Contact residue contributions to interaction energies between SARS-CoV-1 spike proteins and human ACE2 receptors

Jorge H. Rodriguez (jhrodrig@purdue.edu)

Computational Biomolecular Physics Group, Department of Physics & Astronomy, Purdue University, West Lafayette, IN 47907, USA

Akshita Gupta

Purdue University

Research Article

Keywords: SARS, SARS-CoV-1, SARS-CoV-2, COVID-19, cryo-EM, X-ray crystallography, ACE2, binding energy, spike proteins, interaction energies, SARS ACE2 interaction

Posted Date: October 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-95153/v1

Version of Record: A version of this preprint was published on January 13th, 2021. See the published version at https://doi.org/10.1038/s41598-020-80942-6.

Contact residue contributions to interaction energies between SARS-CoV-1 spike proteins and human ACE2 receptors

Jorge H. Rodriguez* and Akshita Gupta

Computational Biomolecular Physics Group, Department of Physics and Astronomy

Purdue University, West Lafayette, Indiana 47907-2036, USA

*E-mail: jhrodrig@purdue.edu

ABSTRACT

Several viruses of the *corona* family interact, *via* their spike (S) proteins, with human cellular receptors. Spike proteins of SARS-CoV-1 and SARS-CoV-2 virions, being structurally related but not identical, mediate attachment to the human angiotensin-converting enzyme 2 (hACE2) receptor in similar but non-identical ways. Molecular-level understanding of interactions between spike proteins and hACE2 can aid strategies for blocking attachment of SARS-CoV-1, a potentially reemerging health threat, to human cells. We have uniquely identified dominant molecular-level interactions, some attractive and some repulsive, between the *receptor binding domain* of SARS-CoV-1 spike proteins (S-RBD) and hACE2. We performed fragment-based quantum-biochemical calculations which directly relate biomolecular structure with the hACE2...S-RBD interaction energy. Consistent with X-ray crystallography and cryo-EM, the interaction energy between hACE2 and S-RBD (≈ -26 kcal/mol) corresponds to a net intermolecular attraction which is significantly enhanced by inclusion of *dispersion* van der Waals forces. Protein fragments, at the hACE2...S-RBD interface, that dominate host-virus attraction have been identified together with their constituent amino acid residues. Two hACE2 fragments which include residues (GLU37, ASP38, TYR41, GLN42) and (GLU329, LYS353, GLY354), respectively, as well as three S-RBD fragments which include residues (TYR436), (ARG426) and (THR487, GLY488, TYR491), respectively, have been identified as primary attractors at the hACE2...S-RBD interface.

INTRODUCTION

The severe acute respiratory syndrome coronavirus, SARS-CoV-1, represents a potentially reemerging and not fully understood health threat [1, 2] that originated in late 2002. While the threat from SARS-CoV-1 faded with the aid of effective health mitigaton policies, other coronaviruses have recently emerged including the genetically related SARS-CoV-2. [3–6] Although related, SARS-CoV-1 and SARS-CoV-2 display important differences in their host-binding structures, namely the *receptor binding domains* (RBD) of their spike (S) proteins. [7] Both, public health concerns and RBD structural variations, underscore the need to study molecular-level interactions of each particular coronavirus with host-cell receptors. Such studies can elucidate the physico-chemical origins and residue-level mechanisms of viral infection. This work presents a quantitative structure-based analysis of key interactions, largely responsible for an attractive host-virus binding energy, between SARS-CoV-1 and the human ACE2 receptor.

Virions of the coronaviruses SARS-CoV-1 and SARS-CoV-2 have characteristic roughly-spherical shapes, on the order of 100 nm in diameter, with petal-shaped spikes which project outwards from their surfaces. [4] Coronaviruses encode three types of surface proteins, [5, 8] namely membrane (M), envelope (E) and, of particular importance, the so-called spike (S) which are positioned in their membrane envelopes. In addition another, nucleocapsid (N), structural protein is encoded. Spike proteins play crucial roles in a virion's infection of host-cells, both by binding to their cellular receptors and, subsequently, by promoting fusion with their cellular membranes. [4, 9, 10] To interact with their host-cell receptors, spike proteins undergo conformational motions that either hide or expose their structural determinants of receptor binding which correspond to their *down* or *up* states, respectively. [11] Spike glycoproteins of coronaviruses play crucial roles in the initial stages of host-cell infection and are major targets for virus-neutralizing antibodies. [12] Thus, identifying and studying in a quantitative way the physico-chemical interactions between the *up*-state spike conformations, which are receptor accessible, [11, 13] and their host receptors is of interest not only for elucidating the molecular-level origins of host-virus binding but also for developing therapeutic countermeasures.

Structurally, the spikes of coronaviruses are globular trimers, of about 150 $\mathrm{\AA}$ in diameter, linked to the virion envelope by a narrow stalk. The spikes of SARS-CoV-1, in particular, are fairly massive ($\approx$ 500 kDa) in comparison to other type I viral spike proteins. [9] Spike proteins of both

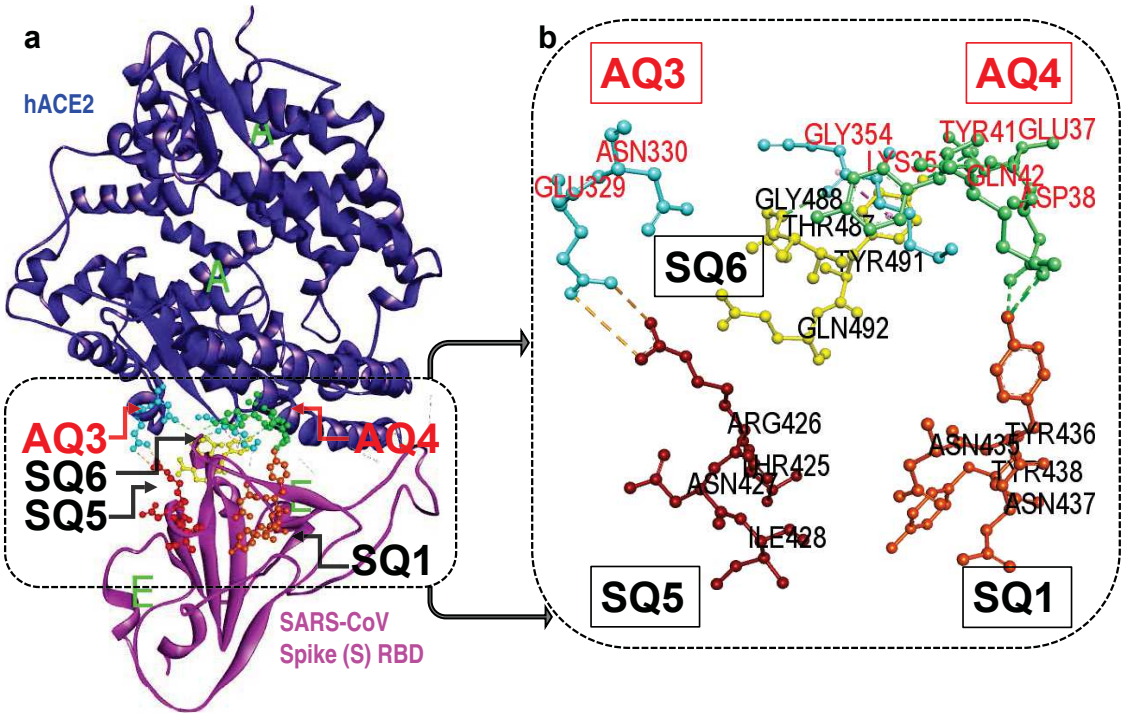

FIG. 1. | **Identification of four-residue fragments (i.e. *quartets*) which produce attractive interaction energies between hACE2 and the SARS-CoV-1 S-RBD. a,** Structure of hACE2 receptor (Chain A) in complex with SARS-CoV-1 spike protein (Chain E). [3] The key *quartets*, at the hACE2...S-RBD interface, promoting host-virus binding are shown in the dashed box. **b,** Magnified view of the hACE2 (**AQ3**, **AQ4**) and S-RBD (**SQ1**, **SQ5**, **SQ6**) residue *quartets* which mostly contribute to the attractive hACE2...S-RBD interaction energy (shown in ball and stick).

