Modeling the interaction of SARS-CoV-2 binding to the ACE2 receptor via molecular theory of solvation

Kobryn, Alexander E.; Maruyama, Yutaka; Velázquez-Martínez, Carlos A.; Yoshida, Norio; Gusarov, Sergey

Publisher's version / Version de l'éditeur:
https://doi.org/10.1039/D1NJ02015C
New Journal of Chemistry, 45, 34, pp. 15448-15457, 2021-07-20

ARTICLE

Modeling the Interaction of SARS-CoV-2 Binding to the ACE2 Receptor by Molecular Theory of Solvation

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Alexander E. Kobryn,*[a] Yutaka Maruyama,[b] Carlos A. Velázquez Martínez,[c] Norio Yoshida,[d] Sergey Gusarov*[a]

The angiotensin-converting enzyme 2 (ACE2) protein is a cell gate receptor for the SARS-CoV-2 virus, responsible for the development of symptoms associated with the Covid-19 disease. Pharmacological inhibition of the SARS-CoV-2 spike receptor binding domain (RBD) and ACE2 interaction is one of the most attractive ways to prevent viral replication in human cells. Unfortunately, at this stage there is no complete picture of this process and so the computational modelling might provide an important insight valuable for the development of new and efficient inhibitors. In this work we propose to use the molecular theory of solvation to study the nanomorphology of the Spike-ACE2 binding formed by complex solvent (water, ions, dissolved drug-like molecules) and leading to viral protein with cell membrane receptors. In contrast to the typical molecular dynamics, the statistical-mechanical theory of solvation directly provides distributions of complex solvents around the binding location as well as the thermodynamics of solvation. We present an example of the application of the three-dimensional theory of solvation to model the nanomorphology formed by solvent environment around binding surfaces of interacting proteins. The results of our calculations are compared with other published data. Our recent developments allow to extend the application of the methodology for potential drug screening and virulence analysis.

1. Introduction

The coronavirus pandemic declared by the WHO in March 2020[1,2] has already triggered a significant global health and economic crisis affecting many areas of human lifestyle as well. This pandemic has also prompted an explosion of research on the etiological agent, namely the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) virus and the corresponding disease that it causes. This research covers a range of subjects including viral replication mechanisms, clinical features, the molecular structure of target viral proteins, and the mechanisms of viral infection. Despite the progress achieved in the development of new vaccines[3] the design of new and efficient drugs remains one of the key goals focused to slow down viral infections. To date, only a few drugs like Remdesivir[4,5] and Dexamethasone[6] have showed some degree of efficacy in treating the infection and its complications. This is because the development of new drug entities (NDEs) is an expensive, long, and complicated multistage process. For this reason, the overwhelming majority of drugs being tested against the SARS-CoV-2 virus are the result of repurposing currently approved drugs.

[a.] *Nanotechnology Research Centre, National Research Council Canada, 11421 Saskatchewan Drive NW, Edmonton, Alberta, T6G 2M9, Canada.*
[b.] *Architecture Development Team, FLAGSHIP 2020 Project, RIKEN Center for Computational Science, Kobe, Hyogo 650-0047, Japan.*
[c.] *2142-L Katz Group Centre for Research, University of Alberta, 11315-87 Avenue NW, Edmonton, Alberta, T6G 2H5, Canada*
[d.] *Department of Chemistry, Graduate School of Science, Kyushu University, 744, Motooka, Nishi-ku, Fukuoka, 819-0395, Japan*
* Corresponding authors; E-mail addresses: alex.kobryn@nrc-cnrc.gc.ca, sergey.gusarov@nrc-cnrc.gc.ca

The SARS-CoV-2 virus is a lipid-enveloped positive-sense RNA virus, possessing a characteristic structure on the lipid-envelope called a spike protein, which plays an essential role in the recognition and membrane fusion with the host cell[7,8] Therefore, the study of the structure and function of the spike protein is extremely important for the development of effective new drugs for SARS-CoV-2. In this regard, the Spike protein is made of two subunits, namely the S1 and S2 subunits. The S1 subunit contains the *N*-terminal domain and the receptor-binding domain (RBD), which is responsible for the binding with the angiotensin-converting enzyme 2 (ACE2) receptor on the host cell. The affinity of SARS-CoV-2 with ACE2 has been reported to be up to 20 times higher than that of SARS-CoV that caused the outbreak back in 2002, and this may explain why the SARS-CoV-2 virus is highly infectious and poses a much more serious health risk[9-11] A number of computational studies have been carried out and some others are underway; these studies aim to understand the mechanism of viral infection as it relates to the development of new small-molecule drugs that inhibit Spike RBD-ACE2 binding[12-23] Computational studies can be classified according to the viral target(s) they focus on. For example, most reports study the binding affinity of RBD with ACE2 using homology modeling, docking analysis and all-atom molecular dynamics (MD)[12-16]. Other studies analyze the allosteric inhibition produced by ligands blocking sites other than the RBD-ACE2 interaction[17-18]. These studies suggested the potential use of several drug candidates on the RBD-ACE2 binding process. Finally, a number of scientific reports have described the structural stability of the Spike protein and the target ACE2 cell receptor proteins in the presence of ligands with different binding properties to each of these two proteins[19-23].

On the other hand, due to the high computational cost and large conformational space to explore[24] the molecular dynamics-based approaches cannot reproduce the full details of the process of

binding and the fusion of coronavirus with the host cell membrane. There is still a number of questions related to the nanomorphology of environment to protein-protein binding interactions, such as the effect of solvent and ionic strength, additions, thermodynamics etc. Fortunately, they can be answered by the statistical mechanical theory of molecular liquids[25-27]. For example, recently an approximation based on this theory was used to obtain the solvation thermodynamics and structural maps of SARS-CoV-2 targets[28].

