

***Ab initio* electronic dielectric “constant” of proteins: a baseline for electrostatic interaction in biomolecular systems**

Puja Adhikari¹, Rudolf Podgornik^{2,3,4,5}, Bahaa Jawad¹, Wai-Yim Ching¹

1. Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City Missouri, USA

2. School of Physical Sciences and Kavli Institute of Theoretical Science, University of Chinese Academy of Sciences, Beijing 100049, China

3. CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100090, China

4. Wenzhou Institute of the University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325000, China.

5. Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia

KEYWORDS: *Electrostatic interaction, dielectric constant, proteins, explicit ab initio modeling, random phase approximation.*

ABSTRACT:

The protein dielectric constant reflects the molecular heterogeneity of the proteins and can be decomposed into different components depending on the size, structure, composition, locality, and environment of the protein in general. The long history of its computation and measurement attest to the vital importance of electrostatic interactions in protein physics that engendered diverse theoretical approaches based often on scattered methodologies with various adjustable parameters. We present a new robust computational method anchored in rigorous *ab initio* quantum mechanical calculation of explicit atomistic models, without any indeterminate parameters to compute and gain insight into the *electronic component* of the static dielectric constants of small proteins under different conditions. We implement the new methodology to the 20 canonical amino acids individually, a polypeptide RGD-4C (1FUV) in different environments, and the SD1 domain in the Spike protein of SARS-COV-2. The calculated electronic dielectric constants for 1FUV and SD1 in vacuum are 28.06 and 50.02 respectively. They decrease in the presence of aqueous bathing solution.

1. INTRODUCTION

The dielectric properties and specifically the static *dielectric constant* of proteins in aqueous solutions have been studied for several decades [1] and play a crucial role in protein electrostatics,

specifically in the protein-protein and protein-nucleic acid interactions, folding pathways, solution behavior and stability [2]. Quite recently electrostatic contributions to the binding free energy have been invoked as a possible source of the differences between the binding pattern of the spike proteins of SARS-CoV-2 and SARS-CoV to the ACE2 human receptor [3] as well as the principal interaction component between the SARS-CoV-2 virus and the electret fibers in personal protective gear [4]. The problem of the macroscopic protein dielectric “constant” is exacerbated by the fact that, as in other biomolecular systems such as lipid membrane bilayers composed of different segregated molecular components, the underlying microscopic dielectric properties are heterogeneous [5] and the continuum assumptions become to say the least, problematic on a molecular level [6]. Consequently, the protein dielectric constant appears more as a phenomenological parameter, or even a phenomenological function of the position inside the protein core [7], that depends on the model used to describe the heterogeneous molecular structure of proteins, then a single universal constant grounded in some fundamental theory [8]. Therefore, rather than referring to a single dielectric constant Warshel and coworkers suggest that the basic assumptions of different models should be listed explicitly so that it is clear which part of the dielectric response is taken into account explicitly and which is just a phenomenological parameter that can be fitted from experiments [1]. Breaking down the various mechanisms that enter the dielectric response of proteins [9], a model which explicitly includes everything could start with the protein dielectric constant as 1, a value which would progressively grow as more atomic/molecular mechanisms are explicitly included, starting with electronic degrees of freedom, atomic/ionic degrees of freedom, induced multipoles, permanent multipoles, internal and vicinal solvating water molecules, polar group protonation/deprotonation fluctuations and the bathing medium solvation effect where the dielectric constant would saturate at its bulk water value of 80 [10].

Realizing the importance of electrostatic interaction in biomolecular systems and after early suggestions in the literature [11-13], it has become an accepted wisdom that in order to have a good fit for electrostatic properties of proteins, such as the pK_a of different dissociable amino acids, redox potentials and ligand binding energies to name just a few, one needs to consider a rather high value of the protein dielectric constant of 20 [14], or even as high as 40, instead of 2-4 as would be on first sight perhaps more plausible for the very non-polar protein interior with only electronic and atomic degrees of freedom [2]. These are important issues as many continuum computational approaches [15] as well as simulation methodologies to protein electrostatics [16] hinge crucially on the numerical choice for the dielectric constant of the protein interior, sometimes not only quantitatively but in many cases even qualitatively [17]. One can argue that the jump from 4 to 40 can be possibly due to configurational fluctuations of polar side chains and/or protonation/deprotonation fluctuations of the dissociable amino acid groups that would thus provide the “missing mass”, decreasing the dielectric inhomogeneity between the protein interior and the external aqueous solution or even completely eliminate it [18, 19]. While these contributions certainly need to be included, one would also like to know what is the *fundamental gauge* for the dielectric constant of proteins, stemming purely from the electronic degrees of freedom, and what part of the presumed anomalously large dielectric function of the protein interior remains to be explained by other mechanisms. It thus seems of some importance to *obtain a proper baseline* for the dielectric constant of proteins by calculating as accurately as possible its electronic part, so that higher level coarse grained or molecular models can be based on a solid number.

Motivated by the past success in calculating the electronic part of the dielectric function at full spectral resolution for many crystals and amorphous solids, we propose an *ab initio* quantum mechanical (QM)-based computational scheme, anchored in random phase approximation (RPA) [20] for the optical transitions from occupied to unoccupied states with explicit inclusion of the dipole transition matrix from the *ab initio* wave functions. We use this full spectral resolution of the dielectric response of proteinaceous molecules to calculate the static value of their dielectric constant. We label our method as quantum mechanical random phase approximation (QMRPA) method, that has been originally applied to inorganic crystals of infinite extension based on the Bloch Theorem [21], the cornerstone of the electronic structure theory of condensed matter physics. There are of course major differences between the standard implementation of this methodology in solid state physics and its application to biomolecules, specifically to proteins, which are due to their finite in size, the fact that they possess no periodically replicated unit cell and finally that they are usually immersed in a bathing aqueous medium of its own molecular structure.

Here, we implement the *ab initio* QMRPA approach to obtain the *electronic part of the dielectric constant* applicable to different biomolecules such as amino acids, polypeptides, and proteins. We calculate the static dielectric constant of the 20 dry canonical amino acids, a small RGD (1FUV) polypeptide in different environments, and the subdomain SD1 of the SARS-CoV-2 spike protein [22]. Our calculations, that are free of adjustable parameters and scalable for larger biomolecules, show that the relatively high dielectric constant that has been variously postulated for the protein interior is *grosso modo* consistent with purely electronic contribution, with other mechanisms providing just additional numerical variations on this universal background. While some *ab initio* calculations of molecular polarizability in response to external electric field have been attempted, mostly such calculations are focused restrictively to very small molecules such as amino acids (AAs) [23-25], and may include electronic as well as vibrational effects. To our knowledge, *ab initio* calculation has not been attempted to small proteins consisting of different numbers of AAs which are of course far more challenging [26].

2. RESULTS

The QMRPA method for inorganic crystalline and amorphous materials has been successfully used by us over the last 30 years [27]. Notable examples are the timely calculation of dielectric functions in the two exceptional materials, the high temperature ceramic superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7.8}$ (YBCO) in 1987 [28, 29] and the Buckminster fullerene C_{60} FCC crystals in 1991 [30, 31]. More recent examples of complex materials include mixed inorganic glass $(\text{a-SiO}_2)_{1-x}(\text{GeO}_2)_x$ [32] and amorphous zeolitic imidazolate framework (a-ZIF) showing the metal-insulator transition under compression [33, 34].

In our group, the calculation of dielectric functions on biomaterials started nearly 20 years ago but was not focused on the dielectric constants itself. For example, the calculation of optical transition in vitamin B₁₂ Adenosyl-cobalamin has shown good agreement with experimental spectra [35, 36]. Similar calculations of the electronic structure and X-ray absorption spectra were carried out on hydroxyapatite and different calcium apatite crystals [37, 38]. A particularly interesting example is the calculation of imaginary part of the dielectric function $\epsilon_2(\omega)$ of herapathite, a large complex

dichroic crystal $(C_{20}H_{24}N_2O_2H_2)_4 \cdot C_2H_4O_2 \cdot 3SO_4 \cdot 2I_3 \cdot 6H_2O$ [39], verifying the strong linear dichroism with anisotropy factor of 385 that has numerous applications for its polarizing properties [40]. Other examples include application to long range interactions based on dielectric spectra for biomolecular systems [41-43] and the effect of optical absorption of cytochrome c (Cyt c) macromolecules at the interface or surface for high performance photodetection [44, 45]. Last but not the least is the special peptide (RGD-4C) that will be highlighted in later sections. In the present paper, we present the results of using QMRPA method for proteins starting with 20 canonical AAs in the following subsection.

2.1 Twenty amino acids

Amino acids (AAs) are the basic structural units in biomolecular protein systems tantamount to atoms in crystalline solids. As of 1983, of the several hundred naturally occurring amino acids only 20 canonical amino acids appear in the genetic code. Any discussion of the structure, properties, and functionalities of polypeptide systems must therefore start with the 20 AAs [46]. (see **Fig.S1**)

We used the QMRPA method to calculate the electronic static dielectric constants $\epsilon_1(0)$ for the 20 AAs as for other types of materials that we already calculated in the past [27-38, 40-45]. We used a dry model for each AA individually, where all the atoms are fixed after optimization, and these results are used solely to illustrate the efficacy of the systematic computational modeling using QMRPA for a fixed set of biomolecules. Dry AAs cannot be used to gain insight into the actual behaviors of AAs in the 3D protein structure. **Fig. S2** shows the calculated total density of states (DOS) of the canonical AAs and **Fig. S3** shows the corresponding complex dielectric spectra $\epsilon(\omega) = \epsilon_1(\omega) + i\epsilon_2(\omega)$. The imaginary parts of $\epsilon(\omega)$ spectra, or $\epsilon_2(\omega)$, are calculated first and $\epsilon_1(\omega)$ is deconvoluted from $\epsilon_2(\omega)$ via the Kramers-Kronig transformation (see Eq. (6) in **Method Section D**). $\epsilon_2(\omega)$ spectra are very complex with multiple peaks of different shapes stemming from transitions from states below HOMO to those above LUMO. The complex character in $\epsilon_2(\omega)$ is rooted in the transition from the occupied molecular level below HOMO to the unoccupied levels above LUMO (see **Fig. S2**). An important characteristic of $\epsilon_2(\omega)$ is that it can be roughly divided into two features: (i) the more broadened peak centered around 15 eV and (ii) a series of scattered peaks at lower frequencies, whose relative strength and actual locations then determine the final dielectric constant $\epsilon_1(0)$ (see **Method Section D**). The prevalence of the type (ii) features relative to those from type (i) component will result in a decrease of $\epsilon_1(0)$. Such complex and intermingled characteristics of the dielectric spectrum originate in quantum characteristic processes determined by the wave functions of each molecular orbital and as such are impossible to parameterize. More specific dielectric spectra for the selected AAs are described in the caption of **Fig. S3**.

