Y. Wang,; M. Liu,; J. Gao, *Proceedings of the National Academy of Sciences* **2020**, *117*, 13967–13974.
- [16] N. Rojas-Valencia,; S. Gómez,; S. Montillo,; M. Manrique-Moreno,; C. Cappelli,; C. Hadad,; A. Restrepo, *Journal of Physical Chemistry B* **2020**, *124*, 79–90.
- [17] G. A. DiLabio,; A. Otero-de-la Roza, *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2016; Chapter 1, pp 1–97.

- [18] E. R. Johnson,; S. Keinan,; P. Mori-Sánchez,; J. Contreras-García,; A. J. Cohen,; W. Yang, *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.
- [19] W. D. Cornell,; P. Cieplak,; C. I. Bayly,; I. R. Gould,; K. M. Merz,; D. M. Ferguson,; D. C. Spellmeyer,; T. Fox,; J. W. Caldwell,; P. A. Kollman, *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.
- [20] S. Dapprich,; I. Komáromi,; K. Byun,; K. Morokuma,; M. J. Frisch, *Journal of Molecular Structure: THEOCHEM* **1999**, *461-462*, 1 – 21.
- [21] B. Korber,; W. M. Fischer,; S. Gnanakaran,; H. Yoon,; J. Theiler,; W. Abfalterer,; N. Hengartner,; E. E. Giorgi,; T. Bhattacharya,; B. Foley, et al. *Cell* **2020**, *182*, 812–827.
- [22] E. Schrödinger, *What is Life? The Physical Aspect of the Living Cell*. 1944; Cambridge at the University Press.
- [23] J. England, *Journal of Chemical Physics* **2013**, *139*, 121923.
- [24] F. Wu,; S. Zhao,; B. Yu,; Y.-M. Chen,; W. Wang,; Z.-G. Song,; Y. Hu,; Z.-W. Tao,; J.-H. Tian,; Y.-Y. Pei, et al. *Nature* **2020**, *579*, 265–269.
- [25] B. Bjellqvist,; G. J. Hughes,; C. Pasquali,; N. Paquet,; F. Ravier,; J. C. Sanchez,; S. Frutiger,; D. Hochstrasser, *Electrophoresis* **1993**, *14*, 1023–1031.
- [26] S. Gauci,; B. van Breukelen,; S. Lemeer,; J. Krijgsveld,; A. J. R. Heck, *Proteomics* **2008**, *8*, 4898–4906.
- [27] E. Gasteiger,; A. Gattiker,; C. Hoogland,; I. Ivanyi,; R. D. Appel,; A. Bairoch, *Nucleic Acids Research* **2003**, *31*, 3784–3788.
- [28] B. J. Cargile,; J. R. Sevinsky,; A. S. Essader,; J. P. Eu,; J. L. Stephenson, *Electrophoresis* **2003**, *29*, 2768–2778.

- [29] Y. Perez-Riverol,; E. Audain,; A. Millan,; Y. Ramos,; A. Sanchez,; J. A. Vizcaíno,; R. Wang,; M. Müller,; Y. J. Machado, *Journal of Proteomics* **2012**, *75*, 2269–2274.
- [30] G. Lang,; D. Lide, *CRC Handbook of Chemistry and Physics*; ed. DR Lide, 72nd Edition, CRC Press, Boca Raton, 1991; pp 4–133.
- [31] Q. Wang,; Y. Zhang,; L. Wu,; S. Niu,; C. Song,; Z. Zhang,; G. Lu,; C. Qiao,; Y. Hu,; K.-Y. Yuen, et al. *Cell* **2020**, <https://doi.org/10.1016/j.cell.2020.03.045>.
- [32] S. Jo,; T. Kim,; V. G. Iyer,; W. Im, *Journal of computational chemistry* **2008**, *29*, 1859–1865.
- [33] J. Lee,; X. Cheng,; J. M. Swails,; M. S. Yeom,; P. K. Eastman,; J. A. Lemkul,; S. Wei,; J. Buckner,; J. C. Jeong,; Y. Qi, et al. *Journal of chemical theory and computation* **2016**, *12*, 405–413.
- [34] W. L. Jorgensen,; J. Chandrasekhar,; J. D. Madura,; R. W. Impey,; M. L. Klein, *The Journal of chemical physics* **1983**, *79*, 926–935.
- [35] J. Huang,; S. Rauscher,; G. Nawrocki,; T. Ran,; M. Feig,; B. L. de Groot,; H. Grubmüller,; A. D. MacKerell Jr, *Nature methods* **2017**, *14*, 71.
- [36] O. Guvench,; S. S. Mallajosyula,; E. P. Raman,; E. Hatcher,; K. Vanommeslaeghe,; T. J. Foster,; F. W. Jamison,; A. D. MacKerell Jr, *Journal of chemical theory and computation* **2011**, *7*, 3162–3180.
- [37] M. J. Abraham,; T. Murtola,; R. Schulz,; S. Páll,; J. C. Smith,; B. Hess,; E. Lindahl, *SoftwareX* **2015**, *1*, 19–25.
- [38] R. Kumari,; R. Kumar,; . Open Source Drug Discovery Consortium,; A. Lynn, *Journal of Chemical Information and Modeling* **2014**, *54*, 1951–1962.
- [39] L. Duan,; X. Liu,; J. Z. Zhang, *Journal of the American Chemical Society* **2016**, *138*, 5722–5728.

- [40] W. Humphrey;; A. Dalke;; K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33–38.
- [41] C. Riplinger;; F. Neese, *The Journal of Chemical Physics* **2013**, *138*, 034106.
- [42] C. Riplinger;; B. Sandhoefer;; A. Hansen;; F. Neese, *The Journal of Chemical Physics* **2013**, *139*, 134101.
- [43] F. Neese, *WIREs Computational Molecular Science* **2012**, *2*, 73–78.
- [44] E. D. Glendening;; J. K. Badenhoop;; A. E. Reed;; J. E. Carpenter;; J. A. Bohmann;; C. M. Morales;; P. Karafiloglou;; C. R. Landis;; F. Weinhold, NBO 7.0. 2018; Theoretical Chemistry Institute, University of Wisconsin, Madison, WI.
- [45] T. Keith, AIMALL (*version 19.10.12*). 2019; TK Gristmill Software, Overland Park KS, USA, aim.tkgristmill.com.
- [46] C. Lee;; W. Yang;; R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
- [47] A. D. Becke, *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- [48] R. A. Kendall;; T. H. Dunning;; R. J. Harrison, *The Journal of Chemical Physics* **1992**, *96*, 6796–6806.
- [49] T. H. Dunning, *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.
- [50] D. Creed, *Photochemistry and Photobiology* **1984**, *39*, 563–575.
- [51] L. H. Fornander;; B. Feng;; T. Beke-Somfai;; B. Nordén, *The Journal of Physical Chemistry B* **2014**, *118*, 9247–9257.
- [52] M. J. Frisch;; G. W. Trucks;; H. B. Schlegel;; G. E. Scuseria;; M. A. Robb;; J. R. Cheeseman;; G. Scalmani;; V. Barone;; G. A. Petersson;; H. Nakatsuji, et al. Gaussian 16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.



- [53] J. Contreras-García,; E. R. Johnson,; S. Keinan,; R. Chaudret,; J.-P. Piquemal,; D. N. Beratan,; W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625–632.