viruses, SARS-CoV-1 and SARS-CoV-2, contain two domains labelled S1 and S2. The S1 domain mediates initial virus binding to target cell receptors whereas S2 is involved in the fusion of virus and target cell membranes. [1] As shown in Fig. 1 spike (S) proteins interact with host receptors *via* their *receptor-binding domain* (RBD) which is herein referred to as S-RBD. A representative S-RBD which is closely related to that of SARS-CoV-1, namely that of SARS-CoV-2, has a molecular weight of $\approx$ 21 kDa and its prefusion cryo-EM structure has been recently reported. [11]

Coronavirus spike proteins, such as those from SARS-CoV-1, need to interact with receptors of their target cells to initiate infection. The human angiotensin-converting enzyme 2 (hACE2), attached to the outer surface of host-cells, has been identified as an efficient binder of the S1 domain of SARS-CoV-1 spike proteins. [10, 14] The hACE2 motif was also identified as an entry receptor for the novel SARS-CoV-2 coronavirus S-protein. [15–17] Thus, blocking S-protein

interaction with hACE2 or promoting hACE2 conformational changes that render it inefficient as a receptor are possible antiviral countermeasures. Conversely, identification of hACE2 receptors as viral entry points highlighted the role of spike protein RBDs as possible target epitopes of S1-protein-based vaccines. [14]

An important determinant of infectivity is the cognate interaction between viral attachment proteins and their host-cell receptors. [18] The structural basis for host receptor recognition has been reviewed for several coronaviruses and binding similarities as well as differences have been highlighted. [19] The main contact residues at the interface of the SARS-CoV-1 S-RBD with ACE2 receptors from several species, including human ACE2, have been structurally identified. [3] The crystallographic structure of the SARS-CoV-1 S-RBD, in complex with hACE2, was reported at 2.9 Å resolution [3] and more recent Cryo-EM structures provide additional insight about prefusion to postfusion conformational changes. [13, 20] Likewise, X-ray diffraction and Cryo-EM structures of the structurally related S-RBD of SARS-CoV-2 [11] and its complex with hACE2 receptors [21, 22] have been reported.

The interaction energy, as defined by Eqs. 1-3, is a measure of the propensity of the viral S-RBD to attach itself to an ACE2 receptor. Although the contact residues at the hACE2...S-RBD interface have been structurally identified, no quantitative assesment of their contributions, individually or as part of a fragment, to the overall host-virus interaction energy has been reported. Techniques such as X-ray crystallography (XRC) and cryogenic electron microscopy (cryo-EM) can identify the contact residues at the hACE2...S-RBM interface. However, these techniques cannot unequivocally determine which S-RBM residue fragments are attractive or which are repulsive relative to hACE2. Likewise, XRC or cryo-EM cannot accurately quantify partial or total strengths of hACE2...S-RBM interaction energies. By contrast, such information, helpful for antiviral or vaccine development, can be obtained via rigorous quantum biochemical calculations as shown in the present study. Quantum biochemical calculations [23] can, to a large extent, explain the origin of attractive energies between spike proteins, in their *up* prefusion state, and host-cell receptors. We implemented a fragment-based quantum biochemical method that evaluates the strength and detailed nature, i.e. attractive or repulsive, of ACE2 interactions with S-protein *receptor binding domains*. We used the SARS-CoV-1...hACE2 crystallographic structure [3] to perform such fragment-based calculations that clearly identify which contact residue fragments give rise to the attractive hACE2...S-RBD interaction energy and, therefore, promote viral infection.

The *receptor binding motif* (S-RBM) of spike proteins, an integral and main functional component of their S-RBD, is at the interface which potentially binds to a host receptor such as hACE2. Importantly, despite a sequence identity of about 72-73% between the *domains* (S-RBD) of SARS-CoV-1 and SARS-CoV-2, the identity of their respective *motifs* (S-RBM) is significantly lower, only about 47.8% [24]. Thus, although structural similarities may produce *some* similar interaction mechanisms between the S-RBD of SARS-CoV-1 and SARS-CoV-2 with hACE2, their S-RBM structural differences [7] will likely produce other, concomitant but different, attractive or repulsive hACE2...S-RBD interactions. To develop therapeutic drugs and to understand the action of antibodies [25] which target viral spike proteins, it is useful to study each specific viral S-RBD and their interactions with hACE2. In this work we focus on identifying the main, molecular level, interactions between the S-RBD of SARS-CoV-1, a potentially reemerging public health threat, and hACE2.

The ability of coronaviruses to recognize their host-cell receptors is a first and crucial determinant of their host range and infectivity. It has been realized that the process of recognition is not due to accidental or random intermolecular events but to viral-RBD and host-receptor structural complementarity. [26] Less attention has been paid, however, to specific and concomitant energetic complementarities which favor non-covalent attraction at the viral-host interface. Here, we establish such quantitative link between structural complementarity and concomitant physicochemical viral-host non-covalent interactions. We implemented a fragment-based quantum biochemical method to study the hACE2...S-RBD interface. We report, in units of kcal/mol, the *total* interaction energy between contact residues of hACE2 and the SARS-CoV-1 S-RBD. In addition, we evaluate *partial* interaction energies between specific sets of four hACE2 residues, herein called *quartets*, with their neighboring S-RBD residues. Thus, we identify which hACE2 *quartets* are attractive and which are repulsive relative to the SARS-CoV-1 S-RBD. Likewise, we identify which S-RBD residue *quartets* are attractive or repulsive relative to the hACE2 receptor. Our results enhance the understanding of molecular-level mechanisms of hACE2 and S-RBD recognition and, in addition, identify potential therapeutic targets and SARS-CoV-1 epitopes.

TABLE I. | **Energies[a] of the human receptor (hACE2), spike protein binding domain (S-RBD) and their interaction energies without (E_{Int}^{DFT}) and with (E_{Int}^{DFT-DD}) van der Waals *dispersion* corrections [DD].[b]**

$E_{hACE2...S-RBD}$	E_{hACE2}	E_{S-RBD}	E_{Int}^{DFT}	E_{Int}^{DFT}	E_{Int}^{DD}	E_{Int}^{DFT-DD}
[Hartrees]	[Hartrees]	[Hartrees]	[Hartrees]	[kcal/mol]	[kcal/mol]	[kcal/mol]
-22240.4565923	-10953.2725200	-11287.1426313	-0.0414410	-26.00	-378.26	-404.26

[a] DFT energies computed with the B3LYP [27] functional and 6-31+G* basis in the gas-phase.
[b] Distance-dependent (DD) *dispersion* evaluated with the B3LYP-DD semiempirical method. [23]

RESULTS

Total and partial interaction energies between hACE2 and the SARS-CoV-1 S-RBD were computed, in the low temperature limit, *via* quantum biochemical calculations and the supermolecular approach. [23] A fragment-based methodology, by which proteins are divided into fragments, was used to evaluate *partial* interaction energies and identify the dominant, attractive or repulsive, sets of residues at the hACE2-S-RBD interface. All calculations were based on all-electron dispersion-corrected [23] density functional theory. [27, 28]

Attractive nature of the hACE2...S-RBD interaction. Table I shows that the net interaction between hACE2 and the S-RBD is attractive as indicated by the negative sign of their interaction energy (E_{Int}^{DFT-DD}) . This finding confirms and is consistent with the tendency of the SARS-CoV-1 prefusion S-RBD to bind to the hACE2 receptor. [10, 14] The attractive nature of the interaction energies is also consistent with the structure of the virus-receptor interface, as displayed by the crystallographic structue, [3] which corresponds to a thermodynamically favorable conformation.