The values of the electronic part of the static dielectric constant, $\epsilon_1(0)$, for the 20 canonical AAs are presented in **Fig. 1** and listed in **Table S1**. Concurrently, we also show the pertinent results from the literature for comparison, either based on explicit calculation of the static dielectric constant [18, 19, 23] or those obtained from the mean dipole polarizabilities, using different approach and methods also claimed to be *ab initio* [23-25]. Our calculated static dielectric constants range from 39.32 in Trp to 143.06 in Gly and can be roughly divided into 3 groups: (1)

$\epsilon_1(0) < 50$: (Met, Phe, Trp, Tyr); (2) $50 < \epsilon_1(0) < 85$; (Arg, Asn, Asp, Cys, Gln, Glu, His, Ile, Leu, Lys, Pro, Thr, Val) and (3) $\epsilon_1(0) > 85$: (Ala, Gly, Ser), the average value for the 20 AAs is 70.69. The values obtained by the QMRPA methodology are comparable to those available in the literature [18, 19, 23-25, 47, 48] for 8 out of 20 AAs (Ala, Arg, Gly, Met, Phe, Ser, Trp, Tyr) being moderately to substantially different for the rest.

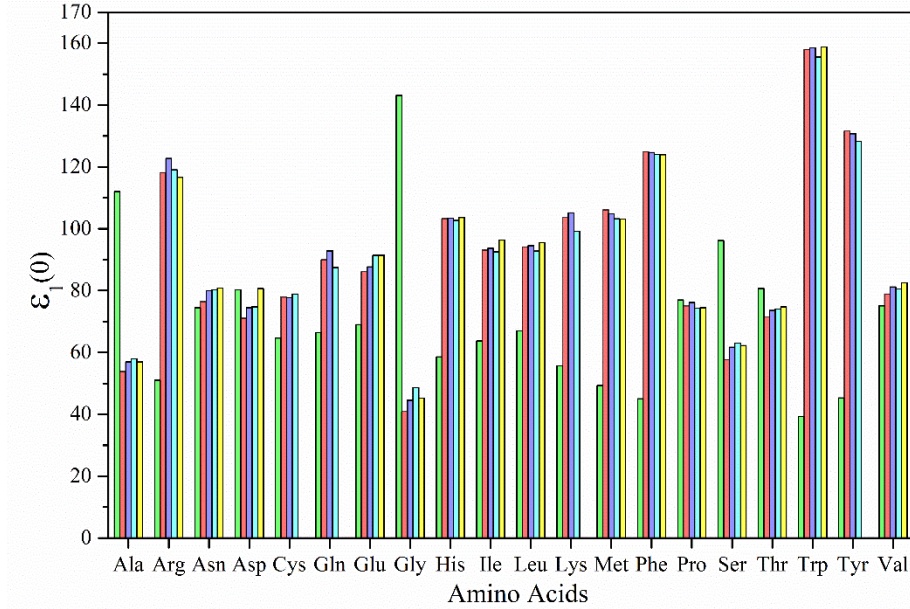


Fig. 1: Comparing dielectric constant $\epsilon_1(0)$ values for 20 AAs. Green: calculated, red: Swart *et al* [24], blue: Millefiori *et al* [23], cyan: Dos *et al* [48], yellow: Khanarian *et al* [47].

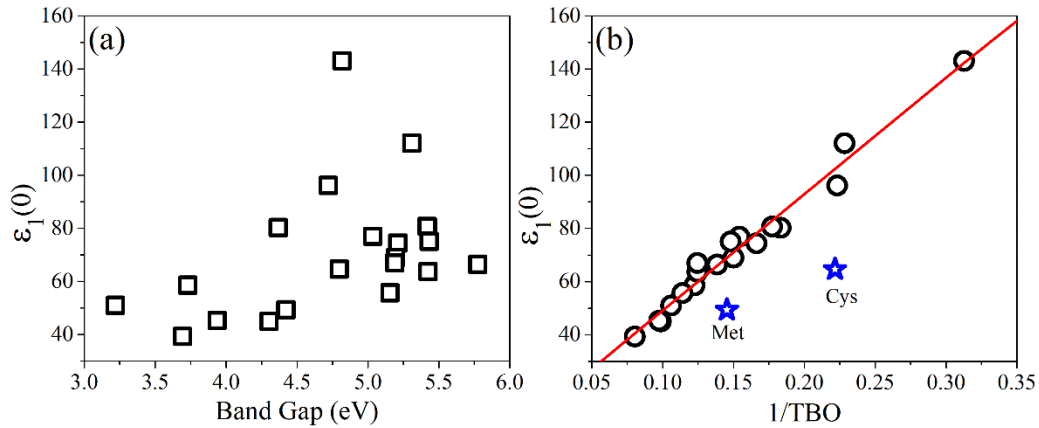


Fig. 2: Correlation plot of calculated electronic dielectric constant $\epsilon_1(0)$, for 20 AAs vs: (a) Band gap and (b) total bond order (TBO).

Fig. 2 shows the correlation plots for the calculated dielectric constant with the band gap and the total bond order (TBO) values (See **Method Section C**). It is interesting to note that there is no correlation with the band gap but there is a near-perfect linear correlation with the inverse of TBO:

$\epsilon_1(0) = 0.28 + 400.47 \times (1/TBO)$ with R^2 value of 0.97 if the 2 outlying data points with anomalously small dielectric constant values Met and Cys are excluded. This is an important new finding showing that the quantum mechanically calculated bond order values (Eq. (2) in **Method section C**) in biomolecular system can predict the electronic portion of the dielectric constant. The reason that the Met and Cys are below the near-perfect linear fit can be traced to the presence of the Sulphur (S) atom. S is a larger atom with $3s^23p^2$ electronic configuration compared to oxygen (O) with the $2s^22p^2$ configuration. As will be clear from the discussion in the following section, the protein RGD (1FUV) peptide has a relatively low electronic dielectric constant since it contains 4 Cys AAs.

It should be made clear that we used a dry model in vapor for each of the AA individually, where all the atoms are fixed after optimization and therefore the results are used only to illustrate the efficacy of systematic computational modeling using QMRPA for a fixed set of biomolecules. They cannot be used to explain the actual behaviors of AAs in a 3D solvated protein structure for two reasons: first, in 3D protein structure, the hydrophobic AAs (Leu, Val, etc.) are often packed in the interior of the protein away from facing the bathing solution, while the polar or charged AAs are exposed to solvent, and second, each AA in the 3D structure interacts with non-vicinal AAs along the primary sequence. The environmental effect involving solvation and salt addition on small proteins will be presented later.

2.2 RGD (1FUV) peptide

Peptide sequences with arginine-glycine-aspartic acid (RGD) motif display a strong affinity and selectivity for a membrane protein called integrin, which is a key cell surface receptor mediating cell adhesion to extracellular matrices (ECM) [49, 50]. Integrins are actively expressed on vascular endothelial cell surfaces and play a key role in tumor metastasis, leukocyte migration, and angiogenesis [51], making them an ideal target for treating inflammatory diseases and cancer. Because integrin receptors recognize RGD as a primary sequence, RGD peptides are used to target cancer cells [52]. In addition, RGD peptides provide numerous applications in biological and biomedical devices, being frequently incorporated into biomaterials designed to facilitate wound healing [53], serving as candidates for radiotracers in imaging [54], being used in implantable medical devices [55], mimicking the activities of the ECM proteins in culture cells [56].

We use the QMRPA method to calculate the static electronic dielectric constant of a cyclic tripeptide RGD-4C to demonstrate its applicability for small protein. The initial structure of the RGD-4C was obtained from the Protein data bank (PDB ID: 1FUV) based on nuclear magnetic resonance (NMR) data [57]. The electronic structure, dielectric response, and the surface partial charge distribution of the RGD peptide has been previously studied by us [58], while here we present improved accuracy calculations of the static electronic dielectric constant $\epsilon_1(0)$. 1FUV has 11 AAs (ACDCRGDCFCG) of 6 different types: Gly (2), Cys (4), Phe (1), Asp (2), Arg (1), and Ala (1).