In the present work we investigate the solvation effects on the RBD-ACE2 binding, Figure 1, by following the approach based on the molecular theory of solvation called Three-Dimensional Reference Interaction Site Model (3D-RISM)[25,26]. Needless to say, the solvent effect is one of the most important factors in the protein folding and protein-protein binding. The 3D-RISM theory is a powerful tool for considering the solvation structure and thermodynamics of the biological molecules including molecular recognition, drug-binding and protein-protein interaction[29-32]. It could be used in different ways as a standalone tool to study the structure and thermodynamics of biosystems[33] or to map interaction between macromolecules[34] as well as in combination with other approaches[35,36]. It has been successfully incorporated into the well-known computational chemistry packages (ADF, AMBER, NWChem, Quantum Espresso, AutoDock, MOE, etc.). There was also a number of attempts of its application to the docking procedure[37] and fragmented drug design[29] that were not frequently used in applications, probably due to high computational cost and a need for the specific expertise. In particular, dealing with complex solvent mixtures requires a large amount of memory and computation time, which makes it difficult to apply to morphology in protein complex formation. In this work we use a new version of 1D/3D-RISM software called "RISM for HPC"[38] which realizes our methodology developments[39] which requires a combination of different types of OpenMPI parallelization algorithms to solve a system of integral equations. The software allows us to handle the complex solvent mixtures including water, ions, and organic co-solvents. The full description will be published elsewhere. In a broader prospective we also intend to use it to quickly estimate the role of particular mutations in the virulence of the coronavirus.

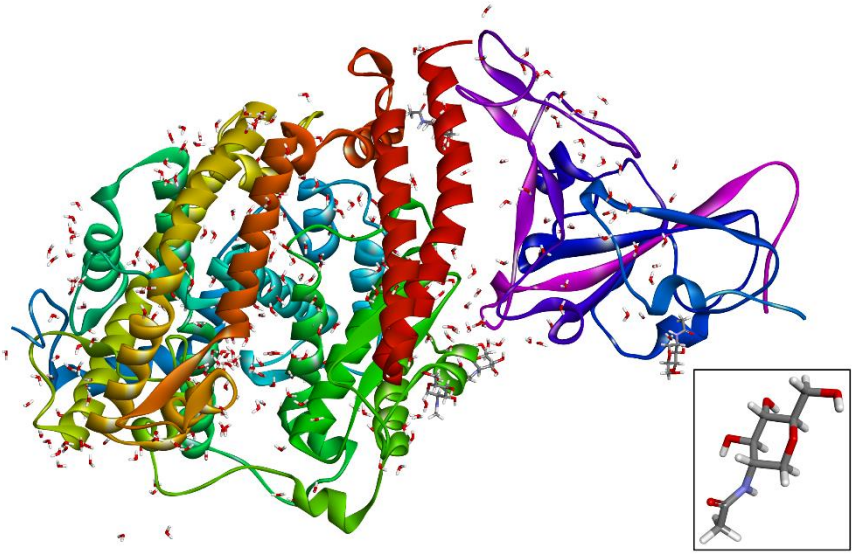

Figure 1. A sketch of a fragment of ACE2 (left) and SARS-CoV-2 (right) proteins in close proximity to each other, based on the experimental measurements reported in the literature and the protein databank. One can also notice the presence of water and small organic molecules (drawn as sticks). A small insert in the bottom right is a magnified sketch of the molecule, with the conventional use of gray color to show carbon atoms, white for hydrogen, red for oxygen and blue for nitrogen.

The remaining part of the manuscript is organized as follows. Section 2 is dedicated to the modeling approach. This is done without going into description of the methodology components except of its basics and important aspects. Instead, we refer to already published studies where the methodology is provided in detail[25-27] and successfully tested on a number of various systems[27, 40-42] including proteins. In Section 3 we describe the system preparation and computation details. In Section 4 we present and discuss the results of the modeling for studied systems. This is done basically through the analysis of 3D solvation shells of solvent interaction sites around solute structures and a tendency for some thermodynamic quantities. Conclusions are found in Section 5 and followed by acknowledgements and references.

2. Method

In statistical-mechanical based approach, the solvation structure is represented by the probability density $\rho_\gamma g_\gamma(\mathbf{r})$ of finding interaction site γ of solvent molecules at 3D space position $\mathbf{r}$ around the solute molecule, as given by the average number density ρ_γ in the solution bulk times the 3D distribution function of solvent interaction sites around the solute molecule $g_\gamma(\mathbf{r})$. The former is determined from the 3D-RISM integral equation[26,27]

$$h_\gamma(\mathbf{r}) = \sum_\alpha \int d\mathbf{r}' c_\alpha(\mathbf{r}-\mathbf{r}') x_{\alpha\gamma}(r'), \quad (1)$$

where $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ are respectively the 3D total and direct correlation functions of solvent site γ around the solute molecule, while $\chi_{\alpha\gamma}(r)$ is the site-site solvent susceptibility and is an input from 1D-RISM theory[25], indices α and γ enumerate all interaction sites on all solvent species. The density distribution function and the total correlation function are related as $g_\gamma(\mathbf{r})=h_\gamma(\mathbf{r})+1$. This gives $h_\gamma(\mathbf{r})$ the meaning of the normalized probability density of 3D spatial correlations between the solute and the solvent molecules, or normalized deviations of solvent site density around the solute molecule from its average value in the solution bulk. For the direct correlation function, the leading term of its asymptotic expansion beyond the short-range region of the solute-solvent repulsive core and attractive well, i.e., the first solvation shell, is given by the 3D interaction potential $u_\gamma(\mathbf{r})$ between the whole solute molecule and solvent interaction site γ, scaled by the Boltzmann factor k_B times temperature T, $c_\gamma(\mathbf{r})\sim -u_\gamma(\mathbf{r})/k_BT$. Inside the repulsive core $c_\gamma(\mathbf{r})$ strongly deviates from its asymptotic form and assumes the values related to the solvation free energy of the solute molecule plunged into the solvent. To be solved, the 3D-RISM equation must be complemented with one more relation, called closure, to reduce the number of the unknown functions. Over the years of work with a number of macro and bio molecules, of practical interest is the so called KH closure[26] of the form

$$g_\gamma^{\mathrm{uv}}(\mathbf{r}) = \begin{cases} \exp\{-\beta u_\gamma^{\mathrm{uv}}(\mathbf{r}) + h_\gamma^{\mathrm{uv}}(\mathbf{r}) - c_\gamma^{\mathrm{uv}}(\mathbf{r})\} & \text{for} \quad d_\gamma^{\mathrm{uv}}(\mathbf{r}) \le 0, \\ 1 - \beta u_\gamma^{\mathrm{uv}}(\mathbf{r}) + h_\gamma^{\mathrm{uv}}(\mathbf{r}) - c_\gamma^{\mathrm{uv}}(\mathbf{r}) & \text{for} \quad d_\gamma^{\mathrm{uv}}(\mathbf{r}) > 0, \end{cases} \tag{2}$$