The hACE2...S-RBD interaction energy was calculated, separately, in gas and solvent phases with both results corresponding to a net intermolecular attraction. In addition, Vander Waals *dispersion* corrections were evaluated *via* the accurate B3LYP-DD methodology [23] which, in the gas phase, added a significant attractive contribution. The gas-phase interaction energies, in the absence and presence of *dispersion* corrections, were on the order of -26 kcal/mol and -404 kcal/mol, respectively, when evaluated with the 6-31+G* basis set (Table I). Similar trends were found from calculations with other basis sets as shown in Supplementary Table S1. It should be

noted that *partial* electrostatic contributions to the interaction energy can be attractive or repulsive which tends to lower the net additive magnitude of this mechanism. By contrast, *dispersion* contributions are additively attractive which explains the large energetic contribution of *dispersion* (E_{Int}^{DD}). *Dispersion* contributions were calculated at the short intermolecular distances corresponding to hACE2...S-RBD noncovalent attachment as displayed by the crystallographic structure. [3] At these short distances Vander Waals forces are particularly strong.

In contrast to gas-phase *dispersion*-corrected interaction energies (E_{Int}^{DFT-DD}), which in that limit are generally accurate to better than 1 kcal/mol [23], the calculation of solvent-phase interaction energies introduces greater uncertainties. Thus, the solvent-phase energies given in Supplementary Table S1 should be considered as rough approximations which illustrate the still attractive, although weaker, hACE2...S-RBD intermolecular interactions when solvation effects are taken into account.

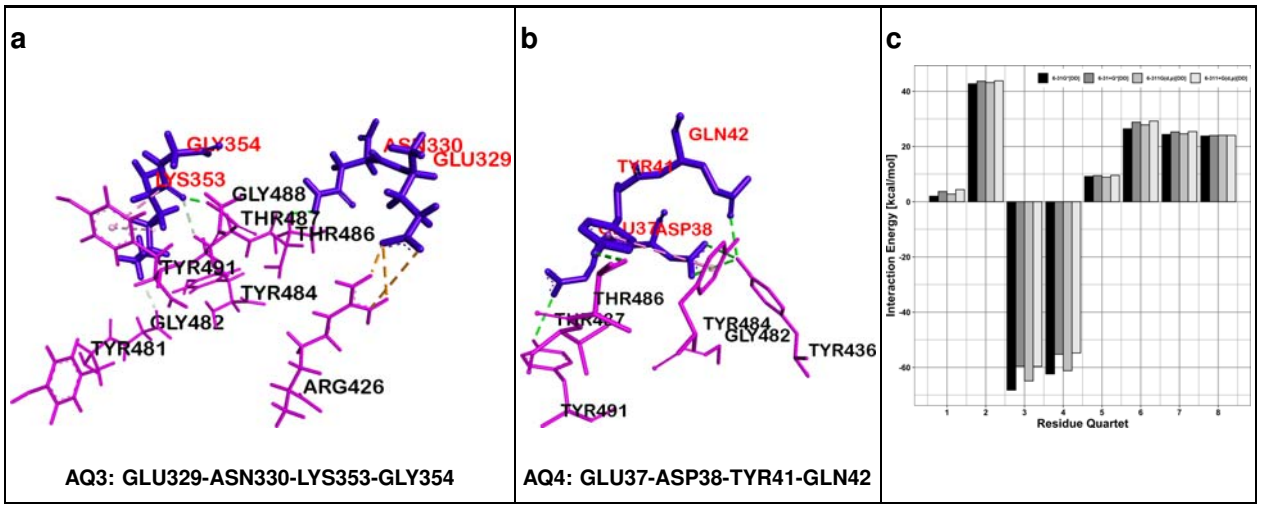

FIG. 2. | **The two hACE2-centered fragments producing a net attractive interaction towards S-RBD. a-b**, ACE2 *quartet* residues (shown in blue) and neighboring S-RBD residues (shown in pink) corresponding to the dominant *attractive* ACE2...S-RBD interactions. ACE2 is the structural reference. **c**, Main repulsive (positive) and attractive (negative) interactions [kcal/mol] between *quartets* of the human ACE2 receptor, used as structural references, and neighboring residues of the receptor binding domain of the SARS corona virus spike protein (S-RBD). The four adjacent vertical bars for each *quartet* correspond, from left to right, to *dispersion*-corrected [DD] [23] interaction energies evaluated with the 6-31G*, 6-31+G*, 6-311G(d,p) and 6-311+G(d,p) basis sets.

TABLE II. | **Human ACE2 receptor (hACE2) *quartets* and their interaction energies [kcal/mol][a] with neighboring[b] virus S-RBD residues.**

Quartet	Human ACE2 Receptor Residues	E_{Int}^{DFT}	E_{Int}^{DD}	E_{Int}^{Total}
AQ1	ASP30-LYS31-ASN33-HIS34	+17.50	−13.10	+4.41
AQ2	GLN24-ALA25-LYS26-THR27	+51.12	−7.32	+43.80
AQ3	GLU329-ASN330-LYS353-GLY354	−30.82	−28.81	−59.63
AQ4	GLU37-ASP38-TYR41-GLN42	−38.48	−16.30	−54.78
AQ5	LEU91-THR92-GLN325-GLY326	+11.82	−2.25	+9.57
AQ6	MET82-TYR83-GLN89-ASN90	+34.46	−5.23	+29.23
AQ7	SER44-LEU45-ALA46-SER47	+28.01	−2.60	+25.41
AQ8	SER77-THR78-LEU79-ALA80	+26.03	−2.01	+24.02

[a] DFT energies computed at 6-311+G(d,p)/B3LYP level; *Dispersion* (DD) corrections evaluated with semiempirical method. [23]

[b] All S-RBD residues within 4.5 Å of each ACE2 *quartet* were included.

Structural separation of the hACE2...S-RBD interface into *quartet* residue fragments. It is of great interest to identify the dominant sets of contact residues involved in physico-chemical attraction or repulsion between hACE2 and the S-RBD. It was determined that partitioning each protein structure into sets of four residues allowed for a qualitatively meaningful determination of intermolecular interaction energies. Protein fragments of smaller size did not include a minimum of nearest-neighbor and next-nearest-neighbor interactions between protein residues to allow for accurate estimates of *partial* contributions to the overall hACE2...S-RBD interaction energy. Therefore, sets of four ACE2 contact residues, herein referred to as ACE2 *quartets*, were selected together with their neighboring, i.e. within a range of 4.5 Å, viral S-RBD residues. An hACE2-centered *supermolecular fragment* is herein defined as a particular hACE2 residue *quartet* and its neighbouring S-RBD residues. Thus, any S-RBD residue localized in a region of strong noncovalent interaction with a particular ACE2 *quartet* was included in a respective *supermolecular fragment* of the hACE2...S-RBD complex. Examples of such hACE2...S-RBD *supermolecular fragments* are shown in Fig. 2,**a-b**. These construits were then used to compute *partial*, attractive or repulsive, interaction energies between particular hACE2 *quartets* and the S-RBD as reported in Table II and Fig. 2,**c**. Similarly, an S-RBD-centered *supermolecular fragment* constitutes a particular S-RBD residue *quartet* and its neighbouring hACE2 residues with examples given in Fig. 3,**a-c**.

