The inhibitory effect of a Corona virus spike protein fragment with ACE2

Emanuel K. Peter[1] and Alexander Schug[1,*]

[1]John von Neumann Institute for Computing and Julich Supercomputing Centre, Institute for Advanced Simulation, Forschungszentrum Jülich, 52425 Jülich, Germany

December 18, 2020

Abstract

In this paper, we investigate the binding processes of a fragment of the Coronavirus spike protein receptor binding domain (CoV RBD), the hexapeptide YKYRYL on the angiotensin converting enzyme 2 (ACE2) receptor and its inhibitory effect on the binding and activation of the Coronavirus-2 spike protein receptor binding domain CoV-2 RBD at ACE2. In agreement with an experimental study, we find a high affinity of the hexapeptide to the binding interface between CoV-2 RBD and ACE2, which we investigate using 20 independent equilibrium Molecular Dynamics (MD) simulations over a total of 1 μs and a 200 ns enhanced correlation guided MD (CORE-MD) simulation. We then evaluate the effect of the hexapeptide on the assembly process of the CoV-2 RBD to ACE2 in long-time enhanced CORE-MD simulations. In that set of simulations, we find that CoV-2 RBD does not bind to ACE2 with the binding motif shown in experiments, but it rotates due to an electrostatic repulsion and forms a hydrophobic interface with ACE2. Surprisingly, we observe that the hexapeptide binds to CoV-2 RBD, which has the effect that this protein only weakly attaches to ACE2, so that the activation of CoV-2 RBD might be inhibited in this case. Our results indicate that the hexapeptide might be a possible treatment option which prevents the viral activation through the inhibition of the interaction between ACE2 and CoV-2 RBD.

Statement of Significance

A novel coronavirus, CoV-19 and a later phenotype CoV-2 were identified as primary cause for a severe acute respiratory syndrome (SARS CoV-2). The spike (S) protein of CoV-2 is one target for the development of a vaccine to

prevent the viral entry into human cells. The inhibition of the direct interaction between ACE2 and the S-protein could provide a suitable strategy to prevent the membrane fusion of CoV-2 and the viral entry into target cells. Using MD simulations, we investigate the assembly process of a Coronavirus Spike protein fragment, the hexapeptide YKYRYL on the ACE2 receptor and its inhibitory effect on the aggregation and activation of the CoV-2 spike receptor protein at the same receptor protein.

Introduction

In December 2019, a novel respiratory disease appeared in Wuhan, Hubei, China. Although it is still under debate, there are strong indications that a first cluster of infections occurred at the Huanan seafood market [40, 41, 20]. A novel coronavirus, CoV-19 and a later phenotype CoV-2 were identified as primary cause for a severe acute respiratory syndrome (SARS) [14, 6]. Within few days, the viral disease spread over whole China and within the following weeks, the local epidemic grew to a global pandemic with an exponentially growing infection rate. At present, the number of infected humans reached 3.855.788 with a number of 265.862 deaths associated with SARS-CoV-19 and CoV-2 [38]. This global pandemic will have an unprecedented economic, sociological and political impact, in contrast to prior outbreaks of CoV related SARS epidemics [7]. While a huge number of trials are still ongoing to develop a successful vaccination strategy against CoV-2, a direct medication of infected patients can have the potential to save lives and to stabilize the situation. The spike (S) protein of CoV-2 is the major target for the development of a vaccine or a potential strategy to tackle the viral entry into human cells [8, 11, 39]. The S-protein forms trimers at the protrusions of the virus and comprises two functional subunits : S1 and S2. In the cascade of the viral entry, The S1 unit of the spike (S) protein facilitates the attachment of the virus at the surface of the cell [37]. The S2 subunit, responsible for membrane fusion, employs TMPRSS2 for the S protein priming, while it uses ACE2 as entry receptor for membrane fusion [27, 16, 17, 10, 13, 24]. One of the key factors for its infectious potential for humans is the high conservation of ACE2 in different mammalian organisms [18], which allows its transmission from animals to humans. The receptor binding domain (RBD) of the S1 subunit contains five antiparallel beta strands, while alpha-helical and loop motifs form the connecting entities between the beta sheets. Between two beta sheets, an extended insertion forms the receptor binding motif (RBM), which binds to ACE2 at its N-terminal helix [25, 26, 4]. Among a large number of potential targets, the inhibition of the direct interaction between ACE2 and the S-protein (SARS CoV-S) provides a suitable strategy to prevent the membrane fusion of CoV-2 and the viral entry into human cells [22, 1]. In a combined experimental