which combines the HNC closure applied to the regions of density profile depletion and the linearization of the whole exponent reducing to the MSA closure applied to the regions of enrichment. Here $\beta=1/k_BT$, superscripts "u" and "v" are used to distinguish between the solute and solvent, respectively, and $d_\gamma^{\mathrm{uv}}(\mathbf{r})$ is introduced to abbreviate the expression in the exponent. The KH closure is known to appropriately describe such features of the distribution functions as long-range enhancement tails for the critical regime and high peaks for association effects in molecular fluids in a wide density range from gas to liquid. In addition, the HNC counterpart eliminates the substantial drawback of the MSA producing negative peaks of the distribution functions at high densities and strong association. Once the 3D-RISM equation is solved, the result can be used to calculate the solvation thermodynamics. In particular, the solvation free energy or excess chemical potential (ECP) $\Delta\mu$ for the solute molecule is obtained as

$$\Delta\mu = \sum_\gamma \int_V d\mathbf{r}\rho_\gamma k_B T \left[\tfrac{1}{2}h_\gamma^2(\mathbf{r})\theta\left(-h_\gamma(\mathbf{r})\right) - \tfrac{1}{2}h_\gamma(\mathbf{r})c_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})\right], \tag{3}$$

where the sum goes over all sites of all solvent species, integration is carried over the whole accessible volume V, and $\theta(x)$ is the Heaviside step function. ρ_γ denotes the average density of solvent site γ in bulk. The integrand in the expression above is interpreted as the solvation free energy density coming from interaction site γ of solvent molecules around solute. On the other hand, the direct correlation function inside the repulsive core is related to the free energy of creation of a cavity to accommodate the solute excluded volume. In this way, the integrand characterizes the intensity of effective solvation forces in different 3D spatial regions of the solvation shells and indicates where they contribute more/less to the entire solvation free energy. The 3D-RISM equation, the KH closure, and the excess chemical potential are key expressions in the present work. The remaining part of the article is based on their application to the specific biomolecular system and to the analysis of the results. In addition, an interaction between two biomolecules can be characterized by the free energy changes as a function of reaction coordinates, such as the distance between them r_{12} and relative orientation Ω_2 (the orientation of molecule 1 is fixed). In the gas phase, the free energy would simply be an interaction pair potential U_{12} and, in liquid phase, the concept of the potential of mean force (PMF) W is useful. It incorporates solvent effects as well as the intrinsic interaction between the two molecules and describes an average over all the conformations of the surrounding solvent molecules. This can be done in a variety of ways. For example, in statistical mechanics the PMF can be typically expressed as a first guess in terms of a distribution function[43] as $W(r_{12}, \Omega_2) = -k_BT\ln[g(r_{12}, \Omega_2)]$, where, as before, r_{12} is the distance between molecules and Ω_2 is their relative orientation. However, in this work we employ a simpler representation of the PMF and use the separation between two subsystems as the reaction coordinate[34]. This way allows one to directly represent the process of binding the Spike to the ACE2 protein by mapping the PMF landscape for a sequence of arrangements.

System Setup

The three-dimensional structure of the target system represented by RBD of spike glycoprotein of SARS-CoV2 and ACE2 receptor complex with resolution 2.45 Å was obtained from Research Collaboratory for Structural Bioinformatics – Protein Data Bank (RCSB – PDB, ID: 6LZG). The system was retrieved from PDB and prepared further by inserting missing loops, protonation and assigning CHARMM force fields with Momany-Rone partial charges, as it is implemented in "Discovery Studio 2020".[44] Next the resulting structure was minimized with Generalized Born using Molecular Volume (GBMV) solvation model[45] as it was implemented into "Discovery Studio 2020".

The 3D-RISM calculations were performed for a system of SARS-CoV-2 and ACE2 in water with add-ons of simple ions Na^+ and Cl^- and small organic molecules of the types of N-acetyl-*D*-glucosamine (NAG), the difference is that one OH group on the ring is missing, as reflected in the pdb-file, both at the infinite dilution, with assigning the CHARMM force field to the molecule and halide and alkali metal cation potential parameters for ions[46]. In other words, a system where solute is a complex between SARS-CoV2 and ACE2, while solvent is water, simple ions, and drug-like molecule. The NAG molecule is known for influential structural roles that it plays at the cell surface. Some studies[47] name it an important agent for treatment of cell-surface proteins and signal transferring. The water model used in all calculations is the TIP3P model[48]. The site-site solvent susceptibility $\chi_{\alpha\gamma}(r)$ was obtained by calculating the corresponding 1D-RISM equation in which dielectric properties of

the solvent are taken into account through the so-called dielectrically consistent version of the theory known as DRISM.[49,50] In numerical aspect 1D-RISM equation was discretized on a uniform radial grid of 16384 nodes with 0.05 Å resolution and converged to a numerical solution by using the modified direct inversion in iterated space (MDIIS) accelerated numerical solver[26]. The result was used as an input to converge 3D-RISM equations in a 3D cubic box of the size 128 × 128 × 256 Å discretized in each dimension on a uniform grid of 256 × 256 × 512 nodes with 0.5 Å resolution and converged to a numerical solution by MDIIS. Both 1D- and 3D-RISM equations were solved with the KH closure. Finally, the excess chemical potential was estimated by applying to its usual expression a correction in a form of a linear function of the partial molar volume (PMV)[51]. Similar setups for RISM or its modifications have been successfully applied in our previous studies of systems composed of both organic and inorganic molecules in solutions to describe the structure, explain the behavior or to predict thermodynamic dependent properties[40-42].