Evaluation of *partial*, attractive or repulsive, hACE2...S-RBD interactions. The net attractive intermolecular interaction promotes the thermodynamic stability of the hACE2...S-RBD complex. Despite the net interaction being attractive, the calculated interaction energies ($\mathrm{E}_{\mathrm{Int}}^{\mathrm{DFT-DD}}$) can be interpreted as the combined result of several *partial* interactions, some attractive and some repulsive, between particular sets of hACE2 and S-RBD residues. The evaluation of *quartet*-centered *partial* contributions to the interaction energy allow the identification, as illustrated by Fig. 1, of which protein fragments are primarily responsible for binding energy of the ACE2...S-RBD complex. In addition, evaluation of *partial* interactions between hACE2 and S-RBD fragments, whether of attractive or repulsive character, provide molecular-level and energetic insight about the related processes of host-virus recognition and attachment.

Our results show that some *supermolecular fragments* at the ACE2...S-RBD interface are intrinsically attractive and thus directly favor the formation of the human receptor's complexation with the virus S-protein. Although other supermolecular fragments were found to be intrinsically repulsive, these too play a concomitant and important role in the formation of the ACE2...S-RBD complex. In fact the repulsive fragments, together with their attractive counterparts, help to guide the process of intermolecular recognition which ultimately leads to attachment. Among the supermolecular fragments that produce attractive hACE2...S-RBD interactions, *dispersion* forces were also found to play an important role. The latter correspond to *partial* contributions to the *dispersion* energy and are consistent with the importance of the Vander Waals mechanism previously uncovered for the total interaction energy (Table I) of the entire host-virus contact interface.

Identification of key hACE2-centered *quartet* interactions with S-RBD. Table II and Fig. 2,**c** display *partial* interaction energies between hACE2 *quartets* and their neighboring S-RBD residues. There are two ACE2 *quartets*, **AQ3** (GLU329-ASN330-LYS353-GLY354) and **AQ4** (GLU37-ASP38-TYR41-GLN42), whose interactions with S-RBD are strongly attractive as indicated by the magnitudes and negative signs, -59.63 and -54.78 kcal/mol, respectively, of their interaction energies. Fig. 2,**a-b** shows the structural composition of these two ACE2 *quartets* and their closely interacting S-RBD residues. The physico-chemical origin of the attractive nature of these *partial* ACE2...S-RBD interactions is not only related to conventional electrostatic effects, including hydrogen bonding, but also to sizable *dispersion* contributions (Table II). For *quartets* **AQ3** and **AQ4** *dispersion* contributions are on the order of -28.81 and -16.30 kcal/mol, respectively, corresponding to $\approx 48\%$ and $\approx 30\%$ of their *partial* interaction energies. Additional

electronic structure calculations were done using the same protocol but using other, closely related, computational basis sets. Supplementary Tables S2-S4 list the corresponding energies which display similar trends, thus confirming the intrinsically attractive nature of ACE2 *quartets* **AQ3** and **AQ4** with respect to S-RBD.

The quantum mechanical (*ab-initio*) character of the present calculations takes into account, at the same time, intermolecular interactions in the low temperature regime. Therefore, the present calculations do not separate or distinguish, contrary to traditional classifications, between particular types of intermolecular forces with the exception of *dispersion* contributions to Vander Waals forces. However, qualitatively, it is possible to relate some of our results to traditional classifications. To this effect, Fig. 2 shows some qualitative (color coded) assignments which include: i) amid-π interactions in (a) and π-π interactions in (b) (dotted pink lines) which, involving six-membered aromatic rings, more fundamentally correspond in the present work to *dispersion* forces; ii) conventional (dotted green lines) and non-conventional (dotted white lines) hydrogen bonds; and iii) electrostatic interactions (dotted yellow lines).

Despite the overall ACE2...S-RBD interaction as well as the dominant *partial* contributions being attractive, Fig. 2,**c** also shows that several of the ACE2 *quartets* are actually repulsive relative to S-RBD. ACE2 *quartet* **AQ2** is the most repulsive with a *partial* interaction energy of about $+43.80$ kcal/mol (Table II) which includes a large repulsive contribution ($\approx +51.12$ kcal/mol) and only a small (≈ -7.32 kcal/mol) *dispersion* component. The structure of the corresponding *supermolecular fragment* is shown in Supplementary Fig. S1.

Identification of key S-RBD-centered *quartet* interactions with hACE2. Table III and Figure 3,**d** show *partial* energies corresponding to spike protein (S-RBD) *quartets* interacting with neighbouring hACE2 residues. There are two S-RBD *quartets*, **SQ5** (THR425-ARG426-ASN427-ILE428) and **SQ6** (THR487-GLY488-TYR491-GLN492), which dominate the attractive interactions with hACE2 and lead to *partial* interaction energies of -57.57 and -42.15 kcal/mol, respectively. Consistent with the absence of six-membered rings no significant *dispersion* contribution was evaluated for *quartet* **SQ5** but, for the opposite reason, *dispersion* contributions were much more prominent for the intermolecular interaction of *quartet* **SQ6** (≈ -28 kcal/mol). In addition, S-RBD *quartet* **SQ1** (ASN435-TYR436-ASN437-TYR438) produced a substantially weaker attraction relative to hACE2. Additional electronic structure calculations were done with the same protocol but using other, closely related, computational basis sets. Supplementary Tables S5-S7 list the

corresponding energies which display similar trends, thus confirming a dominant and intrinsically attractive nature of S-RBD *quartets* **SQ5** and **SQ6** with respect to hACE2. The weaker attractive nature of S-RBD *quartet* **SQ1** was also confirmed by the data in the Supplementary Tables.