and theoretical study, a hexapeptide ((438)YKYRYL(443)) of the receptor domain of SARS-CoV S has been identified as an efficient inhibitor of the interaction between the S-protein and ACE2 [34]. In vitro infection of Vero E6 cells by SARS coronavirus (SARS-CoV) was blocked by the hexapeptide. It also has been shown that the peptide inhibits the proliferation of CoV-NL63. Interestingly, the fragment (438)YKYRYL(443) carries the dominant binding epitope and binds to ACE2 with a high affinity of K(D)=46 μM. Its binding mode was further characterized by saturation transfer difference (STD), NMR spectroscopy and Molecular Dynamics simulations. Based on this information, the peptide can be used as lead structure to design potential entry inhibitors against SARS-CoV and related viruses.

In this article, we present Molecular Dynamics (MD) simulations to investigate the effect of the hexapeptide on the binding of the spike protein receptor binding domain of SARS CoV-2 (CoV-2 RBD) with ACE2 and quantify its affinity to the binding site shown by Struck *et al.* [34]. Second, we applied enhanced correlation guided MD (CORE-MD) simulations to measure the free energy of binding of the hexapeptide to its preferential binding site at ACE2. In the third stage of the study, we investigated the effect of the hexapeptide on the interaction of CoV-2 RBD with ACE2. In our simulations, we observe that the hexapeptide binds to the N-terminal region of ACE2 with a high affinity to 3 clusters that are located at the interface at which CoV-2 RBD binds to the receptor as revealed in X-ray structures [25, 26]. In the enhanced MD simulations, we observe that CoV-2 RBD relaxes into an energy minimum which differs fundamentally from the X-ray structure : We find that CoV-2 RBD rotates and binds to ACE2 at the N-terminal region by a hydrophobic patch, which is between residues Thr351 and Leu535. Our simulations reveal that the energetic minimum does not favour a hydrophilic interaction as shown in X-ray crystallography. In the simulations of binding of CoV-2 RBD to ACE2 in the presence of the hexapeptide, the hexapeptide binds preferentially to CoV-2 RBD in the vicinity to the ACE2 binding segment. Surprisingly, we find that the binding of the hexapeptide changes the assembly process of CoV-2 RBD, such that the activation of ACE2 is inhibited by the hexapeptide. Our simulations are in agreement with the experimental study and demonstrate the potential of the hexapeptide YKYRYL as a possible 'new modality' treatment option [35] which prevents the viral entry into human cells. Due to a damping effect by the cleavage of the peptide by proteases, a chemical modification which hinders that cellular process might increase the therapeutic potential of this peptide [3, 35, 33].

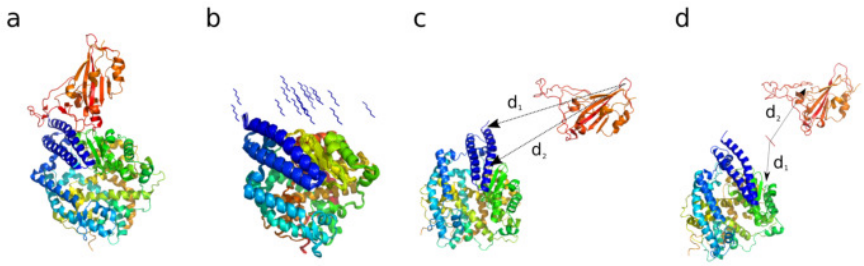

Figure 1: (a) Crystal structure of the SARS CoV-2 spike protein receptor binding domain CoV-2 RBD bound with ACE2 (PDB : 6M0J [25]). ACE2 shown in blue, cyan, yellow and green color, CoV-2 RBD is shown in orange. (b) 20 different hexapeptide (YKYRYL) conformations used as initial starting models of independent 50 ns equilibrium MD simulations in explicit solvent. (c) Starting structure of the enhanced CORE-MD simulation of CoV-2 RBD to ACE2. (d) Starting structure of the enhanced CORE-MD simulation of the same system in the presence of the hexapeptide. The distances d_1 and d_2 used in the free energy projections (see Figure 4) are indicated in the panels (c) and (d).