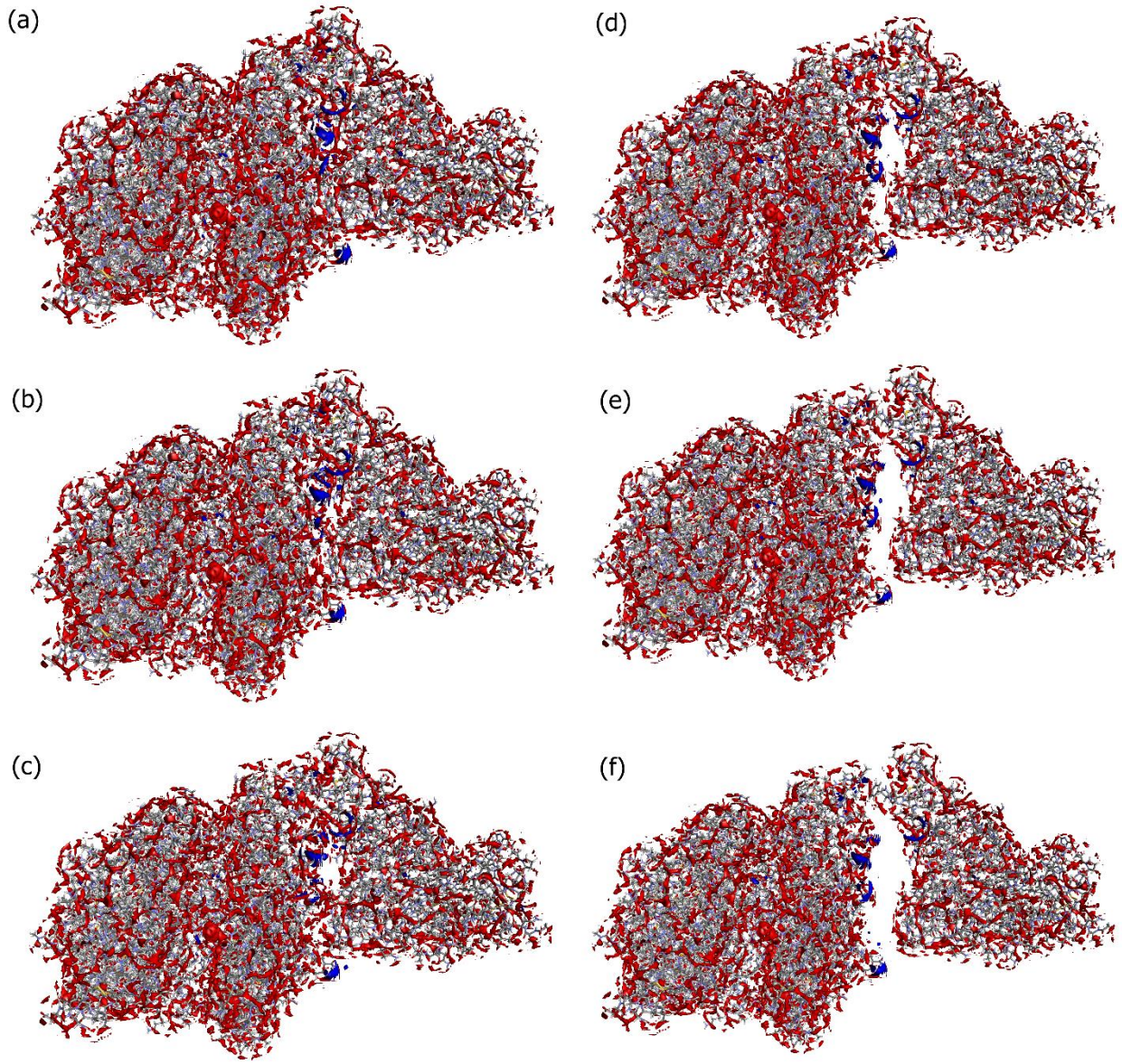

Figure 2. Isosurfaces of 3D disctribution of oxygen of water (red) and nitrogen of *N*-acetyl-*D*-glucosamine (blue) around proteins SARS-CoV-2 and ACE2 (sticks) at a series of separations between them: (a) corresponds to 0 Å distance, (b) 2 Å, (c) 4 Å, (d) 6 Å, (e) 8 Å, and (f) 10 Å. Isosurface values are 3.5 for oxygen and 2.5 for nitrogen. In both cases the values are close to their maximum for the current 3D-RISM solution which means that they show the most probable positions of the respected solvent sites around the proteins.

Results and Discussion

As an example of data which can be assessed by 3D-RISM approach we first present distribution of the water oxygen (red) around the contacting surfaces for different intermolecular distances between the corresponding proteins, Figure 2. Because of the relatively high isosurface value (3.5) the distribution of oxygen atoms shows only preferential mutual orientations of water molecules with interaction sites of proteins while distribution with this high value is not seen in the bulk. The same figure shows distribution of the nitrogen (represented in blue) present in NAG. It is interesting to observe that with high probability the presence of the NAG molecule is identified mostly in the area where protein surfaces come to the contact with each other, irrespectively of the protein separation. Some studies suggest that such "glycosylation-like" molecules play an important role in protein-ligand interactions as well as shielding the spike protein surface[52]. However, recent MD simulation[53] did not find any strong interaction between the NAG and the Spike. The 3D-distributions in Figure 2 are in agreement with conclusions of this work as they show that the NAG molecule prefers remaining close to the ACE2 receptor.

In Figure 3 we present the distribution of Cl^- and Na^+ ions for the same protein separations. Once again, to underline the most favorable distribution of ions we plot the corresponding isosurfaces for a comparably high value (6.5). An interesting effect is that the positive ions are mostly located around connecting edges of interacting surfaces for close distances between proteins, and remain on their positions on these surfaces when the separation increases. This effect suggests a potentially important role of simple cations in the binding recognition process of the Spike protein and the ACE2 receptor and very well can be a subject for a separate study.