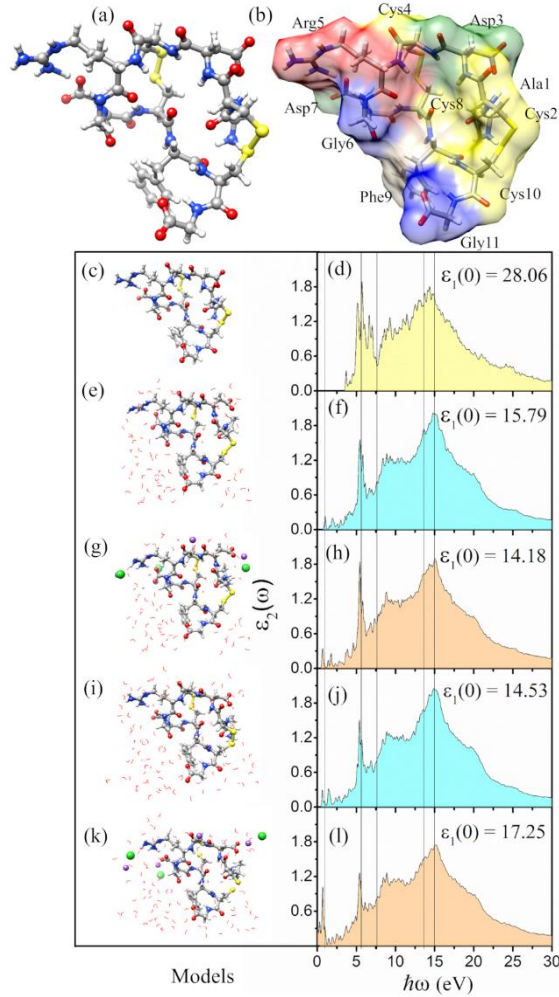


Fig 3: Dry, solvated, and salted models for 1FUV with their $\epsilon_2(\omega)$ spectra and $\epsilon_1(0)$ values. Ball and stick illustration of (a), (c) dry 1FUV with (b) surface shown with different colors. $\epsilon_2(\omega)$ for: (d) dry 1FUV; (f) 1FUV with 80 H₂O, (h) 1FUV with 80 H₂O and 0.15 NaCl, (j). 1FUV with 100 H₂O, and (l) 1FUV with 100 H₂O and 0.15 NaCl. The ball and stick figure for respective cases are in the left panel. Grey: C, red: O, blue: N, white: H, green: Cl, and purple: Na.

For its 135 atoms, including H atoms, it is somewhat clustered but otherwise a rather typical small peptide. It is one of the most used RGD variants because of the presence of 4 Cys residues which allow for the two rare disulfide bonds. **Fig. 3 (a), (b) and (c)** show ball and stick structure of 1FUV after VASP relaxation. The 11 AAs in 1FUV are marked in **Fig. 3 (b)**. There are considerable non-local interactions between non-nearest neighbor (NN) AAs in the primary sequence. This aspect of 1FUV will be further elaborated in **Section 3.2**. **Fig. 3 (d)** shows the final imaginary part $\epsilon_2(\omega)$

up to 30 eV for dry 1FUV and will be discussed again in the following section together with the spectra (**Fig. 3 (e), (f), (i), and (j)**) for the purely solvated and (**Fig. 3 (g), (h), (k), and (l)**) for salted models. The dielectric constant for dry 1FUV $\epsilon_1(0)$ is found to be 28.06. As can be seen, the general feature of $\epsilon_2(\omega)$ of 1FUV are like those of the AAs in **Fig. S2** with the broader peak at 14.4 eV and discernable multiple sharp peaks below 5.7 eV. Nevertheless, these peaks are much

less spiky than for most of the AAs. This also explains that the final value of the dielectric constant of dry 1FUV. We now proceed to the wet and/or salted models that are solvated by 80 and 100 water molecules with added 0.15 M NaCl (4 Na and 3 Cl ions) (see **Method section A**).

2.3 Modelling Specific local environmental effect in RGD (1FUV) peptide

For biomolecular systems and proteins in particular, environment effects such as the presence of solvent, fixed pH reservoir, fixed salt content and *protonation/deprotonation equilibrium* etc. must

ideally be considered. In the explicit modeling for QMRPA calculation, which is a single point calculation, a specific model for each of the environmental effects needs to be constructed individually with increased size and complexity. The construction of four solvated and salted models are described in the **Method section A**. The atomic structures of these four solvated models are further optimized using VASP and the final ball and stick sketch for these four models are displayed in left panel of **Fig. 3**. These 2D projections of the solvated structures of 1FUV clearly show that the water molecules are evenly distributed with random orientation and no penetration of water into the interior of the protein. The locations of the Na and Cl ions depend on the VASP optimization routines, corresponding to the most energetically favorable positions. There is no clear evidence of any protonation/deprotonation in these models but the formation of many HBs is obvious.

We now present the results for these four solvated and salted 1FUV models together with the dry models discussed previously in **Section 2.2**. The calculated electronic dielectric constants for the five 1FUV models are listed in **Table S2**, while **Fig. S4** shows the calculated DOS and PDOS for all the 5 models. For the dry 1FUV, the TDOS is also resolved into 11 different AAs. For example, the 4 S-containing Cys residues have separate peak in the LUMO which is not the case for individual Cys in **Fig. S2**. This feature comes from the interaction of Cys with the other AAs in 1FUV. We also observe that there are no so-called *gap states* between the HOMO-LUMO gap in the dried 1FUV model, unlike those in most of the isolated AAs where sporadic gap states do appear, especially close to the top of the HOMO. In **Fig. S4 (b) – (e)** for the solvated and salted models, the PDOS are resolved into those from 1FUV, the H₂O molecules and the salt ions. It is of interest to note that the TDOS do have many gap states but the PDOS can show the interaction between different AAs in solvated 1FUV, the individual H₂Os and the Na or Cl ions of the dissolved salts.

The calculated dielectric spectra $\epsilon_2(\omega)$ of the 4 solvated 1FUV models are shown in **Fig. 3(f), 3(h), 3(j), and 3(l)**. It can be seen that the general features in the solvated models are similar to the dry model in **Fig. 3 (d)** except for the appearance of many small transition peaks below 6 eV arising from the transitions from or to the gap states as evidenced in the DOS presented in **Fig.S4**. The increase in the smaller peaks below 6 eV results in the reduction of the electronic dielectric constant $\epsilon_1(0)$ (marked in **Fig. 3 (f), (h), 3(j), and 3(l)**). When the salts Na and Cl ions are added, the $\epsilon_2(\omega)$ and $\epsilon_1(\omega)$ spectra are changed even further in a more complex manner depending on the actual location of the gap states and the strength of their dipole transition matrix elements (Eq. (5) in **Methods section D**).

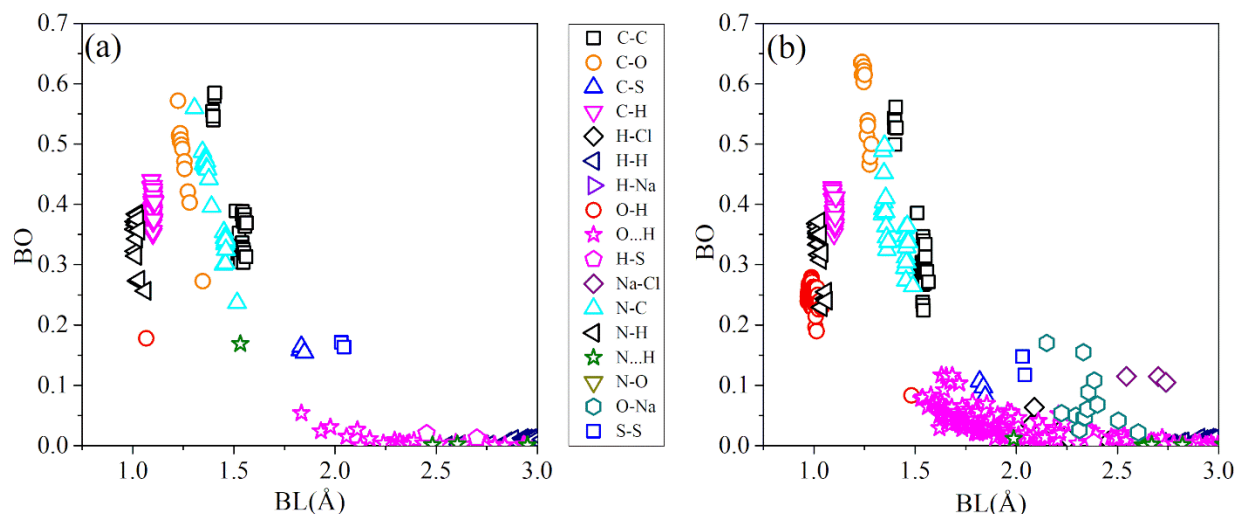


Fig. 4 Distribution of BO vs BL in the two 1FUV models: **(a)** dry 1FUV; **(b)** 1FUV with 100 H₂O and 0.15 NaCl

As stated earlier, there exists a strong correlation between the static dielectric constant and the TBO for the 20 AAs. The bond order (BO) for every pairs of atoms calculated from Eq. (2) in **Method section C**), is an integral part of the description of biomolecular materials [27, 59, 60]. **Fig. 4** shows BO versus bond length (BL) plots for the dry 1FUV model and the more complex solvated model with 100 H₂O molecules and salts. These bonds include those within each AA, between different AAs, and between atoms in each of the AA with H₂O molecules or salt ions in the solvated models. Similar plots for the other three models, solvated 1FUV with 80 and 100 water molecules and salted models in model with 80 H₂O molecules are shown in **Fig.S5**.

Here, we briefly describe the main features observed in these two models in **Fig. 4**:

(1) Dry 1FUV: As it is well known, the strongest bonds are the covalent bonds containing C (C-C, C-O, N-C, C-H). Their BL varies slightly but bond strength or BO values can vary rapidly,

scaling with BL. The BO depends on the nature of the covalent bond (single or double), the atoms it bonds and to a lesser extent, the local environment of the bonded pair. The covalent and C-H bonds within or between different AAs have BL slightly larger than 1.0 Å and exhibit varying BO values. In **Fig. 4(a)**, there is a singular O-H bond that can be traced to O in residue Asp7 and H in residue Arg5 which can be interpreted also as an exceptionally strong HB. The hydrogen bonds (HB) O...H or N...H are of vital importance in any biological system, being weaker than covalent bonds but ubiquitous especially in the presence of solvents. The BLs of HBs generally range between 1.5 Å to 3.0 Å and can have BO values larger than 0.1 e⁻ in some cases. **Fig. 4 (a)** shows the presence of a very strong N...H bond with BO of 0.016 e⁻ and BL of 1.532 Å with H and N, pertaining to the same residue Arg5. Other HBs are all O...H bonds with a much longer BL above 1.6 Å. The other important bonds are the two S-S bonds near 2.2 Å and short C-S covalent bonds at 1.75 Å involving S atom in Cys residues. These are all very stable bonds with BO values of 0.16 e⁻ in different environments.