Methods

Correlation guided dynamics (CORE-MD)

CORE-MD uses the path along the reduced action L_i as function of the momenta p_i and coordinates q_i for an atom with the index i [23, 9, 30, 29, 31] :

$$L_i = \oint p_i dq_i \ , \tag{1}$$

where a finite time summation is applied over the path over momenta and displacements along the trajectory. For the calculation of the momentum $p_i = m_i v_i$, we use a uniform atomic mass. A path-dependent correlation function $C_i(t)$ is calculated as follows :

$$C_i(t) = \frac{1}{\tau} \sum_{t \leq \tau} \frac{(L_i' - \langle L_i \rangle)(L_i - \langle L_i \rangle)}{|L_i' - \langle L_i \rangle||L_i - \langle L_i \rangle|} \ , \tag{2}$$

where $\langle ... \rangle$ denotes the time average, and L_i' is determined at a time t' with a probability $\mathcal{P}_i(t')$:

$$\mathcal{P}_i(t') = \frac{1}{1 + e^{-C_i(t')}}, \tag{3}$$

at each timestep. In order to define a correlation dependent probability density, the space of the correlation function is discretized into a histogram

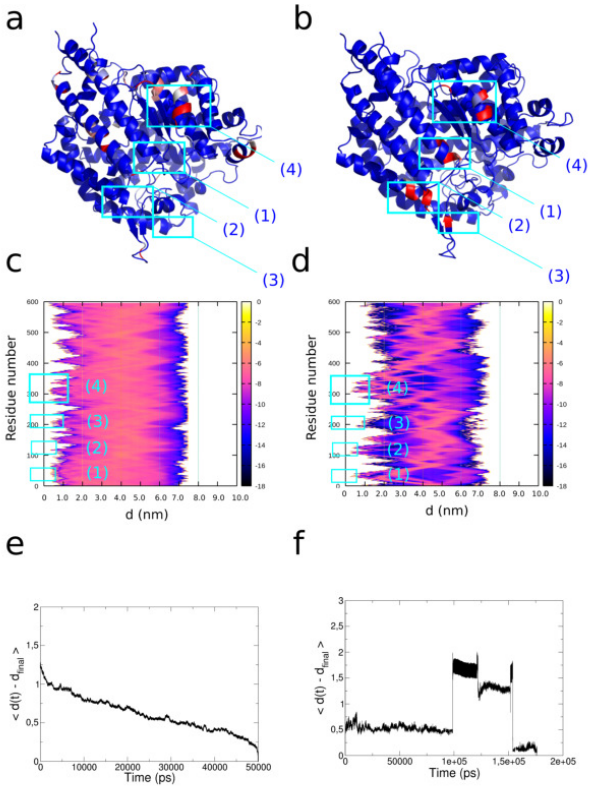

Figure 2: (a, b) Result from 20 equilibrium MD simulations over 50 ns of the hexapeptide in the vicinity of ACE2. (a) Color assigned structure of ACE2 with the affinity of the hexapeptide for regions at ACE2 indexed with a color gradient ranging from 0 (blue) to 100 % (red) corresponding to the highest affinity. (c) Log-plot of the relative affinity ϵ averaged over the set of 20 MD simulations. The affinity expresses the propensity of finding the hexapeptide as function of the residue number and the distance. (b, d) Result from the enhanced CORE-MD simulation over 200 ns of the hexapeptide in the vicinity of ACE2. (b) B-factor assigned structure of ACE2 with the affinity of the hexapeptide for regions at ACE2 indexed with a color gradient ranging from 0 (blue) to 100 % (red). (d) Log-plot of the relative affinity ϵ of finding the hexapeptide as function of the residue number and the distance averaged over the correlation guided MD simulation. (e) Average root mean square deviation of all measured distances of the hexapeptide and ACE2 as function of time taken from the set of 20 independent NPT simulations. (f) Average root mean square deviation of all pairwise distances in the CORE-MD simulation as function of time.