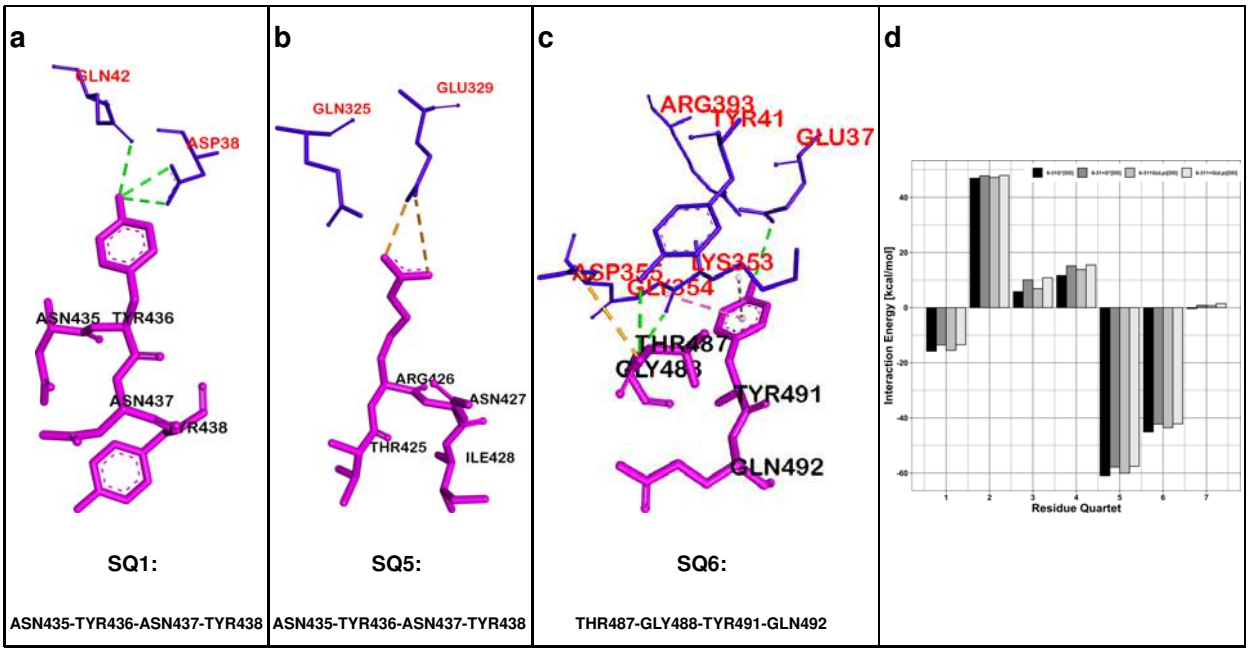

FIG. 3. | **The three S-RBD-centered fragments producing a net attractive interaction towards hACE2. a-c**, S-RBD *quartet* residues (shown in pink) and neighboring ACE2 residues (shown in blue) corresponding to the dominant *attractive* ACE2...S-RBD interactions. S-RBD is the structural reference. **d**, Main repulsive (positive) and attractive (negative) interactions [kcal/mol] between *quartets* of the SARS-CoV-1 S-RBD, used as structural references, and neighboring residues of the human hACE2 receptor. The four adjacent vertical bars for each *quartet* correspond, from left to right, to *dispersion*-corrected [DD] [23] interaction energies evaluated with the 6-31G*, 6-31+G*, 6-311G(d,p) and 6-311+G(d,p) basis sets.

DISCUSSION

Relationship between biomolecular structure and quantum-mechanical non-covalent hACE2...S-RBD interactions. Interaction energies, as defined in Eqs. 2-3, can be positive or negative and are a measure of the tendency of two biomolecular structures to repel or attract each other, respectively. Within the present quantum biochemical framework interaction energies are the combined result of several physico-chemical effects, incorporated in Eqs. 1-3, some of which

TABLE III. | **S-RBD-centered *quartets* and their interaction energies [kcal/mol][a] with neighboring[b] hACE2 residues.**

Quartet	SARS-CoV-1 S-RBD Residues	E_{Int}^{DFT}	E_{Int}^{DD}	E_{Int}^{Total}
SQ1	ASN435-TYR436-ASN437-TYR438	−10.05	−3.40	−13.45
SQ2	LYS439-TYR440-LEU478-ASN479	+52.04	−4.09	+47.95
SQ3	PHE483-TYR484-THR485-THR486	+31.72	−20.90	+10.81
SQ4	PRO470-ALA471-LEU472-ASN473	+24.48	−8.99	+15.49
SQ5	THR425-ARG426-ASN427-ILE428	−55.25	−2.32	−57.57
SQ6	THR487-GLY488-TYR491-GLN492	−14.06	−28.08	−42.15
SQ7	TYR442-LEU443-TYR475-TRP476	+18.51	−17.02	+1.48

[a] DFT energies computed at 6-311+G(d,p)/B3LYP level; *Dispersion* (DD) corrections evaluated with semiempirical method. [23]

[b] All ACE2 residues within 4.5 Å of each S-RBD *quartet* were included.

are intrinsically attractive whereas others are repulsive. For example, intermolecular Coulomb interactions between atoms whose charge has the same(different) sign are repulsive(attractive), respectively, whereas intermolecular *dispersion* van der Waals forces are additively attractive. *Dispersion* forces correspond to the attractive portion of intermolecular van der Waals potentials [23] and were carefully evaluated and incorporated in this work.

The structural details, at the molecular level, of host-virus interfaces are crucial for determining the strength and relative importance of the various types of intermolecular forces since these are dependent on different powers of interatomic distances (r_{ij}). For example, Coulomb interaction energies between two atomic centers i and j, separated by a distance r_{ij}, scale as $\frac{1}{r_{ij}}$. By contrast, at short intermolecular distances (i.e. the nonretarded regime), attractive *dispersion* contributions to van der Waals energies scale inversely to the sixth power ($\frac{1}{r_{ij}^6}$) of the distances. [23, 29, 30] Thus, the relative importance of each type of noncovalent intermolecular interaction is highly dependent on intermolecular distances with Coulomb interactions being longer range and *dispersion* interactions playing critical roles at shorter ranges. In this work we focus on evaluation of host-virus interactions corresponding to the intermolecular distances of the non-covalently bound hACE2...S-RBD sructure determined by crystallography. [3] That is, we focus on key hACE2...S-RBD interactions at the crucial structural, as opposed to temporal, stage when hACE2 has formed, upon completion of a process of intermolecular recognition, a thermodynamically favorable noncovalent complex with the prefusion conformation of the SARS-CoV-1 spike protein.

Particularly important attractive residues at the hACE2...S-RBD interface. The parallel evaluation of two sets of interaction energies, hACE2-centered *quartets* interacting with S-RBD and S-RBD-centered *quartets* interacting with hACE2, allows the identification of contact residues of particular importance to the host-virus binding energy. Tables II and III provide complementary information and suggest a number of residues which dominate the hACE...S-RBD attractive energy. Most hACE2 residues belonging to *quartet* **AQ3** (GLU329, LYS353 and GLY354) and all hACE2 residues making up *quartet* **AQ4** (GLU37, ASP38, TYR41 and GLN42) are involved in significant attractive interactions as determined by both, hACE2-centered and S-RBD-centered, energetic calculations. Similarly, residue TYR436 from S-RBD *quartet* **SQ1**, residue ARG426 from S-RBD *quartet* **SQ5** and most residues from S-RBD *quartet* **SQ6** (THR487, GLY488 and TYR491) are likely primary attractors, with respect to hACE2, based on a similar analysis.

Some of the previous results are consistent not only with available crystallographic data but also with functional and substitutional studies. For example the strong (salt bridge) interaction between hACE2(GLU329) and S-RBD(ARG426) has been noticed [22] from structural analysis whereas the importance, for hACE2 binding, of S-RBD residues ARG426 and THR487 was suggested from mutation substitutional studies. [24] In addition, S-RBD residue TYR484 has been postulated as an important hACE2 binder. [3, 24] In this work this residue is part of S-RBD *quartet* **SQ3** which produces a net weak repulsion relative to hACE2. However Table III shows that, due to the presence of its phenolic group, TYR484 likely contributes an attractive *dispersion* interaction consistent with the ≈ -29.90 kcal/mol *dispersion* energy of the entire *quartet*. Thus, this residue can potentially be an important attractor even though the evidence in the present study is somewhat indirect.

CONCLUSION

SARS-CoV-1 is a potentially-reemerging [1, 2] highly-pathogenic virus and substantial gaps remain in our understanding of its molecular-level mechanisms of transmissibility. [2] Spike proteins of coronaviruses interact, *via* their *receptor binding domains*, with human ACE2 receptors. The identification of protein fragments, at the hACE2...S-RBD interface, which are primarily responsible for close-range attractive or repulsive interactions is of importance i) fundamentally for elucidating the physico-chemical origin of host-virus attachment and ii) for identifying specific therapeutic

targets and viral epitopes. Among the various anti-coronavirus therapeutic strategies there are two which may, in particular, benefit from this study. Namely, therapies which target the human ACE2 receptor and therapies which attempt to block SARS-CoV-1 spike proteins. The present studies, complementary to those based on X-ray crystallography or cryo-EM, have uniquely identified which protein fragments, herein referred to as residue *quartets*, are involved in the strongest, attractive or repulsive, hACE2...S-RBD interactions. The dominant residue fragments of attractive nature are shown in Fig. 1.