(2) *1FUV with 100 H₂O and salt ions*: **Fig. 4 (b)** shows the BO vs. BL plots for the largest and more complicated 1FUV model with 100 H₂O molecules and 7 salt ions (4 Na and 3 Cl). The main difference is the appearance of the internal O-H covalent bond in H₂O and the new bonds (H-Cl, H-Na, Na-Cl and O-Cl) with the salt ions at much larger BL. Other changes are less remarkable but it is notable that the single O-H bond and the strong N \cdots H HB in **Fig. 4 (a)** have both disappeared, showing the changes in the atomic positions in residues Asp and Arg due to solvation. The appearance of many more O \cdots H HBs with shorter HB distances is expected. In addition, there are some minor changes in the N-C bonds while the C-S and S-S bonds become weaker but remain at similar separations. The C-H bonds retain similar separation since the AAs in 1FUV remain relatively intact. The number of O-H bonds is observed to increase with concomitant increase in the BO values. This could imply the possible occurrence of protonation when some H atoms in water molecules rearrange themselves under increased complexity of the environment although we do not yet have solid evidence for this interpretation.

The *ab initio* quantum mechanical calculation allows for a very detailed analysis of the interatomic bonding in dry as well as solvated 1FUV, directly affecting the electronic portion of the dielectric constant in proteins. Quantitative evaluation of the HB in solvated model is extremely important and universally agreed by all researchers in biomolecular science, yet, few have gone deep enough in search for quantitative details. Most of them are characterized qualitatively based on the distance of separations between H and O or N without quantitative value for their bond strength, especially those using classical molecular dynamic methodology.

2.4 SD1 of Spike-protein

The last example for the calculation of the electronic dielectric constant in proteins is a part of the Spike-protein in SARS-CoV-2 virus. The *ab initio* calculation of the electronic structure and interatomic bonding of this large biomolecular system has been already discussed [22]. The Chain A of the Spike protein (PDB ID: 6VSB) has seven structural domains: N-terminal domain (NTD), receptor binding domain (RBD), subdomain 1 (SD1), subdomain 2 (SD2), fusion peptide (FP), heptad repeat 1 with central helix (HR1-CH) and connector domain (CD) [61]. SD1 is the smallest of the seven structural domains having 24 AAs and 391 atoms in the dry environment. Unlike 1FUV, it has an elongated structure mimicking a very irregularly shaped mini-protein as shown in **Fig. 5 (a)-(c)**. It is therefore highly instructive to apply the QMRPA method to investigate its static electronic dielectric constant and dielectric spectra of SD1. Other information on the electronic structure and intra- and inter-atomic bonding of SD1 can be found in Ref. [22].

The solvation effects in SD1 have been addressed by adding 300 vicinal water molecules, as illustrated in **Fig. 5(e)**. The solvated structure is then fully optimized by using VASP in the same way as the solvated model of 1FUV. The calculated dielectric spectra for the dry and solvated SD1 are shown in **Fig. 5(d)** and **(f)** with the static dielectric constants $\epsilon_1(0)$ 50.02 and 29.05, respectively. By comparing the same two values 28.06 and 14.53 in the case of dry and solvated 1FUV models, shown in **Fig. 3(d)** and **(j)**, we conclude that the electronic dielectric constant in SD2 is about twice as large as in the 1FUV but the ratio of reduction due to solvation is similar.

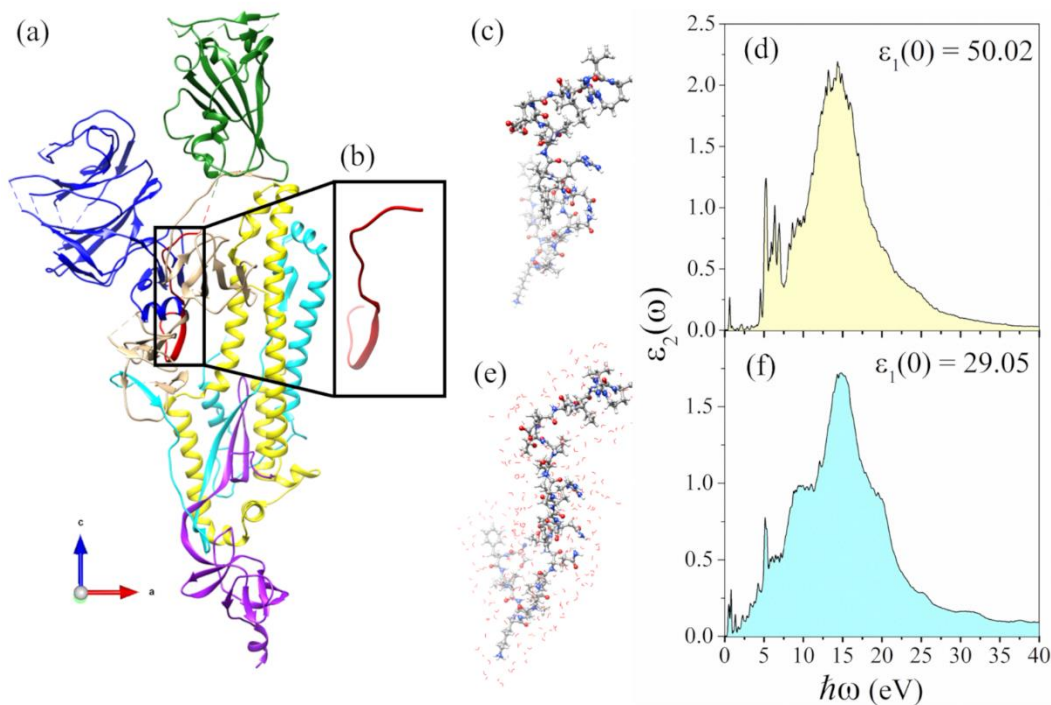


Fig. 5 (a) Chain A of Spike protein (6VSB), (b) ribbon for SD1, (c) ball & stick model of SD1, (d) $\epsilon_2(\omega)$ for dry SD1 model, (e) SD1 with dry model, and (f) $\epsilon_2(\omega)$ with 300 H₂O molecules.

3. DISCUSSION

There is growing interest in determining the optimum value of the protein dielectric constant to better understand and estimate its electrostatic properties. However, the dielectric constant is heterogeneous quantity or a function that depends on various factors such as the nature, size, and structure of the protein, the location where this constant is defined (interior, exposed surface, or interface), and so on. In the present study, we report the implementation of QMRPA scheme to calculate the static dielectric constants for various biomolecular systems including 20 canonical AAs, 1FUV in five different environments, and SD1 of SARS-CoV-2 Spike-protein in dry and solvated cases. The most interesting finding is that the electronic component of the static dielectric constants is inversely relation to total bond order (TBO) as shown in **Fig. 2**. In small proteins 1FUV and SD1, our results indicated that the electrostatic dielectric constants should be much higher than the values 2-4 that have been employed in many continuum approaches. Additionally, our calculated values revealed that electrostatic dielectric constants are reduced in the presence of bathing solutions (water and ions). These values may be used in classical MD simulation to compute the electrostatic interactions more accurately for larger biomolecular systems. Combining the present approach using *ab initio* calculations for the electronic dielectric constants in proteins with the classical MD simulations has important consequences for computational biomolecular science. Problems related to protein folding and stability, conformational change, mutation, glycosylation, etc. could be formulated on a more solid basis. However, the application of QMRPA methodology to a larger protein could be challenging due to computational limitations but are certainly not impossible. Also, a single dielectric “constant” of a heterogeneous protein is

meaningless since its value depends on the local environments. It is much more effective to use QMRPA method to a selected number of specific larger macromolecules or semi-macromolecules to provide a more accurate parameter-free data than just relying on intrinsic continuum level modeling.

3.1 Role of partial charge.

In addition to the static electronic dielectric constant, the partial charge (PC) distribution is also crucial for determining and elucidating the electrostatic effects that play major role in catalysis [62], drug engineering [63] etc. For different models studied here, the PC values are calculated based on the reliable orthogonalized linear combination of atomic orbitals (OLCAO) methodology [27] to understand the electrostatic effects but also the impact due to aqueous environments. We discuss here the comparison of PC distributions in the five 1FUV models under different local environment.

The calculated PCs on each of the 11 AAs in the five 1FUV models ((i) dry, (ii) solvated with 80 H₂O (iii) 100 H₂O, (iv) salted model of (ii), and (v) salted model of (iii)) are displayed in **Fig. 6**. It can be observed that the variations of PC in five 1FUV models are minor for all amino acids with the exception of the negatively charged AAs (Asp3, Asp7) and Gly11, largely affected by presence of the aqueous solution (water and ions). In particular, Phe9 residue in salted models with 100 H₂O or pure 100 H₂O (models (v) and (iii)) differs from other models in which its PC gains a small charge. This behavior is even more pronounced in model (v) when the PC of Phe9 is flipped from

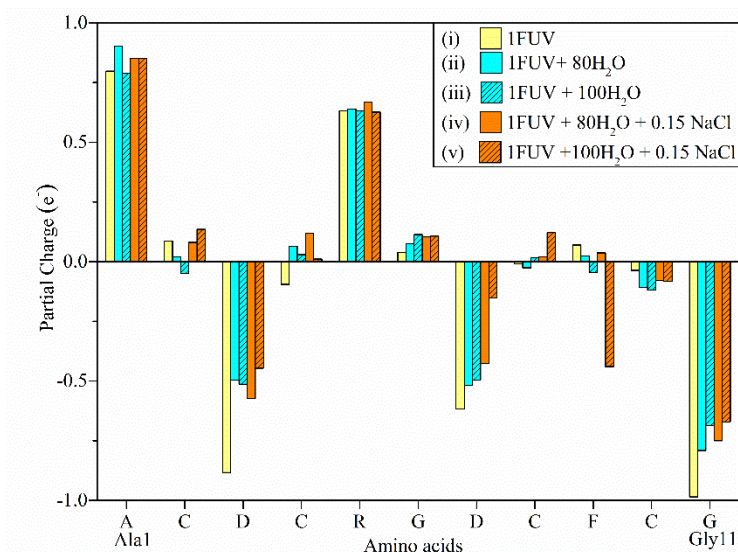


Fig. 6. Partial charge (PC) distributions in terms of AAs for five 1FUV models (i)-(v) in various environments as shown in legends.

a negligible $0.07 e^-$ in the dry case to a negative value of $-0.44 e^-$. A close examination of such behavior reveals that the oxygen atom in the backbone of Phe9 can form a HB with one water molecule in all four solvent models, but its separation distance varies. In model (v), this HB having a much shorter distance of 1.65 \AA as compared to 1.83 , 1.78 , and 2.52 \AA for models (ii), (iii), and (iv) respectively. This finding clearly indicates that local environments have a direct effect on PC distributions. It is noteworthy that positive and negative PCs in Ala1 and the Gly11 at both terminals of 1FUV result from the NH_2^+ in N-terminal and COO^- in C-terminal respectively and these opposite PCs neutralize each other.