ranging from -1 to +1 and a probability density $\rho_i(t)$ is defined at the time t for the history dependent number of states $N_{C_i(t)}$ and the total number of states in this state of the correlation function $C_i(t)$:

$$\rho_i(t) = \frac{N_{C_i(t)}}{\sum_i N_{C_i(t)}} \ . \tag{4}$$

That definition allows for the discretization of the path-dependent correlation and the definition of a log likelihood function (see below). The correlation dependent density $\rho_i(t)$ as function of the correlation function $C_i(t)$ for an atom with index i is then defined by :

$$\rho_i(t) = \frac{\sum_t \sum_{C_\mu=-1}^{C_\mu=1} e^{-\frac{(C_i(t)-C_\mu)^2}{2\sigma}}}{\sum_i \sum_t \sum_{C_\mu=-1}^{C_\mu=1} e^{-\frac{(C_i(t)-C_\mu)^2}{2\sigma}}} \ , \tag{5}$$

where σ defines the width of the Gaussian function (Due to the fact that we apply a histogram over 10^2 bins, we apply $\sigma = 2 \times 10^{-2}$). Subsequently, a log pseudo likelihood function l of the correlation dependent density is defined, which describes a form of a correlation dependent potential :

$$l_i(t) = -\log(\rho_i(t)) \ , \tag{6}$$

which leads to the corresponding bias A_i with an additional parameter α with the units of an energy :

$$A_i = \alpha \nabla l_i(t) \ , \tag{7}$$

as the derivative along a unit vector with a unit length, due to the dimensionality of the correlation function. As a consequence, the bias gradient evolves as the gradient of the potential of the history dependent probability density $\rho_i(t)$, which is described by the log functional in equation 6. That way, the correlation dependent likelihood is maximized in analogy to the principle of maximum entropy [28].

As a second element of the CORE-MD algorithm, a factorization of the total gradient by a factor r_i is introduced. The application of the bias-gradient using only the bias derived from the path-dependent correlation requires the sufficient sampling of the correlation space. The correlation space of the path-correlation is sampled along a first-order rate equation :

$$\dot{C}_i(t) = -k_{i_1} t \ , \tag{8}$$

where k_1 stands for the first-order rate constant. In order to reach a sufficient sampling efficiency of the correlation space, the resulting gradient is scaled by a correlation dependent factor r, in order to enhance the decay of the auto-correlation and to achieve a faster access of the conformation space. As a consequence of the factorization, the time-dependent behavior

of the correlation function is then described by a second order rate equation :

$$\dot{C}_i(t) = -(k_{i_1} + k_{i_2})t \ , \tag{9}$$

We define the factor $r_i(t)$ as :

$$r_i(t) = e^{-\beta C_i(t)}(1 + \beta C_i(t)) \ , \tag{10}$$

where β stands for a second dimensionless constant. In the global picture, the log likelihood function converges to the global log likelihood of the total correlation dependent density Ξ :

$$\lim_{t\to\infty} l_i(t) = -\log(\Xi_i) \ , \tag{11}$$

where Ξ_i is approximated as the probability function P_i of the path dependent correlation :

$$P_i \approx \Xi_i \ . \tag{12}$$

The last expression shows that the CORE-MD algorithm samples the global free energy in the infinite time limit, due to the defintion of $\Delta F_i = -k_B T \log(P_i)$.