Our results are based on the three-dimensional biomolecular structures of the human ACE2 receptor and the SARS-CoV-1 spike protein. Thus, the present identification of specific, attractive and repulsive, biomolecular fragments as well as the quantification of their interaction energies is particular to this system, namely hACE2 interacting with the prefusion conformation of the SARS-CoV-1 spike protein. Our results also suggest possible interaction mechanisms of hACE2 with other similar, but not structurally identical, spike protein RBDs such as those from SARS-CoV-2. The fact that the sequence identity of the *domains* (S-RBD) from SARS-CoV-1 and SARS-CoV-2 is about 72-73$\%$ whereas the identity of their *motifs* (S-RBM) is only about 48$\%$ [24, 31] suggests similarities as well as differences in the relative importance of their specific amino acid residues towards hACE2 binding energies. This would be consistent with structural differences between their respective S-RBM and their non-identical binding affinities towards hACE2. [7] Interaction studies of hACE2 with SARS-CoV-2 must take into account the sequence and structural details of its own S-RBD. Indeed, some key hACE2-interacting S-RBD residues in SARS-CoV-1 may not play an equivalent role in SARS-CoV-2. [32] Interaction energy studies for SARS-CoV-2 to determine similarities as well as differences in hACE2...S-RBD binding, relative to SARS-CoV-1, are needed and are currently in progress in our laboratory.

REFERENCES

[1] He, Y. *et al.* Receptor-binding domain of SARS-COV spike protein induces highly potent neutralizing antibodies: implication for developing subunit vaccine. *Biochemical and Biophysical Research Communications* **324**, 773 – 781 (2004).

[2] Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y. & Yuen, K. Y. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews* **20**, 660–694 (2007).

[3] Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).

[4] Masters, S. & Perlman, S. Coronaviridae. In Fields, B. N., Knipe, D. M. & Howley, P. M. (eds.) *Fields Virology*, chap. 28 (Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2013).

[5] Simmons, G., Zmora, P., Gierer, S., Heurich, A. & Pohlmann, S. Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. *Antiviral Research* **100**, 605 – 614 (2013).

[6] Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet Infectious Diseases* **20**, e238–e244 (2020).

[7] Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

[8] Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virology Journal* **16**, 69 (2019).

[9] Beniac, D. R., Andonov, A., Grudeski, E. & Booth, T. F. Architecture of the SARS coronavirus prefusion spike. *Nature Structural & Molecular Biology* **13**, 751–752 (2006).

[10] Belouzard, S., Millet, J. K., Licitra, B. N. & Whittaker, G. R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **4**, 1011–1033 (2012).

[11] Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).

[12] Godet, M., Grosclaude, J., Delmas, B. & Laude, H. Major receptor-binding and neutralization

362 determinants are located within the same domain of the transmissible gastroenteritis virus
363 (coronavirus) spike protein. *Journal of Virology* **68**, 8008–8016 (1994).
364 [13] Gui, M. *et al.* Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal
365 a prerequisite conformational state for receptor binding. *Cell Research* **27**, 119–129 (2017).
366 [14] Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus.
367 *Nature* **426**, 450–454 (2003).
368 [15] Hoffmann, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked
369 by a clinically proven protease inhibitor. *Cell* **181**, 271 – 280.e8 (2020).
370 [16] Walls, A. C. *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein.
371 *Cell* **181**, 281 – 292.e6 (2020).
372 [17] Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage
373 for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology* **5**, 562–569
374 (2020).
375 [18] Fields, B. N., Knipe, D. M. & Howley, P. M. *Fields virology* (Wolters Kluwer Health/Lippincott
376 Williams & Wilkins, Philadelphia, 2013).
377 [19] Reguera, J., Mudgal, G., Santiago, C. & Casasnovas, J. M. A structural view of coronavirus-
378 receptor interactions. *Virus Research* **194**, 3 – 15 (2014).
379 [20] Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike
380 glycoprotein in complex with its host cell receptor ACE2. *PLOS Pathogens* **14**, 1–19 (2018).
381 [21] Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2.
382 *Science* **367**, 1444–1448 (2020).
383 [22] Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–
384 224 (2020).
385 [23] Deligkaris, C. & Rodriguez, J. H. Correction to DFT interaction energies by an empirical
386 dispersion term valid for a range of intermolecular distances. *Phys. Chem. Chem. Phys.* **14**,
387 3414–3424 (2012).
388 [24] Yi, C. *et al.* Key residues of the receptor binding motif in the spike protein of SARS-CoV-2
389 that interact with ACE2 and neutralizing antibodies. *Cellular & Molecular Immunology* **17**,
390 621–630 (2020).
391 [25] Renn, A., Fu, Y., Hu, X., Hall, M. D. & Simeonov, A. Fruitful neutralizing antibody pipeline
392 brings hope to defeat SARS-Cov-2. *Trends in Pharmacological Sciences* (2020).

[26] Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *Journal of Virology* **89**, 1954–1964 (2015).

[27] Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).

[28] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).

[29] Roth, C., Neal, B. & Lenhoff, A. Van der Waals interactions involving proteins. *Biophysical Journal* **70**, 977 – 987 (1996).

[30] Lund, M. & Jönsson, B. A mesoscopic model for protein-protein interactions in solution. *Biophysical Journal* **85**, 2940 – 2947 (2003).

[31] Chen, Y., Guo, Y., Pan, Y. & Zhao, Z. J. Structure analysis of the receptor binding of 2019-nCoV. *Biochemical and Biophysical Research Communications* **525**, 135 – 140 (2020).

[32] Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565 – 574 (2020).

[33] Cossi, M., Rega, N., Scalmani, G. & Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comp. Chem.* **24**, 669–681 (2003).

[34] Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).

METHODS

The biomolecular structure of the hACE2...S-RBD interface corresponding to the SARS-CoV-1 virus, as extracted from the published X-ray crystallographic structure, [35] was studied as a single structure and also separated into *quartet*-based fragments as described in the main text. A locally developed algorithm was used to divide the interacting hACE2...S-RBD molecular structure into *quartet* fragments. It was determined that, either hACE2-centered or S-RBD-centered fragments composed of at least four residues was necessary to evaluate fragment-based interaction energies. Fragments of smaller size, i.e. containing less than four residues, did not include a minimum of nearest neighbor and next nearest neighbor interactions to provide reliable qualitative estimates of *partial* intermolecular interaction energies.

All electron Khon-Sham density functional calculations were done on the overall structure in both, gas and solvent, phases. Similar calculations were done on all host-virus biomolecular fragments which in the main text are referred to as *supermolecular fragments*. Khon-Sham density functional calculations solve, numerically, a quantum mechanical Hamiltonian that includes an approximation to the exact, but unknown, exchange-correlation potential. Energies were obtained, in the low temperature limit, in terms of Eq. 1 for all biomolecular structures described in the text.

$$\begin{aligned} \mathrm{E}^{\mathrm{DFT}} = \mathrm{KE}[\rho(\mathbf{r})] + \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\ + \mathrm{E}_{\mathrm{xc}}[\rho(\mathbf{r})] + \int \mathrm{v}(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} + \mathrm{E}_{\mathrm{NN}} \end{aligned} \tag{1}$$