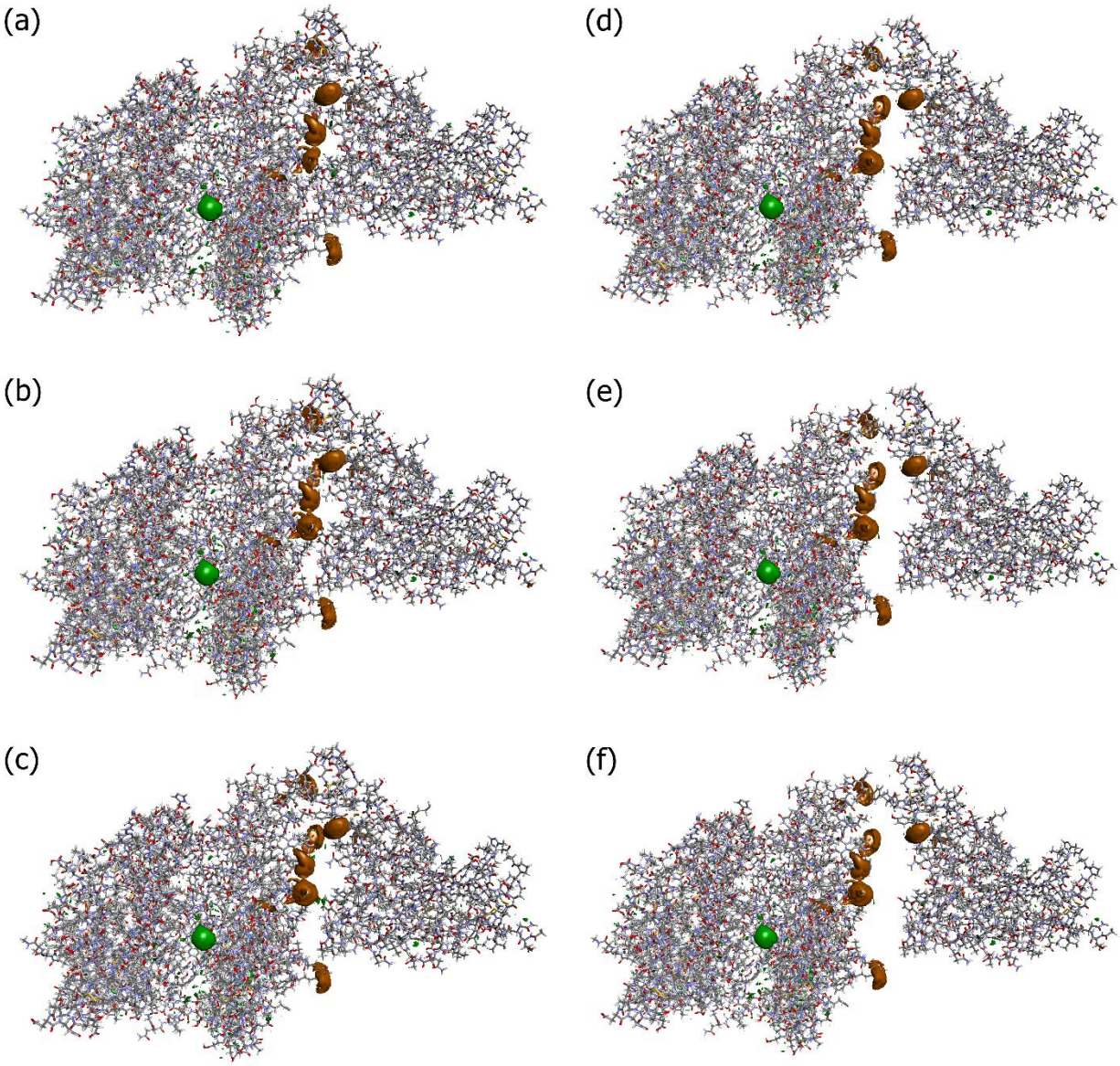

Figure 3. Same as in Figure 2 except that isosurfaces of 3D-distributions are for Na^+ (brown) and Cl^- (green) ions in solution around proteins. For both cases, the isosurface values are 6.5.

According to the currently accepted mechanism of viral protein recognition, the SARS-CoV-2 particles enter human cells by interacting with the ACE2 receptor expressed on epithelial cells in the respiratory tract, however not all the details of this process are fully understood. The hypothesis is that SARS-CoV-2 RBD binds to ACE2 receptors activating it and then entering the cell. After membrane fusion, the SARS-CoV-2 RNA is delivered into the cytosol and the virus starts the replication cycle[54]. Typically, the binding ability is facilitated by hydrogen bonding interactions, salt bridges, hydrophobic and electrostatic interactions. With this respect, Figures 2 and 3 can be regarded as starting points for the analysis of hydrogen bonding and salt/ion bridges: by zooming into the 3D distributions the bonds and the bridges can be identified.

The analysis of hydrophobic interactions is conducted in a similar manner, only that it involves not individual interaction sites of the proteins but also entire amino acid residues. Protein-protein interactions are difficult to modulate with small molecules as the corresponding interacting surfaces are relatively large and flat, with a tendency to lack of deep and well-defined pockets that could be easy target for docking protocol[55-57]. Moreover, some proteins have a very high number of interactions (which is probably the case of Spike-ACE2 interaction) but some of them have only a few and so it is very important to study the energetics and redistribution of the environment along whole surface of interaction (=nanomorphology) by approaching the interacting surfaces.