Simulation parameters and system setup

For all simulations, we used the structures of the CoV-2 RBD - ACE2 complex from the PDB structure 6M0J, (chain A - ACE2, chain E - CoV-2 RBD) [25]. For the first set of 20 independent NPT-MD simulations over 50 ns in explicit solvent, we centered the PDB structure of ACE2 (PDB: 6M0J, chain A) in a triclinic box with dimensions 7.419 x 8.361 x 8.614 nm^3 fill the box with 17.026 SPC/E waters. For the preparation of 20 starting structures , we placed the hexapeptide YKYRYL at 20 different initial positions in the vicinity to the potential binding site of ACE2 (see Figure 1 b). In parts, we reduced the accessible conformation space by the placement of the 20 peptide conformations in the vicinity of the CoV-2 RBD - ACE2 interface in the PDB-structure. For the selection of random orientations, larger simulation times and a larger number of starting structures would have to be applied. For the enhanced sampling simulation of the hexapeptide-ACE2 system in implicit solvent using CORE-MD enhanced sampling, we modeled one hexapeptide-ACE2 conformation to and simulated the system over 200 ns using the parameters α =5.0 and β =0.5. For a third and a fourth CORE-MD enhanced sampling simulation in and without the presence of the hexapeptide, we modeled a separated CoV-2 RBD - ACE2 system with an increased contact distance of approximately 2.2 nm by which the two domains are separated from each other. For the separated

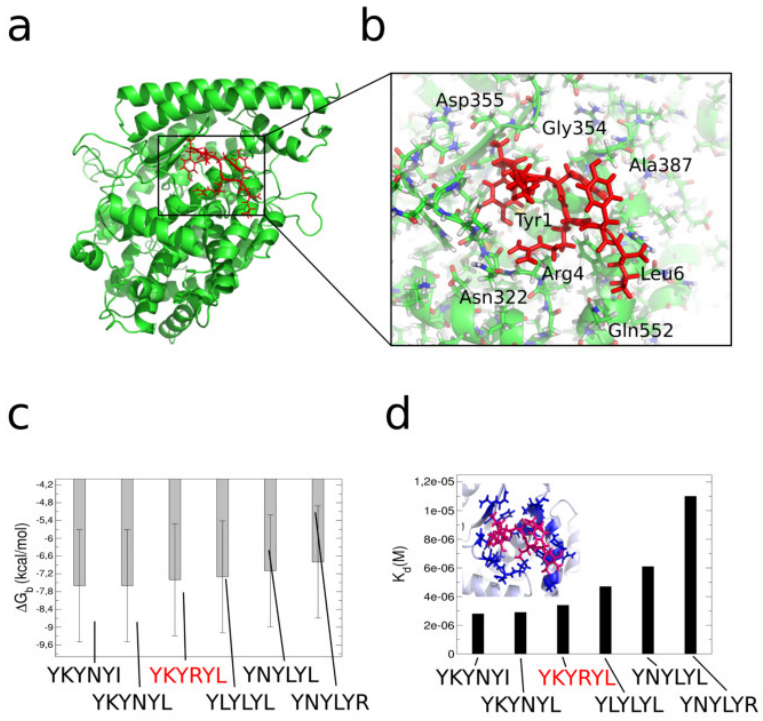

Figure 3: (a) Main conformation of the hexapeptide in complex with the N-terminal helical interface of ACE2 as revealed from equilibrium MD and enhanced sampling MD simulations. (b) Molecular view on the binding site of the hexapeptide at the interface with ACE2. (c) Results from protein binding energy calculations ΔG [36] on the hexapeptide YKYRYL and 5 further variant models in the binding site at ACE2. The relative error to the experiment lies at 1.89 kcal/mol [36], as indicated by the error bars. (d) Dissociation constant from the protein binding energy calculations on the 6 peptide variants.