Here, $\rho(\mathbf{r})$ represents the electron density obtained from solution of the Khon-Sham equations. The B3LYP [36-37] exchange-correlation functional was used in the energy calculations due to its complementarity with the B3LYP-DD *dispersion*-correction methodology. [38] Many exchange-correlation functionals, including B3LYP, fail to properly account for intermolecular *dispersion* van der Waals contributions. Therefore, semiempirical corrections ($\mathrm{E}_{\mathrm{Int}}^{\mathrm{DD}}$) were added to the Khon-Sham interaction calculations *via* the B3LYP-DD methodology [38] which fairly accurately incorporates *dispersion* for a range of intermolecular distances. As reported in the main text and the Supplementary Tables, several basis sets of progressively increasing size [including 6-31G*, 6-31+G*, 6-311G(d,p) and 6-311+G(d,p)] were used in a series of independent energy calculations to ensure qualitative consistency of the numerical results. Interaction energies were computed in the absense ($\mathrm{E}_{\mathrm{Int}}^{\mathrm{DFT}}$) and presence ($\mathrm{E}_{\mathrm{Int}}^{\mathrm{DFT-DD}}$) of *dispersion*, *via* Eqs. 2-3, following the super-

molecular approach as described in the B3LYP-DD reference. [38]

$$\mathrm{E}_{\mathrm{Int}}^{\mathrm{DFT}} = \mathrm{E}_{\mathrm{hACE2...S-RBD}}^{\mathrm{DFT}} - \mathrm{E}_{\mathrm{hACE2}}^{\mathrm{DFT}} - \mathrm{E}_{\mathrm{S-RBD}}^{\mathrm{DFT}} \tag{2}$$

$$\mathrm{E}_{\mathrm{Int}}^{\mathrm{DFT-DD}} = \mathrm{E}_{\mathrm{Int}}^{\mathrm{DFT}} + \mathrm{E}_{\mathrm{Int}}^{\mathrm{DD}} \tag{3}$$

REFERENCES

[35] Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science **309**, 1864–1868 (2005).

[36] Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. J. Chem. Phys. **98**, 1372–1377 (1993).

[37] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. **98**, 5648–5652 1993).

[38] Deligkaris, C. & Rodriguez, J. H. Correction to DFT interaction energies by an empirical dispersion term valid for a range of intermolecular distances. Phys. Chem. Chem. Phys. **14**, 3414–3424 (2012).

ACKNOWLEDGEMENTS

The use of high performance computational resources from the Purdue University ITaP center is gratefully acknowledged.

AUTHOR CONTRIBUTIONS

J.H.R was responsible for the overall design of the project and carried out all electronic structure and numerical calculations. J.H.R also wrote the complete mansucript. A.G. implemented and executed the computational algorithm to create protein residue fragments based on a previously published X-ray crystallographic structure.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https:/doi.org/

Correspondence and requests for materials should be addressed to J.H.R.

Figures

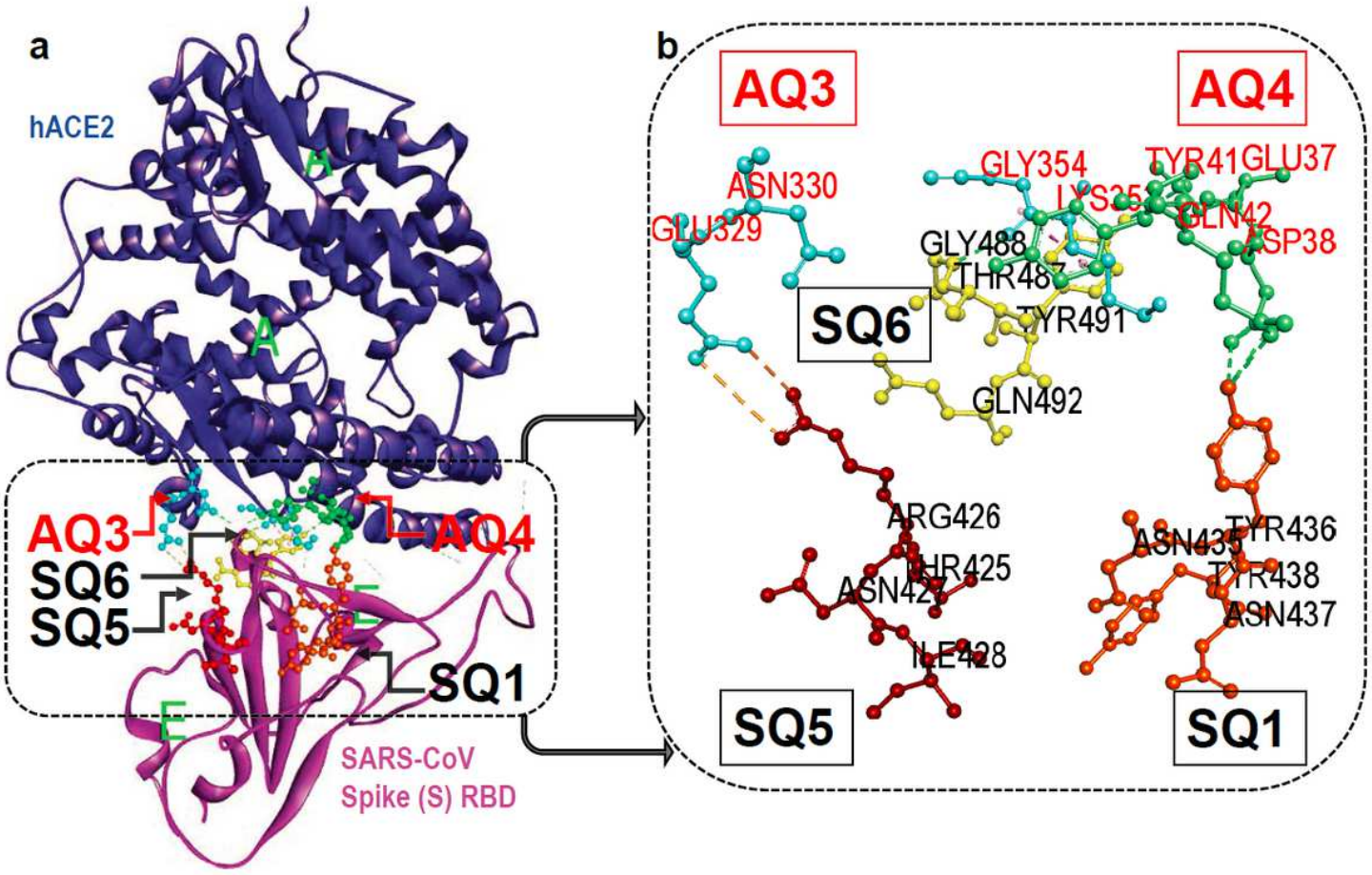

Figure 1

Identification of four-residue fragments (i.e. quartets) which produce attractive interaction energies between hACE2 and the SARS-CoV-1 S-RBD. a, Structure of hACE2 receptor (Chain A) in complex with SARS-CoV-1 spike protein (Chain E). [3] The key quartets, at the hACE2...S-RBD interface, promoting host-virus binding are shown in the dashed box. b, Magnified view of the hACE2 (AQ3, AQ4) and S-RBD (SQ1, SQ5, SQ6) residue quartets which mostly contribute to the attractive hACE2...S-RBD interaction energy (shown in ball and stick).