3.2 Connection to AABP

All polypeptides are composed of amino acids that follow one another along the primary sequence. Because of the unique 3D structure of folded proteins, the AAs interact not only between nearest neighbors (NN) in the primary sequence, but also with other non-NN or spatially not vicinal i.e., nonlocal AAs. These non-local interactions between AAs along the primary sequence allow for the formation of AA-AA bond pairs (AABPs) (see Eq. (3) in **Method Section C**) [22, 60]. The distribution of AABP and their characteristics in the 7 domains of the Spike-protein in SARS-COVID-2 has been described in detail [22]. The same type of analysis can be performed also for the 1FUV protein models and for the SD1 small protein models, revealing the potential implications for the electronic part of their dielectric constant. **Fig. 7** shows the comparison of the calculated

AABP distribution and the 3D bonding network in dry and solvated 1FUV and SD1. They differ substantially mainly because of the differences in their structures. SD1 has a long ribbon-like structure and naturally exhibits less off-diagonal AABP contribution (4 out of 24 AAs or 0.16%), while 1FUV is a rather compact protein with only 11 AAs, displaying a substantial contribution from AABP (8 out of 11 or 73%).

In **Fig.7**, we display and compare the results for the AABP values in dry and solvated modes for 1FUV and SD1. The main observations can be summarized as follows:

1. The AAs in 1FUV models have far more off-diagonal contributions to AABP than the AAs in SD1 as expected from the very different 3D structure of these 2 proteins.
2. The solvated models exhibit a reduced total AABP mainly from the reduced NN interactions.
3. The 5 off-diagonal pairs in dry 1FUV are reduced to 4 pairs when solvated by 80 H_2O molecules. The missing pair is between Ala1-Phe9. Surprisingly, when solvated by 100 H_2O molecules, the number of pairs increases again to 8. The new pairs are Ala1-Asp3, Cys2-Cys8, Asp3-Cys8, Gly6-Phe9.
4. For SD1, the 2 off-diagonal pairs remain the same when solvated by 300 H_2O molecules, but the total AABP values for all AAs are decreased by about 13 % on the average.
5. The decrease in total AABP in solvated models in comparison to the dry model is due to the interactions between AAs; modified by the presence of H_2O molecules.

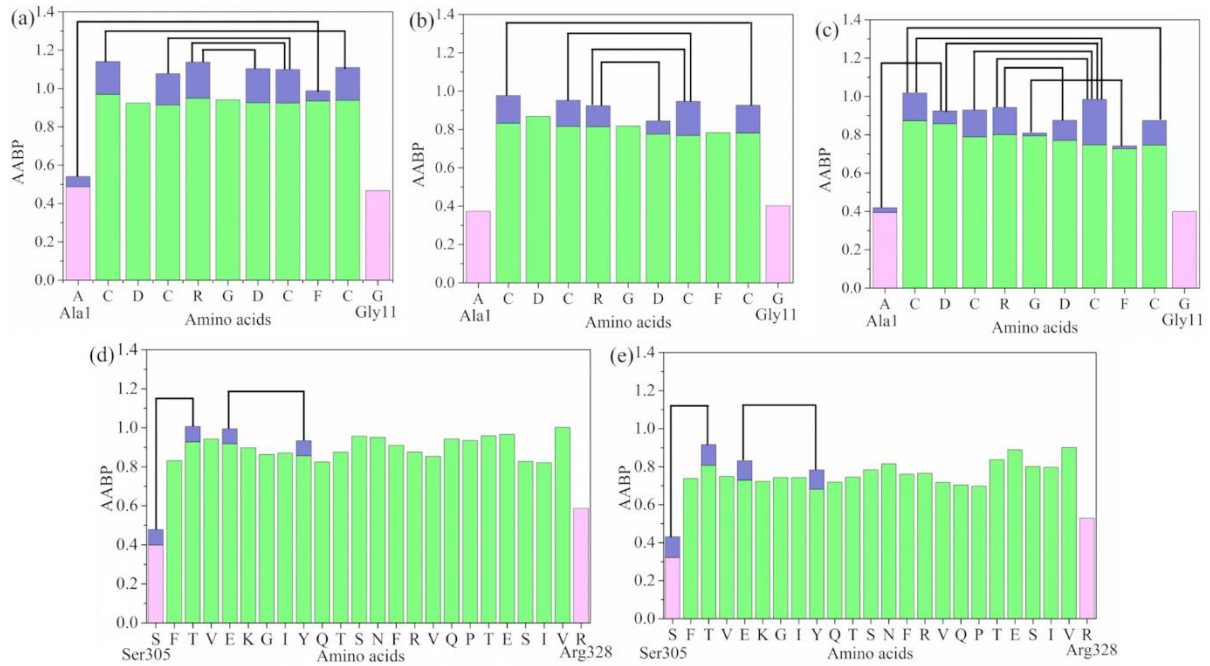


Fig. 7: Distribution of AABP in (a) 1FUUV, (b) 1FUUV in 80 H₂O molecules, (c) 1FUUV in 100 H₂O molecules, (d) SD1, and (e) SD1 in 300 H₂O molecules. The color bar represents following, light pink: sum AABP of AA with single nearest neighbor, green: sum AABP of AA with two nearest neighbors, light blue: AA with off-diagonal AABP. The black curve lines represent off-diagonal bonding between two AAs.

6. It can then be concluded that the total AABP values can be changed in a rather complicated fashion depend on the nature of the protein and the amount of water molecules surrounding them.

3.3 Specific applications

The Spike-protein of SARS-CoV-2 is an ideal target for vaccine development and other therapeutic treatment due to its essential role in the virus life cycle [64, 65] As has already been pointed out in **Section 2.3**, that the structure of Spike-protein is very complex as it exists in a trimeric form with each

protomer consisting of two functional subunits, S1 and S2. The S1 subunit comprises a signal sequence (SS) at the N-terminal end, followed by NTD and RBD and two structurally conserved subdomains (SD1 and SD2). The S2 subunit is subdivided into a fusion peptide (FP), two heptad repeats (HR1 and 2) and a central helix (CH), a connector domain (CD), a transmembrane domain (TM) and the cytoplasmic tail (CT). Between the S1 and S2 subunits, there are two protease cleavage sites (S1/S2 and S2') [61].

Recently, we have succeeded in the largest *ab initio* quantum chemical computation to date on the S-protein by using a *divide and conquer* strategy for all subdomains of the Spike-protein available in 6VSB [22] while here, the electronic dielectric properties of SD1, that acts as a hinge point for the RBD in the down to up transitions [66], in the dry and solvated environments have been

investigated. The same type of calculations for other much larger subdomains is also possible to gain additional insights on the electrostatic interactions in the Spike-protein, and the results could be very useful for COVID-19 researchers working on detection techniques. For example, the biosensor based on imaging ellipsometry was used to detect two neutralizing monoclonal antibodies and serial serum samples from SARS-CoV patients [67, 68]. Ellipsometry is an optical technique closely related to the dielectric properties[68]. We speculate that the calculated dielectric constants of the Spike-protein can be helpful in their design and implementation. On other hand, the electrostatic interactions have been demonstrated to play a major role in enhancing the binding affinity of RBD of SARS-CoV-2 to ACE2 as compared to SARS-CoV [69-72] and the polar solvation free energy of RBD-ACE2 complex is in fact quite sensitive to the electronic dielectric constant. However, all these MD simulations are based on the continuum models with uncertain dielectric constants for RBD or ACE2, that are usually assumed in the range of 1 to 4.

Our work on the RBD-ACE2 interface complex is in progress and will be reported elsewhere. We have high expectations that the calculated dielectric constant based on a more realistic *ab initio* approach could be used to achieve a reasonable value for overall electrostatic interactions (Coulombic and polar solvation energies) as compared to the experimental data.

4. CONCLUSION

In conclusion, we have presented a new method for calculating the electronic part of the dielectric constants for small proteins. A succinct summary of the work presented above is as follows:

1. we introduced the QMRPA method for static electronic dielectric constants $\epsilon_1(0)$ for small proteins.
2. we presented explicit calculations of $\epsilon_1(0)$ for 20 canonical amino acids. There is a strong inverse correlation between $\epsilon_1(0)$ and the total bond order of the amino acid. The presence of the unique S-containing amino acids (Met, Cys) tends to enhance the $\epsilon_1(0)$.
3. we presented $\epsilon_1(0)$ for the RGD (1FUV) peptide and the SD1 subdomain of the Spike-protein of SARS-CoV-2 virus. The $\epsilon_1(0)$ of 1FUV and SD1 decreases in the presence of a solvent environment with and without salt ions.
5. we pointed out possible connections between atomic scale partial charges of AAs in proteins and their specific role in the electrostatic interaction.
6. we described the role of non-local AA-AA interactions *via* AABP values in the 3D structure of the protein and their possible connection to the static electronic dielectric constant.
7. we laid out the roadmap to use QMRPA method for applications to electrostatic interactions in Spike-protein and other biomolecular systems in general.