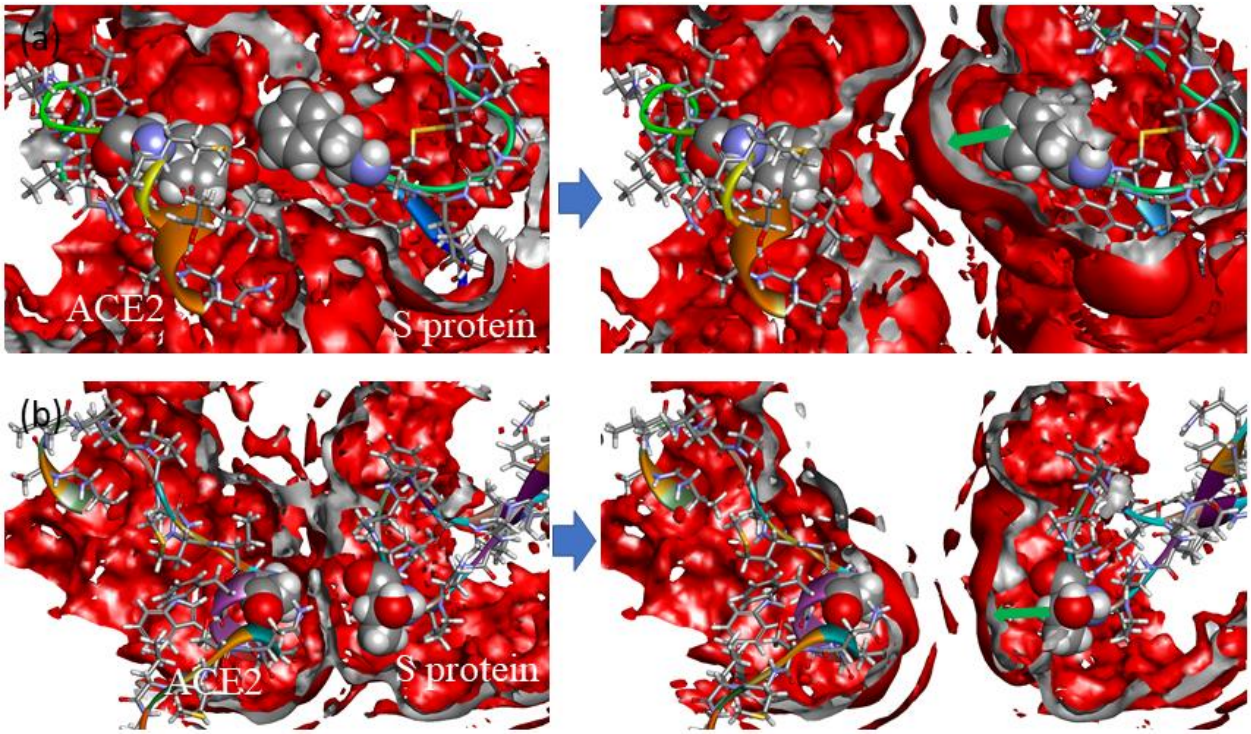

Figure 4. 3D distribution of oxygen of water around fragments of ACE2 and a spike of Sars-CoV-2 in fixed positions at two different separations of proteins: 4 Å (left column) and 14 Å (right column); the isosurface value is 1.3 in all cases. Protein fragments are drawn as ribbon sketches and sticks, selected amino acid residues are displayed as Van der Waals surfaces. Subfigures on top show it for the pair of amino acids TYR83 (ACE2) and PHE486 (spike) while subfigures on bottom show it for the pair GLY326 (ACE2) and THR500 (spike). Green arrows in the right part of the figure direct into a position of first solvation peak around residues. The intensity of the peak is estimated for about 25-30% weaker compared to non-hydrophobic interactions, and its position along the direction appears to be pushed away for about 0.5 Å.

Referring to some recently published studies[58,59] the hydrophobic interactions could play a dominant role on viral recognition and, thus these interactions could define the unusually strong attachments exerted by the SARS-CoV-2 virus. Hydrophobic interactions assume that the protein seeks to decrease in external hydrophobic surface and reduce the undesirable interactions with water[60]. It has been found that the geometry of the water hydrogen bond network within solvation layers differs from that observed with pure water, and this is the result of interaction of water molecules with protein surfaces[61]. When this happens, an unoccupied space between the hydrophobic surface and neighboring solvation layer occurs. The thickness of this region depends on local structure of the water-protein interface

which is a result of maintaining a balance between water-surface interactions and water-water interactions. An existence of this region is one of the main factors that differentiates the hydrophobic hydration from hydration and can be estimated through the change of PMV with geometrical structure[61]. Its presence also explains why the solvent density is greater around the native form of the protein compared to that in the vicinity of the hydrophobic surface and allows to predict the binding affinity[62]. The 3D-RISM theory allows us to directly analyze the solvent distribution as well to estimate the values of PMV for different configurations[63]. In Figure 4 we illustrate the effect of hydrophobic interactions by showing the 3D-distribution of water oxygen around selected pairs of amino acids of the Spike and ACE2 proteins that contribute to the hydrophobic interactions[59]. The first solvation shell between interacting amino acids is hardly visible in the case of a short separation of proteins, and is clearly seen around them when the separation increases, confirming the presence of an excluded volume at short distances. This is accompanied by the change in PMV. Its value for the bound state is estimated at 106164 Å^3 and increases to 106297 Å^3 when the separation reaches 4 Å. After this separation water starts penetrating into the cavity between the residues as it seen in Figure 4. A brief physical explanation of the depleted region of water density at the surface can be explained as follows. Hydrophobic residues are sequestered from water as a result of the unfavorable hydration of these nonpolar groups which minimizes the amount of surface exposed to water[64]. This is in contrast to polar (or charged) amino acids generally forming solute-solvent hydrogen bonds with adjacent waters which lowers their PMV in hydration layer. The nonpolar residues do not form hydrogen bonds which increases the PMV. When two hydrophobic solutes come close together, water molecules are released from the region between two surfaces due to the constraints on the hydrogen bond network. The release of water from the intervening space creates an imbalance in pressure that provides long-range attraction between hydrophobic surfaces and lowering of PMV of the bonded complex[65].

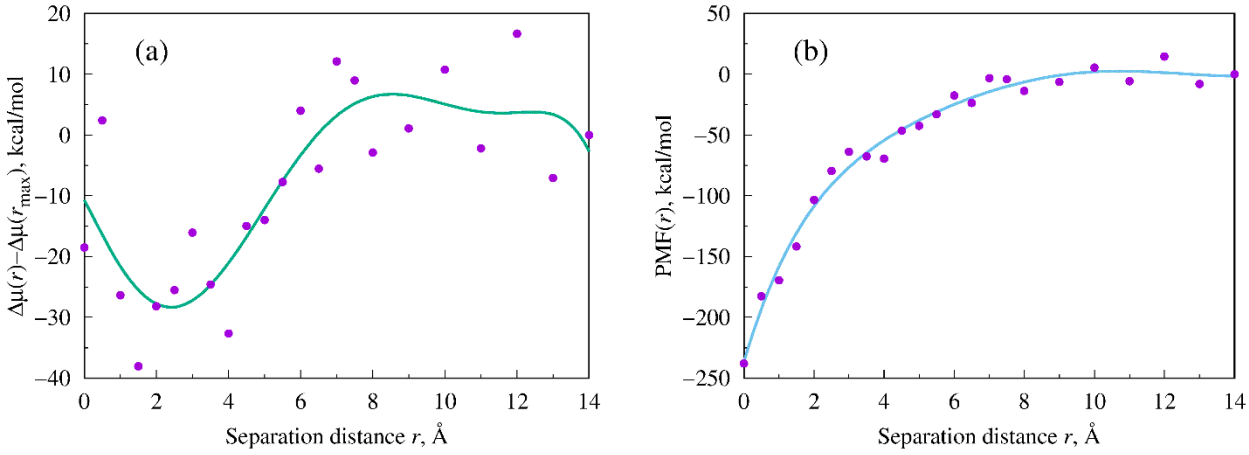

Figure 5. ECP (a) and PMF (b) as a function of separation distance between SARS-CoV-2 and ACE2 relatively to the initial orientation of proteins. The value r_{max} corresponds to maximum separation. Dots show the 3D-RISM calculation results for a series of discrete separations and lines are used to plot the interpolation curve with a simple polynomial of degree n=6.

From a thermodynamic point of view, a strong correlation between the free energy of hydration and the number of water molecules packed around a hydrophobic compound has been reported in the literature[66]. Using the RISM approach, we calculated the solvation free energy, also called ECP. In Figure 5 we present ECP (a) and PMF (b) for a series of separations between proteins. To avoid possible confusions, it should be noted that protein separations are relative to the initial position and are not absolute distances between the objects. The presence of a broad minimum of about -30 kcal/mol deep in ECP at short separations, Figure 5 (a), supports the idea about a significant role the hydrophobic hydration plays in the mechanism of the proteins binding, and the downslope shape of PMF with a deep minimum of about -200 kcal/mol at initial separation, Figure 5 (b), explains the intensity of such binding. As we can see, in our calculations the 3D-RISM contribution to PMF from expression (3) is about -30 kcal/mol. The remaining portion comes from VdW and Coulomb interactions which strongly depend on the force field used (CHARMM in our case) and are not directly related to the presented approach. Still, in numerical aspect our results are in agreement with recently published study[67] reporting comparable values. In particular, by using an alternative molecular mechanics approach (MD simulation and docking) the authors estimate the binding affinity for SARS-Cov-2 - ACE2 around -140±10 kcal/mol, with the hydrophobic contribution estimated at -38±3 kcal/mol. Although we have calculated the potentials for separations along one randomly chosen direction it becomes clear that any attempt to reverse the process of binding will require much energy. A feasible solution to this problem consists of using ligands.