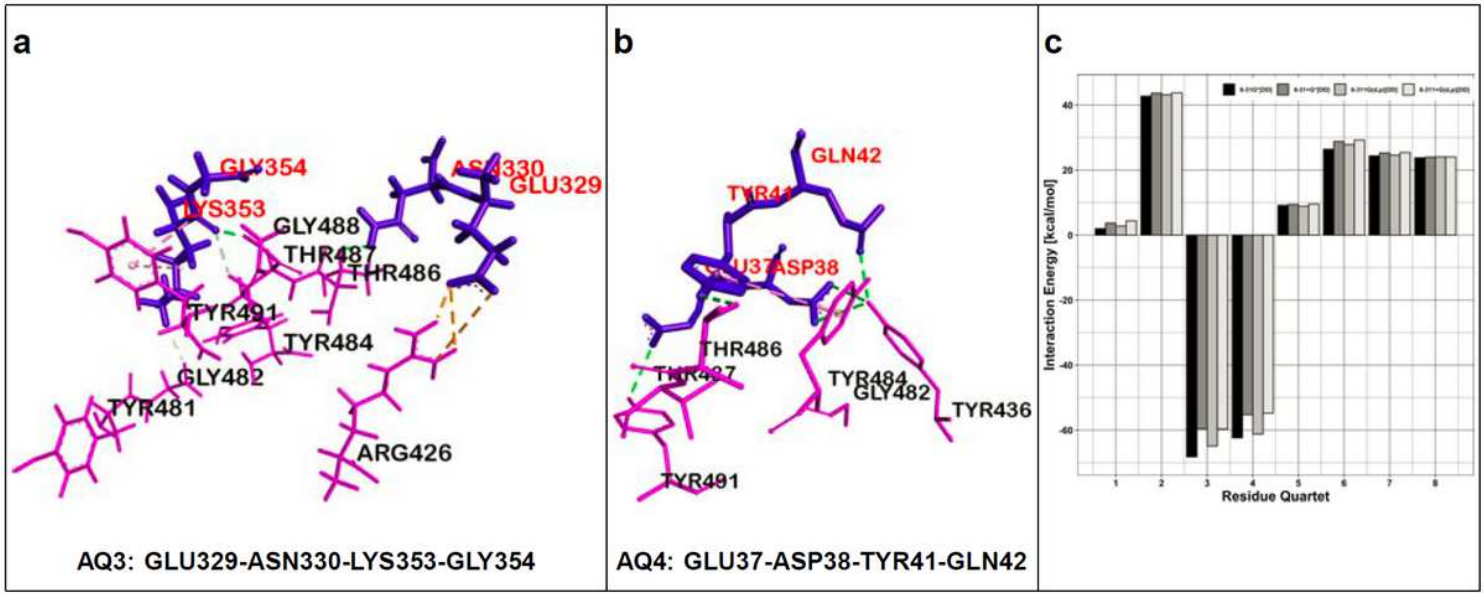

Figure 2

The two hACE2-centered fragments producing a net attractive interaction towards S-RBD. a-b,ACE2 quartet residues (shown in blue) and neighboring S-RBD residues (shown in 158 pink) corresponding to the dominant attractive ACE2...S-RBD interactions. ACE2 is the structural 159 reference. c, Main repulsive (positive) and attractive (negative) interactions [kcal/mol] between 160 quartets of the human ACE2 receptor, used as structural references, and neighboring residues of 161 the receptor binding domain of the SARS corona virus spike protein (S-RBD). The four adjacent 162 vertical bars for each quartet correspond, from left to right, to dispersion-corrected [DD] [23] 163 interaction energies evaluated with the 6-31G*, 6-31+G*, 6-311G(d,p) and 6-311+G(d,p) basis 164 sets.

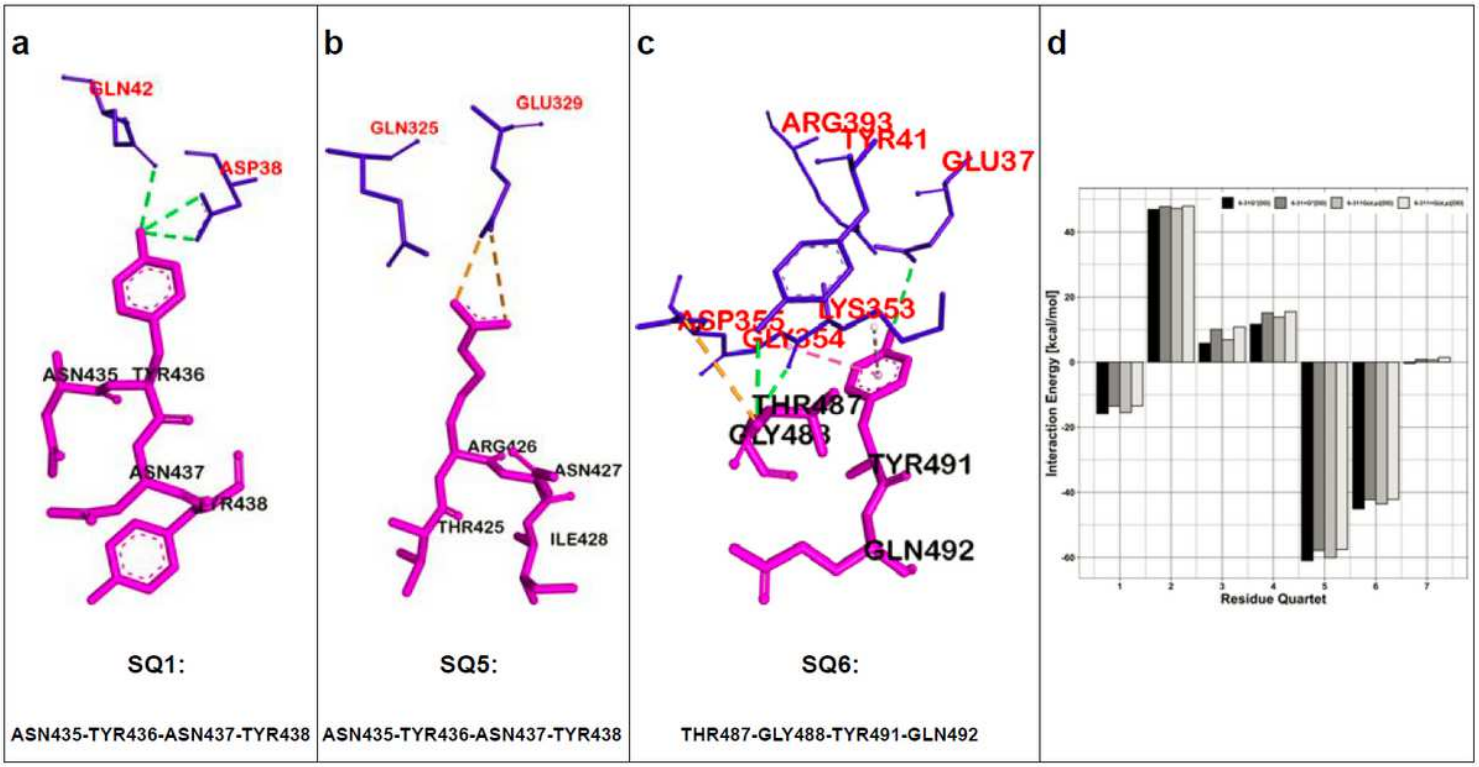

Figure 3

The three S-RBD-centered fragments producing a net attractive interaction to-250 wards hACE2. a-c,S-RBDquartet residues (shown in pink) and neighboring ACE2 residues 251 (shown in blue) corresponding to the dominant attractive ACE2...S-RBD interactions. S-RBD 252 is the structural reference. d, Main repulsive (positive) and attractive (negative) interactions 253 [kcal/mol] between quartets of the SARS-CoV-1 S-RBD, used as structural references, and neigh-254 boring residues of the human hACE2 receptor. The four adjacent vertical bars for each quartet 255 correspond, from left to right, to dispersion-corrected [DD] [23] interaction energies evaluated with 256 the 6-31G*, 6-31+G*, 6-311G(d,p) and 6-311+G(d,p) basis sets.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RodriguezNatureSBOctober2020Supplementary.pdf