APPENDIX

METHODS

A. Model Construction

The initial structures of 20 AAs were obtained from PubChem database [46]. For RGD-4C peptide, the starting structure was downloaded from the RCSB protein data bank (PDB) with ID: 1FUV [57] based on nuclear magnetic resonance (NMR) data. 1FUV contains 11 AAs with 135 atoms including hydrogen atoms. From the initial dry model, two purely solvated models with 80 and 100 H₂O molecules and two salted models by adding salt with concentration of 0.15 M were obtained. The purely solvated models are generated by the three-point charge TIP3P model [73] using LEaP program included in the AMBER 18 package [74]. In total, five RGD-4C models were constructed. One dry, two purely solvated and two salted models with 0.15 M salt concentration (4 Na⁺ and 3 Cl⁻ ions). These ions are systematically placed in a shell around system using a Coulombic potential on a grid using the program LEaP and the number of ions is determined based on the volume of the cells. For dry model of SD1 in Spike-protein of SARS-CoV-2, its fully relaxed structure was obtained from ref. [22] from PDB with ID: 6VSB [61]. The solvated SD1 model is built with 300 water molecules using the same approach as in 1FUV cases. The selection of 300 water molecules is based on UCSF Chimera program [75].

B. Structure optimization using VASP

This initial atomic-scale structure for an individual biomolecule such as amino acid or a small protein is fully optimized by using Vienna ab initio simulation package (VASP) [76]. The pseudopotential plane-wave based VASP package is known for its efficiency in relaxation to the equilibrium structure at minimal energy. We use the projector augmented wave (PAW) method with Perdew-Burke-Ernzerhof (PBE) exchange correlation potential [77] within the generalized gradient approximation (GGA). Our experience and tests suggest the use of following input parameters in VASP: Energy cut-off energy at 500 eV, electronic convergence of 10⁻⁴ eV; force convergence for ionic steps at -10⁻² eV/Å and a single k-point sampling. For the optimization of a small protein SD1 with 24 amino acids and 391 atoms, the total energy decreases from -2370.90 eV to -2379.21 eV or 2.05 kJ/mol per atom. The VASP-relaxed structure is used for the electronic structure, interatomic bonding and dielectric constant calculations described below.

C. DFT calculations using OLCAO

For electronic structure, interatomic interactions and dielectric function calculations, a different DFT method is used, the all-electron orthogonalized linear combination of atomic orbitals (OLCAO) method [27] which was developed in-house. The combination of these two different DFT codes is extremely efficient for large complex materials is well documented [58, 78-80] especially for complex proteins such as the SARS-CoV-2 virus [22, 59, 60]. The key feature of the OLCAO method is the provision for the effective charge (Q^*) on each atom and the bond order (BO) values $\rho_{\alpha\beta}$ between any pairs of atoms. They are obtained from the *ab initio* wave functions with atomic basis expansion calculated quantum mechanically:

$$Q_{\alpha}^* = \sum_i \sum_{m,occ} \sum_{j,\beta} C_{i\alpha}^{*m} C_{j\beta}^m S_{i\alpha,j\beta} \quad (1)$$

$$\rho_{\alpha\beta} = \sum_{m,occ} \sum_{i,j} C_{i\alpha}^{*m} C_{j\beta}^m S_{i\alpha,j\beta} \quad (2)$$

In the above equations, $S_{i\alpha,j\beta}$ are the overlap integrals between the i^{th} orbital in α^{th} atom and the j^{th} orbital in the β^{th} atom. $C_{j\beta}^m$ are the eigenvector coefficients of the m^{th} occupied molecular orbital levels. The partial charge (PC) or ($\Delta Q_{\alpha} = Q_{\alpha}^0 - Q_{\alpha}^*$) is the deviation of the effective charge Q_{α}^* from the neutral atomic charge Q_{α}^0 on the same atom α . The BO values $\rho_{\alpha\beta}$ in Eq. (2) can be calculated for every pair of atoms (α, β) in the optimized structure with precise atomic positions. The BO quantifies the strength of the bond between two atoms and generally scales with the bond length (BL) being also influenced by the surrounding atoms. The calculation of PC and BO are based on the Mulliken scheme [81, 82], hence are basis-dependent. Comparisons of BO values using different methods should be treated with caution. The sum BO of all pairs in the system gives TBO.

In proteins, the focus is on amino acids (AAs) or residues. AAs are essentially biomolecules containing different atoms with different molecular configurations and orientations. Strictly speaking, assigning the distance of separation between two AAs in a protein to describe their interaction is a vague and arbitrary parameter. However, with the quantum mechanically based OLCAO method and with the interatomic interaction between all atoms available, we can define the bonding between two AAs u and v with no ambiguity, which we denote it as *amino acid bond pair* (AABP) [60]:

$$AABP(u, v) = \sum_{\alpha \in u} \sum_{\beta \in v} \rho_{\alpha i, \beta j} \quad (3)$$

where the summations are over atoms α in AA u and atoms β in AA v . This is a far more rigorously defined quantity and can be further extended to different units or groups of AAs if necessary. The merit of the above scheme is that AABP includes all possible bonding between two amino acids such as covalent, ionic, hydrogen bonding (HB) and even their intermediate mixtures [83]. This single quantitative parameter reflects the internal bonding strength among amino acids. Proteins are generally described in the sequence of connected nearest neighbor (NN) AAs. In fact, non-NN amino acids in the 3D structure can also interact. AAPB can be resolved into NN and non-local bonding proving much more penetrating details on inter amino acid bonding.

D. Optical transition and random phase approximation

We use OLCAO method to calculate the optical transitions in a biomolecule from occupied to the unoccupied molecular states within the random phase approximation [20]. The transition explicitly involves the dipole transition matrix elements calculated quantum mechanically from the *ab initio* wave functions which automatically obey the transition rules involved in the transition.

The complex dielectric function is given by:

$$\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega) \quad (4)$$

$$\varepsilon_2(\omega) = \frac{e^2}{\pi m \omega^2} \int_{BZ} dk^3 \sum_{n,l} |\langle \Psi_n(\vec{k}, \vec{r}) | P | \Psi_l(\vec{k}, \vec{r}) \rangle|^2 \delta(E_n(\vec{k}) - E_l(\vec{k}) - E) \quad (5)$$

Here, $\psi_n(\vec{k}, \vec{r})$ is the Bloch wave function for the n^{th} band with energy $E_n(\vec{k})$ at Brillouin zone point k . Momentum matrix elements $\langle \Psi_n(\vec{k}, \vec{r}) | P | \Psi_l(\vec{k}, \vec{r}) \rangle$ from occupied valence band states (l) to empty conduction band states (n) are calculated from *ab initio* wave functions. The E_l and E_n are the energy of occupied state and unoccupied state, respectively. E is the photon energy.

The real part $\varepsilon_1(\omega)$ can be derived from ε_2 by using the Kramers-Kronig relation [84]:

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{s \varepsilon_2(s)}{s^2 - \omega^2} ds \quad (6)$$

The dielectric constant is defined as the zero-frequency limit of $\varepsilon_1(\omega)$, or $\varepsilon_0 = \varepsilon_1(0)$. From the complex dielectric function equation (4), the energy loss function $F(\omega)$ can be obtained.

$$F(\omega) = IM \left(-\frac{1}{\varepsilon(\omega)} \right) = \frac{\varepsilon_2(\omega)}{\varepsilon_1^2(\omega) + \varepsilon_2^2(\omega)} \quad (7)$$

In crystalline solids, the position of the peak of energy loss function $F(\omega)$ at ω_0 is identified as the plasmon frequency which is the energy for the collective excitation of the electrons in the unoccupied bands.

Depending on the structural complexity of the material, the plasmon frequency ω_p is not the position of a single major peak and may involve multiple peaks, but they are all subject to the same

sum rule [93]. So we can use a single effective parameter ω_p in relation to optical transitions in Eq. (5) rooted in the random phase approximation [20] to account for the fundamental differences between the calculation of dielectric constant of periodic infinite systems for crystals of infinite extent via Bloch theorem [21] and the finite biological materials of vastly varying sizes. We use a large supercell for each biological molecule to avoid interaction at atomic level beyond the molecule itself using codes developed for periodic solids. The reciprocal space is reduced to a single k-point (0,0,0). The difference in the calculated electronic dielectric constant between these two entirely different systems is rationalized by a scaling factor based on the same sum rule of the electrons under transitions [85] in the spirit of the random phase approximation [20]

$$\int_0^{\infty} \omega \varepsilon_2(\omega) d\omega = \frac{\pi \omega_p^2}{2} \quad (8)$$

This scaling factor is obtained by carefully matching the area below the energy loss function with the ones using multiple Gaussian fits to minimize the difference. The scaling factor is determined by the weighted average of from the contribution each Gaussian component and the effective plasmon frequency ω_p . So, the scaling factor used is the ratio between the left-hand side (LHS) and the right-hand side (RHS) of Eq. (8). The LHS area is obtained by integrating the energy times $\varepsilon_2(\omega)$ as a function of energy in the periodic supercell calculation for the protein, while the RHS area is obtained using the weighted average from the contributions of the Gaussian components from each peak.

The procedures outlined above can accurately reflect the complex structural difference of the biomolecular systems as apparent in their electronic dielectric constant $\varepsilon_1(0)$. We name it as the quantum mechanical random phase approximation (QMRAP) method.

ASSOCIATED CONTENT

Supporting Information.

Additional figures and Tables are provided in the Supporting Information.

AUTHOR INFORMATION

Corresponding Author

Wai-Yim Ching, Email: Chingw@umkc.edu ORCID 0000-0001-7738-8822

Rudolf Podgornik, Email: rudipod@gmail.com ORCID 0000-0002-3855-4637

Author Contributions

WC and RP conceived the project. PA and WC performed the calculations, PA made most of the figures. WC and RP drafted the paper with inputs from PA, BJ. All authors participated in the discussion and interpretation of the results. All authors edited and proofread the final manuscript.

Funding Sources

This project is funded by the National Science Foundation of USA: RAPID DMR/CMMT-2028803.