Conclusions

We have illustrated and tested a modeling approach capable of studying the nanomorphology and binding of SARS-CoV-2 spike protein with the ACE2 receptor. The core part of this approach is the integral equation theory of molecular liquids in the Three-Dimensional Reference Interaction Site Model – 3D-RISM approximation. This approach has been employed to calculate the nanomorphology of solvation shells and provided 3D distribution functions for all interaction sites of the complex solvent (composed of water, simple ions, and drug-like molecule) around the solute (Spike and ACE2 proteins at varying distances) taking into account their chemical composition and thermodynamic conditions. By visual analysis of 3D distributions at different separations between proteins we were able to track the changes in preferential orientations of the

solvent molecules (water) in such a way that we could identify the role it plays in the protein binding mechanism. According to the current understanding, this process takes place by hydrogen bonding, electrostatic interactions, and ionic bridges, but hydrophobic contacts play a central role in anchoring RBD of Spike protein to the ACE2 Receptor. Our calculations are in agreement with these conclusions.

In addition to nanomorphology, the mapping of reaction coordinates by 3D-RISM has provided an important information about thermodynamics of binding the Spike to ACE2 which is difficult to reach by MD simulation. This includes the information about the shape and deepness of the ECP and PMF and has direct connection to virulence of the particular strain, because the virulence of different Spike mutations can be estimated by simple comparison of the corresponding PMFs. In our study both ECP and PMF have shapes showing a minimum at short separations and reaching a plateau on bigger distances. The presence of the minimum explains why the proteins bind and supports the idea about the dominant role of hydrophobic interactions in this process. In methodological aspect, demonstrated examples directly show the capability of the presented approach for the analysis of the effect of ligand type molecules. By simple replace of the NAG molecule with a drug candidate or by adding prospective ligands as the part of solvent with small concentrations, the approach is resourceful for accurate screening of potential drugs. Compared to docking the methodology is more accurate[68] and allows one to understand the mechanism of Covid-19 penetration through the cell membrane and ways of inhibition. It means that the approach described in this work could be applied, with minor modifications, to study the effects of viral mutations detected in new strains worldwide. Our protocol could quickly and accurately help in the identification of drug molecules with inhibitory properties of the new viral strains This is especially important for the case of protein binding driven by hydrophobic interactions where interacting surfaces are large and do not have specific “hot spots” which could be studied by docking[58]. This can be a good addition to machine learning approaches used for selections of drug candidates but they do not describe mechanisms[69].

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Mr. Renato Caldart, Senior Advisor, International Relations, National Research Council Canada (NRC), for his advice and kind support for this study. We also acknowledge the support from the broader NRC and RIKEN collaboration promotion teams in organizing the High-Performance Computing Workshop that inspired new ideas and created new connections for the authors. G.S. acknowledges the BIOVIA Inc. for providing six-month license to “Discovery Studio”[68,69] for SARS-CoV-2 related studies.

1 WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020.
2 D. Cucinotta and M. Vanelli, *Acta Biomedica*, 2020, **91**, 157-160.
3 B. S. Graham, *Science*, 2020, **368**, 945-946**.**
4 J. H. Beigel, K. M. Tomashek, L. E. Dodd et al, *N. Engl. J. Med.*, 2020, **383**, 1813-1826.
5 R. T. Eastman, J. S. Roth, K. R. Brimacombe, A. Simeonov, M. Shen, S. Patnaik and M. D. Hall, ACS Central Science, 2020, 6, 672-683.
6 B. M. Tomazini, I. S. Maia, A. B. Cavalcanti *et al*, *JAMA*, 2020, **324**, 1307-1316.
7 M. Hoffmann, H. Kleine-Weber, S. Schroeder *et al*, *Cell*, 2020, **181**, 271-280.
8 P. Zhou, X. L. Yang, X. G. Wang *et al*, *Nature*, 2020, **579**, 270-273.
9 A. Spinello, A. Saltalamacchia and A. Magistrato, *J. Phys. Chem. Lett.*, 2020, **11**, 4785-4790.
10 Y. Wang, M. Liu and J. Gao, *Proc. Nat. Acad. Sci.*, 2020, **117**, 13967-13974.
11 C. S. Lupala, X. Li, J. Lei, H. Chen, J. Qi, H. Liu and X. Su, bioRxiv, 2020.03.24.005561.
12 E. P. Barros, L. Casalino, Z. Gaieb *et al*, *Biophysical Journal*, 2021, **120**, 1-13
13 R. Li, H. Li and W. Chen, *Biophysical Chemistry*, 2020, **267**, 06472.
14 B. Dehury, V. Raina, N. Misra and M. Suar, *Journal of Biomolecular Structure and Dynamics*, 2020, **6**, 1-16
15 J. T. Ortega, M. L. Serrano, F. H. Pujol and H. R. Rangel, *EXCLI J.*, 2020, **19**, 410-417.
16 A. Choudhury and S. Mukherjee, *J. Med. Virol.*, 2020, **92**, 2105- 2113.
17 M. Hakmi, E. L. M. Bouricha, J. Akachar, B. Lmimouni, J. E. Harti, L. Belyamani and A. Ibrahimi, *Journal of Biomolecular Structure and Dynamics*, 2020, **8**, 19
18 A. Basu, A. Sarkar and U. Maulik, *Sci; Rep.*, 2020, **10**, 17699.
19 C. G. Benítez-Cardoza and J. L. Vique-Sánchez, *Life Sciences*, 2020, **256**, 117970.
20 P. N. F. Souza, F. E. S. Lopes, J. L. Amaral, C. D. T. Freitas and J. T. A. Oliveira, *Int. J. Biological Macromolecules*, 2020, **164**, 66-76.
21 G. M. Verkhivker, *J. Proteome Research*, 2020, **19**, 4587-4608.
22 L. Casalino, Z. Gaieb, J. A. Goldsmith *et al*, *ACS Central Science*, 2020, **6**, 1722-1734.
23 A. Ali and R. Vijayan, *Sci. Rep.*, 2020, **10**, 14214.
24 S. A. Hollingsworth and R. O. Dror, *Neuron*, 2018, **99**, 1129-1143.
25 F. Hirata. Theory of Molecular Liquids. In: F. Hirata (ed.), Molecular Theory of Solvation. Understanding Chemical Reactivity, vol. 24, pp. 1-60 (Kluwer, Dordrecht, 2003).
26 A. Kovalenko. Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces. In: F. Hirata (ed.), Molecular Theory of Solvation. Understanding Chemical Reactivity, vol 24, pp. 169-275 (Kluwer, Dordrecht, 2003).
27 A. E. Kobryn, S. Gusarov and A. Kovalenko, *J. Phys.: Condens. Matter*, 2016, **28**, 404003.
28 B. Olson, A. Cruz, L. Chen et al, *J Comput. Aided Mol. Des.*, 2020, **34**, 1219-1228.
29 T. Imai, K. Oda, A. Kovalenko, F. Hirata and A. Kidera, *J. Am. Chem. Soc.*, 2009, **131**, 12430-12440.
30 Y. Kiyota, N. Yoshida and F. Hirata, *J. Chem. Theory Comput.*, 2011, **7**, 3803-3815.
31 Y. Maruyama, N. Yoshida, H. Tadano, D. Takahashi, M. Sato and F. Hirata, *J. Comput. Chem.*, 2014, **35**, 1347-1355.
32 N. Yoshida, M. Higashi, H. Motoki and S. Hirota, *J. Chem. Phys.*, 2018, **148**, 025102.
33 T. Imai, A. Kovalenko and F. Hirata, *Molecular Simulation*, 2006, **32**, 817-824.

34 P. Drabik, S. Gusarov and A. Kovalenko, *Biophysical Journal*, 2007, **92**, 394-403.
35 S. Gusarov, T. Ziegler and A. Kovalenko, *J. Phys. Chem. A*, 2006, **110**, 6083-6090.
36 A. Kovalenko and S. Gusarov, *Phys. Chem. Chem. Phys.*, 2018, **20**, 2947-2969.
37 D. Nikolic, N. Blinov, D. Wishart and A. Kovalenko, *J. Chem. Theory Comput.*, 2012, **11**, 3356-3372.
38 S. Wilson, A. E. Kobryn and S. Gusarov, Invention Disclosure "RISM for HPC". National Research Council Canada, 2019.
39 S. Gusarov, B. S. Pujari and A. Kovalenko, *J. Comput. Chem.*, 2012, **33**, 1478-1494
40 A. E. Kobryn, S. Gusarov and K. Shankar, *Polymers*, 2016, **8**, 136.
41 A. E. Kobryn, S. Gusarov and K. Shankar, *J. Molecular Liquids*, 2019, **289**, 110997.
42 A. Kovalenko, A. E. Kobryn, S. Gusarov, O. Lyubimova, X. Liu, N. Blinov and M. Yoshida, *Soft Matter*, 2012, **8**, 1508-1520.
43 J.-P. Hansen and I. R. McDonald, Theory of Simple Liquids: With Applications to Soft Matter, 4th ed. (Elsevier, Amsterdam, 2013).
44 Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2020, San Diego: Dassault Systèmes, 2019.
45 M. S. Lee, F. R. Salsbury and C. L. Brooks, *J. Chem. Phys.*, 2002, **116**, 10606-10614.
46 S. H. Lee and J. C. Rasaiah, *J. Phys. Chem.*, 1996, **100**, 1420-1425.
47 J. B. Konopka, *Scientifica*, 2012, 489208.
48 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
49 J. S. Perkyns and B. M. Pettitt, *Chem. Phys. Lett.*, 1992, **190**, 626-630.
50 J. Perkyns and B. M. Pettitt, *J. Chem. Phys.*, 1992, **97**, 7656-7666.
51 D. S. Palmer, A. I. Frolov, E. L. Ratkova and M. V. Fedorov, *J. Phys.: Condens. Matter*, 2010, **22**, 492101.
52 B. Robson, *Computers in Biology and Medicine*, 2020, **121**, 103749.
53 J. Lan, J. Ge, J. Yu *et al*, *Nature*, 2020, **581**, 215-220.
54 M. Hoffmann, H. Kleine-Weber and S. Pöhlmann, *Molecular Cell*, 2020, **78**, 779-784.
55 D. Bojadzic and P. Buchwald, *Curr. Topics Med. Chem.*, 2018; **18**, 674-699.
56 H. Lu, Q. Zhou, J. He *et al*, *Signal Transduction and Targeted Therapy*, 2020, **5**, 213:1-23.
57 P. Walter, J. Metzger, C. Thiel and V. Helms, *Plos One*, 2013, **8**, 58583:1-12.
58 J. Li, X. Ma, S. Guo *et al*, *Global Challenges*, 2020, **4**, 2000067.
59 E. Taka, S. Z. Yilmaz, M. Golcuk, C. Kilinc, U. Aktas, A. Yildiz and M. Gur, bioRxiv, 2020.09.21.305490.
60 R. H. Garrett and C. M. Grisham, Biochemistry, 4th ed. (Brooks/Cole, Boston, 2010).
61 A. Kuffel and J. Zielkiewicz, *J. Phys. Chem. B*, 2012, **116**, 12113-12124.
62 A. Szymaniec-Rutkowska, E. Bugajska, S. Kasperowicz, K. Mieczkowska, A. M. Maciejewska and J. Poznański, *J. Molecular Liquids*, 2019, **293**, 111527.
63 T. Imai, S. Ohyama, A. Kovalenko and F. Hirata, *Protein Sci.*, 2007, **16**, 1927-1933.
64 K. A. Dill, *Biochemistry*, 1990, **29**, 7133-7155.
65 K. Lum, D. Chandler and J. D. Weeks, *J. Phys. Chem. B*, 1999, **103**, 4570-4577.
66 R. B. Hermann, *J. Phys. Chem.*, 1971, **75**, 363-368.
67 A. Ali and R. Vijayan, *Scientific Reports*, 2020, **10**, 14214:1-12.
68 V. K. Hinge, N. Blinov, D. Roy, D. S. Wishart and A. Kovalenko, *J. Comput. Aided Mol. Des.*, 2019, **33**, 913-926.
69 S. Gusarov and S. Stoyanov, *J. Phys. Chem. Lett.*, 2020, **11**, 21, 9408-9414