ACKNOWLEDGMENT

This research used the resources of the National Energy Research Scientific Computing Center supported by DOE under Contract No. DE-AC03-76SF00098 and the Research Computing Support Services (RCSS) of the University of Missouri System. This project is funded by the National Science Foundation of USA: RAPID DMR/CMMT-2028803. RP acknowledges funding from the Key project #12034019 of the National Natural Science Foundation of China.

REFERENCES

1. Schutz, C.N. and A. Warshel, *What are the dielectric “constants” of proteins and how to validate electrostatic models?* Proteins: Structure, Function, and Bioinformatics, 2001. **44**(4): p. 400-417, DOI: 10.1002/prot.1106.
2. Zhou, H.-X. and X. Pang, *Electrostatic interactions in protein structure, folding, binding, and condensation.* Chemical reviews, 2018. **118**(4): p. 1691-1741, DOI: 10.1021/acs.chemrev.7b00305.
3. Bai, C. and A. Warshel, *Critical Differences Between the Binding Features of the Spike Proteins of SARS-CoV-2 and SARS-CoV.* The Journal of Physical Chemistry B, 2020. **124**(28): p. 5907-5912, DOI: 10.1021/acs.jpcc.0c04317.
4. Javidpour, L., et al., *Electrostatic interaction between SARS-CoV-2 virus and charged electret fibre.* arXiv preprint arXiv:2012.07160, 2020, DOI: arXiv:2012.07160.
5. McLaughlin, S., *The electrostatic properties of membranes.* Annual review of biophysics and biophysical chemistry, 1989. **18**(1): p. 113-136, DOI: 10.1146/annurev.bb.18.060189.000553.
6. Warshel, A., et al., *Modeling electrostatic effects in proteins.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2006. **1764**(11): p. 1647-1676, DOI: 10.1016/j.bbapap.2006.08.007.
7. Li, L., et al., *On the dielectric “constant” of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi.* Journal of chemical theory and computation, 2013. **9**(4): p. 2126-2136, DOI: 10.1021/ct400065j.
8. García-Moreno, B.E., et al., *Experimental measurement of the effective dielectric in the hydrophobic core of a protein.* Biophysical chemistry, 1997. **64**(1-3): p. 211-224, DOI: 10.1016/S0301-4622(96)02238-7.
9. Bibi, F., et al., *A review: Origins of the dielectric properties of proteins and potential development as bio-sensors.* Sensors, 2016. **16**(8): p. 1232, DOI: 10.3390/s16081232.
10. Gitlin, I., J.D. Carbeck, and G.M. Whitesides, *Why are proteins charged? Networks of charge–charge interactions in proteins measured by charge ladders and capillary electrophoresis.* Angewandte Chemie International Edition, 2006. **45**(19): p. 3022-3060, DOI: 10.1002/anie.200502530.
11. Penfold, R., J. Warwicker, and B. Jönsson, *Electrostatic models for calcium binding proteins.* The Journal of Physical Chemistry B, 1998. **102**(43): p. 8599-8610, DOI: 10.1021/jp973420s.
12. Fogolari, F., A. Brigo, and H. Molinari, *The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology.* Journal of Molecular Recognition, 2002. **15**(6): p. 377-392, DOI: 10.1002/jmr.577.

13. Gilson, M.K., *Theory of electrostatic interactions in macromolecules*. Current opinion in structural biology, 1995. **5**(2): p. 216-223, DOI: 10.1016/0959-440x(95)80079-4.
14. Antosiewicz, J., J.A. McCammon, and M.K. Gilson, *Prediction of pH-dependent properties of proteins*. Journal of molecular biology, 1994. **238**(3): p. 415-436, DOI: 10.1006/jmbi.1994.1301.
15. Li, C., et al., *DelPhi suite: New developments and review of functionalities*. Journal of computational chemistry, 2019. **40**(28): p. 2502-2508, DOI: 10.1002/jcc.26006.
16. Koehl, P., *Electrostatics calculations: latest methodological advances*. Current opinion in structural biology, 2006. **16**(2): p. 142-151, DOI: 10.1016/j.sbi.2006.03.001.
17. Simonson, T., *Electrostatics and dynamics of proteins*. Reports on Progress in Physics, 2003. **66**(5): p. 737, DOI: 10.1088/0034-4885/66/5/202.
18. Lund, M., B. Jönsson, and C.E. Woodward, *Implications of a high dielectric constant in proteins*. The Journal of chemical physics, 2007. **126**(22): p. 06B608, DOI: 10.1063/1.2741543.
19. Voges, D. and A. Karshikoff, *A model of a local dielectric constant in proteins*. The Journal of chemical physics, 1998. **108**(5): p. 2219-2227, DOI: 10.1063/1.475602.
20. Ehrenreich, H. and M.H. Cohen, *Self-consistent field approach to the many-electron problem*. Physical Review, 1959. **115**(4): p. 786-790, DOI: 10.1103/physrev.115.786.
21. Bloch, F., *Über die quantenmechanik der elektronen in kristallgittern*. Zeitschrift für physik, 1929. **52**(7): p. 555-600, DOI: 10.1007/BF01339455.
22. Ching, W.-Y., et al., *Ultra-Large-Scale Ab Initio Quantum Chemical Computation of Bio-Molecular Systems: The Case of Spike Protein of SARS-CoV-2 Virus*. Computational and Structural Biotechnology Journal 2021. **19**: p. 1288-1301, DOI: 10.1016/j.csbj.2021.02.004.
23. Millefiori, S., et al., *Electronic and vibrational polarizabilities of the twenty naturally occurring amino acids*. Biophysical chemistry, 2008. **132**(2-3): p. 139-147, DOI: 10.1016/j.bpc.2007.11.003.
24. Swart, M., J.G. Snijders, and P.T. van Duijnen, *Polarizabilities of amino acid residues*. Journal of Computational Methods in Sciences and Engineering, 2004. **4**(3): p. 419-425, DOI: 10.3233/JCM-2004-4317.
25. Abeyrathne, C.D., et al., *An ab-initio computational method to determine dielectric properties of biological materials*. Scientific reports, 2013. **3**(1): p. 1-5, DOI: 10.1038/srep01796.
26. Stöhr, M., T. Van Voorhis, and A. Tkatchenko, *Theory and practice of modeling van der Waals interactions in electronic-structure calculations*. Chemical Society Reviews, 2019. **48**(15): p. 4118-4154, DOI: 10.1039/c9cs00060g.
27. Ching, W.-Y. and P. Rulis, *Electronic Structure Methods for Complex Materials: The orthogonalized linear combination of atomic orbitals*. 2012: Oxford University Press.
28. Ching, W., et al., *Electronic structure and excitonic-enhanced superconducting mechanism in YBa₂Cu₃O_{7- δ}* . Physical review letters, 1987. **59**(12): p. 1333, DOI: 10.1103/PhysRevLett.59.1333.
29. Zhao, G.-L., et al., *Theoretical calculation of optical properties of Y-Ba-Cu-O superconductors*. Physical Review B, 1987. **36**(13): p. 7203, DOI: 10.1103/PhysRevB.36.7203.
30. Ching, W., et al., *First-principles calculation of optical properties of C₆₀ in the fcc lattice*. Physical review letters, 1991. **67**(15): p. 2045, DOI: 10.1103/PhysRevLett.67.2045.
31. Xu, Y.-N., M.-Z. Huang, and W. Ching, *Optical properties of superconducting K₃C₆₀ and insulating K₆C₆₀*. Physical Review B, 1991. **44**(23): p. 13171, DOI: 10.1103/PhysRevB.44.13171.
32. Baral, K., P. Adhikari, and W.Y. Ching, *Ab initio Modeling of the Electronic Structures and Physical Properties of a-Si_{1-x}GexO₂ Glass (x= 0 to 1)*. Journal of the American Ceramic Society, 2016. **99**(11): p. 3677-3684, DOI: 10.1111/jace.14386.
33. Adhikari, P., et al., *Structure and electronic properties of a continuous random network model of an amorphous zeolitic imidazolate framework (a-ZIF)*. The Journal of Physical Chemistry C, 2016. **120**(28): p. 15362-15368, DOI: 10.1021/acs.jpcc.6b06337.

34. Adhikari, P., et al., *Deformation behavior of an amorphous zeolitic imidazolate framework—from a supersoft material to a complex organometallic alloy*. Physical Chemistry Chemical Physics, 2018. **20**(46): p. 29001-29011, DOI: 10.1039/C8CP05610B.
35. Ouyang, L., et al., *Electronic structure and bonding in vitamin B12, cyanocobalamin*. Journal of Molecular Structure: THEOCHEM, 2003. **622**(3): p. 221-227, DOI: 10.1016/S0166-1280(02)00647-4.
36. Ouyang, L., et al., *Accurate redetermination of the X-ray structure and electronic bonding in adenosylcobalamin*. Inorganic chemistry, 2004. **43**(4): p. 1235-1241, DOI: 10.1021/ic0348446.
37. Rulis, P., L. Ouyang, and W. Ching, *Electronic structure and bonding in calcium apatite crystals: Hydroxyapatite, fluorapatite, chlorapatite, and bromapatite*. Physical Review B, 2004. **70**(15): p. 155104, DOI: 10.1103/PhysRevB.70.155104.
38. Rulis, P., et al., *Electronic structure, bonding, charge distribution, and x-ray absorption spectra of the (001) surfaces of fluorapatite and hydroxyapatite from first principles*. Physical Review B, 2007. **76**(24): p. 245410, DOI: 10.1103/PhysRevB.76.245410.
39. Kahr, B., et al., *Herapathite*. Science, 2009. **324**(5933): p. 1407-1407, DOI: 10.1126/science.1173605.
40. Liang, L., et al., *Theoretical study of the large linear dichroism of herapathite*. Physical Review B, 2009. **80**(23): p. 235132, DOI: 10.1103/PhysRevB.80.235132.
41. French, R.H., et al., *Long range interactions in nanoscale science*. Reviews of Modern Physics, 2010. **82**(2): p. 1887, DOI: 10.1103/RevModPhys.82.1887.
42. Hopkins, J.C., et al., *Dielectric response variation and the strength of van der Waals interactions*. Journal of colloid and interface science, 2014. **417**: p. 278-284, DOI: 10.1016/j.jcis.2013.10.040.
43. Schimelman, J.B., et al., *Optical properties and electronic transitions of DNA oligonucleotides as a function of composition and stacking sequence*. Physical Chemistry Chemical Physics, 2015. **17**(6): p. 4589-4599, DOI: 10.1039/C4CP03395G.
44. Gong, Y., et al., *Designing the interface of carbon nanotube/biomaterials for high-performance ultra-broadband photodetection*. ACS applied materials & interfaces, 2017. **9**(12): p. 11016-11024, DOI: 10.1021/acsami.7b00352.
45. Gong, M., et al., *Polarity-Controlled Attachment of Cytochrome C for High-Performance Cytochrome C/Graphene van der Waals Heterojunction Photodetectors*. Advanced Functional Materials, 2018. **28**(5): p. 1704797, DOI: 10.1002/adfm.201704797.
46. Kim, S., et al., *PubChem substance and compound databases*. Nucleic acids research, 2016. **44**(D1): p. D1202-D1213, DOI: 10.1093/nar/gkv951.
47. Khanarian, G. and W. Moore, *The Kerr effect of amino acids in water*. Australian Journal of Chemistry, 1980. **33**(8): p. 1727-1741, DOI: 10.1071/CH9801727.
48. Dos Santos, L.H., A. Krawczuk, and P. Macchi, *Distributed atomic polarizabilities of amino acids and their hydrogen-bonded aggregates*. The Journal of Physical Chemistry A, 2015. **119**(13): p. 3285-3298, DOI: 10.1021/acs.jpca.5b00069.
49. Donev, R., *Protein and Peptide Nanoparticles for Drug Delivery*. 2015: Academic Press.
50. Wang, F., et al., *The functions and applications of RGD in tumor therapy and tissue engineering*. International journal of molecular sciences, 2013. **14**(7): p. 13447-13462, DOI: 10.3390/ijms140713447.
51. Avraamides, C.J., B. Garmany-Susini, and J.A. Varner, *Integrins in angiogenesis and lymphangiogenesis*. Nature Reviews Cancer, 2008. **8**(8): p. 604-617, DOI: 10.1038/nrc2353.
52. Garanger, E., D. Boturyn, and P. Dumy, *Tumor targeting with RGD peptide ligands-design of new molecular conjugates for imaging and therapy of cancers*. Anti-Cancer Agents in Medicinal

- Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents), 2007. **7**(5): p. 552-558, DOI: 10.2174/187152007781668706.
53. Metcalfe, A.D. and M.W. Ferguson, *Tissue engineering of replacement skin: the crossroads of biomaterials, wound healing, embryonic development, stem cells and regeneration*. Journal of the Royal Society Interface, 2007. **4**(14): p. 413-437, DOI: 10.1098/rsif.2006.0179.
 54. Zhou, Y., S. Chakraborty, and S. Liu, *Radiolabeled cyclic RGD peptides as radiotracers for imaging tumors and thrombosis by SPECT*. Theranostics, 2011. **1**: p. 58, DOI: 10.7150/thno/v01p0058.
 55. Meyers, S.R. and M.W. Grinstaff, *Biocompatible and bioactive surface modifications for prolonged in vivo efficacy*. Chemical reviews, 2012. **112**(3): p. 1615-1632, DOI: 10.1021/cr2000916.
 56. Bellis, S.L., *Advantages of RGD peptides for directing cell association with biomaterials*. Biomaterials, 2011. **32**(18): p. 4205-4210, DOI: 10.1016/j.biomaterials.2011.02.029.
 57. *1FUV SOLUTION STRUCTURE OF AN RGD PEPTIDE ISOMER-A*. 2000; Available from: <https://www.rcsb.org/structure/1FUV>.
 58. Adhikari, P., et al., *Electronic structure, dielectric response, and surface charge distribution of RGD (1FUV) peptide*. Scientific reports, 2014. **4**: p. 5605, DOI: 10.1038/srep05605.
 59. Adhikari, P., et al., *Intra- and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding*. Physical Chemistry Chemical Physics, 2020. **22**(33): p. 18272-18283, DOI: 10.1039/D0CP03145C.
 60. Adhikari, P. and W.-Y. Ching, *Amino acid interacting network in the receptor-binding domain of SARS-CoV-2 spike protein*. RSC Advances 2020. **10**: p. 39831-39841, DOI: 10.1039/d0ra08222h.
 61. Wrapp, D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation*. Science, 2020. **367**(6483): p. 1260-1263, DOI: 10.1126/science.abb2507.
 62. Zheng, H., et al., *Electrostatic Effect of Functional Surfaces on the Activity of Adsorbed Enzymes: Simulations and Experiments*. ACS Applied Materials & Interfaces, 2020. **12**(31): p. 35676-35687, DOI: 10.1021/acsami.0c08080.
 63. Ionescu, C.-M., et al., *AtomicChargeCalculator: interactive web-based calculation of atomic charges in large biomolecular complexes and drug-like molecules*. Journal of cheminformatics, 2015. **7**(1): p. 1-13, DOI: 10.1186/s13321-015-0099-x.
 64. Gao, Q., et al., *Development of an inactivated vaccine candidate for SARS-CoV-2*. Science, 2020. **369**(6499): p. 77-81, DOI: 10.1126/science.abc1932.
 65. Chi, X., et al., *A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2*. Science, 2020. **369**(6504): p. 650-655, DOI: 10.1126/science.abc6952.
 66. Henderson, R., et al., *Controlling the SARS-CoV-2 spike glycoprotein conformation*. Nature structural & molecular biology, 2020. **27**(10): p. 925-933, DOI: 10.1038/s41594-020-0479-4.
 67. Qi, C., et al., *Investigation of interaction between two neutralizing monoclonal antibodies and SARS virus using biosensor based on imaging ellipsometry*. Biomedical microdevices, 2006. **8**(3): p. 247-253, DOI: 10.1007/s10544-006-8305-2.
 68. Orooji, Y., et al., *An overview on SARS-CoV-2 (COVID-19) and other human coronaviruses and their detection capability via amplification assay, chemical sensing, biosensing, immunosensing, and clinical assays*. Nano-Micro Letters, 2021. **13**(1): p. 1-30, DOI: 10.1007/s40820-020-00533-y.
 69. Spinello, A., A. Saltalamacchia, and A. Magistrato, *Is the rigidity of SARS-CoV-2 spike receptor-binding motif the hallmark for its enhanced infectivity? Insights from all-atom simulations*. The journal of physical chemistry letters, 2020. **11**(12): p. 4785-4790, DOI: 10.1021/acs.jpclett.0c01148.
 70. Xie, Y., et al., *Spike proteins of SARS-CoV and SARS-CoV-2 utilize different mechanisms to bind with human ACE2*. Frontiers in molecular biosciences, 2020. **7**, DOI: 10.3389/fmolb.2020.591873.

71. Peng, C., et al., *Computational Insights into the Conformational Accessibility and Binding Strength of SARS-CoV-2 Spike Protein to Human Angiotensin-Converting Enzyme 2*. The journal of physical chemistry letters, 2020. **11**: p. 10482-10488, DOI: 10.1021/acs.jpclett.0c02958.
72. Amin, M., M.K. Sorour, and A. Kasry, *Comparing the binding interactions in the receptor binding domains of SARS-CoV-2 and SARS-CoV*. The journal of physical chemistry letters, 2020. **11**(12): p. 4897-4900, DOI: 10.1021/acs.jpclett.0c01064.
73. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of chemical physics, 1983. **79**(2): p. 926-935, DOI: 10.1063/1.445869.
74. Pearlman, D.A., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules*. Computer Physics Communications, 1995. **91**(1-3): p. 1-41, DOI: 10.1016/0010-4655(95)00041-D.
75. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis*. Journal of computational chemistry, 2004. **25**(13): p. 1605-1612, DOI: 10.1002/jcc.20084.
76. VASP - Vienna Ab initio Simulation Package. Available from: <https://www.vasp.at/>.
77. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized gradient approximation made simple*. Physical review letters, 1996. **77**(18): p. 3865, DOI: 10.1103/PhysRevLett.77.3865.
78. Poudel, L., et al., *Implication of the solvent effect, metal ions and topology in the electronic structure and hydrogen bonding of human telomeric G-quadruplex DNA*. Physical Chemistry Chemical Physics, 2016. **18**(31): p. 21573-21585, DOI: 10.1039/c6cp04357g.
79. Poudel, L., et al., *Impact of hydrogen bonding in the binding site between capsid protein and MS2 bacteriophage ssRNA*. The Journal of Physical Chemistry B, 2017. **121**(26): p. 6321-6330, DOI: 10.1021/acs.jpcb.7b02569.
80. Eifler, J., et al., *Charge distribution and hydrogen bonding of a collagen α 2-chain in vacuum, hydrated, neutral, and charged structural models*. International Journal of Quantum Chemistry, 2016. **116**(9): p. 681-691, DOI: 10.1002/qua.25089.
81. Mulliken, R.S., *Electronic population analysis on LCAO–MO molecular wave functions. I*. The Journal of Chemical Physics, 1955. **23**(10): p. 1833-1840, DOI: 10.1063/1.1740588.
82. Mulliken, R., *Electronic population analysis on LCAO–MO molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies*. The Journal of Chemical Physics, 1955. **23**(10): p. 1841-1846, DOI: 10.1063/1.1740589.
83. Dereka, B., et al., *Crossover from hydrogen to chemical bonding*. Science, 2021. **371**(6525): p. 160-164, DOI: 10.1126/science.abe1951.
84. Martin, P.C., *Sum rules, Kramers-Kronig relations, and transport coefficients in charged systems*. Physical Review, 1967. **161**(1): p. 143, DOI: 10.1103/physrev.161.143.
85. Smith, D.Y., *Dispersion Theory, Sum Rules, and Their Application to the Analysis of Optical Data*. , in *Handbook of Optical Constants of Solids*, E. Palik, Editor. 1985, Academic Press: Orlando, Florida.