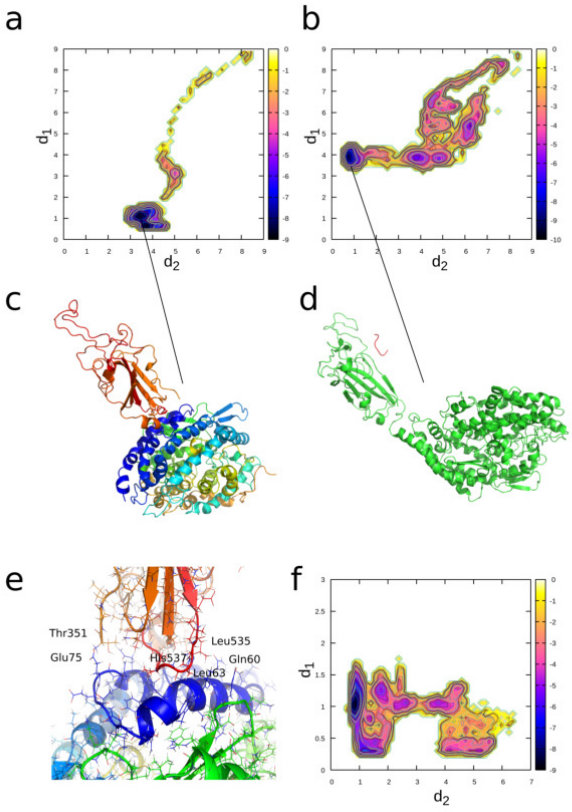

Figure 4: Results from enhanced sampling MD simulations of the binding process of CoV-2 RBD with ACE2 in and without the presence of the hexapeptide. (a) Free energy landscape averaged over the trajectory of the CoV-2 RBD - ACE2 complex without the hexapeptide as function of the order parameters d_1 and d_2, given by the distances between the residues Ile21 CB (ACE2) - Val524 CB (CoV-2 RBD) (d_1) and Ala65 CB (ACE2) - Leu390 CB (CoV-2 RBD) (d_2). (b) Free energy landscape averaged over the trajectory of CoV-2 RBD - ACE2 system in the presence of the hexapeptide as function of the order parameters d_1 and d_2, given by the distances between the residues Ile21 CB (ACE2) - Val524 CB (CoV-2 RBD) (d_1) and Ala65 CB (ACE2) - Leu390 CB (CoV-2 RBD) (d_2) (See figure 1 (c), where the distances d_1 and d_2 are depicted). (c) Final converged state of CoV-2 RBD in complex with the N-terminus of ACE2 in the simulation without the hexapeptide. (d) Final converged state of CoV-2 RBD in the presence of the hexapeptide. (e) Molecular view of the interface of CoV-2 RBD at ACE2. (f) Free energy landscape as function of the order parameters d_1 and d_2, given by the distances between the residues Lys2 NZ (Hexapeptide) - Glu465 OE1 (CoV-2 RBD) (d_1) and Lys2 NZ (Hexapeptide) - Lys335 (ACE2) (d_2). The units in the color bar are given in k_BT (See figure 1 (d), where the distances d_1 and d_2 are depicted).

Simulation	Length	System
20 independent simulations, NPT-MD	50 ns	Hexapeptide - ACE2, explicit solvent (different starting positions)
1 CORE-MD enhanced sampling	200 ns	Hexapeptide - ACE2, implicit solvent
1 CORE-MD enhanced sampling	200 ns	CoV-2 RBD - ACE2, implicit solvent
1 CORE-MD enhanced sampling	200 ns	CoV-2 RBD - ACE2, hexapeptide, implicit solvent

Table 1: MD and CORE-MD enhanced sampling simulations which were performed in this study to investigate the effect of the hexapeptide on the binding process of CoV2-RBD with ACE2.

ACE2-CoV2 RBD complex with the hexapeptide, we inserted the hexapeptide 1 nm away from ACE2. Both systems were centered in a triclinic box with dimensions $7.41900 \times 9.13930 \times 16.23050\ nm^3$ (see Figure 1 c, d).

We used the AMBER99SB forcefield to describe the interactions [19]. For the 20 independent MD simulations over a total of 1 μs, we used the stochastic velocity rescaling algorithm in combination with the berendsen barostat to simulate the system at NPT conditions at 300 K and 1 bar using a timestep of 1 fs [5]. The enhanced sampling simulations of the hexapeptide-ACE2 complex and the separated CoV-2 RBD - ACE2 system with and without the hexapeptide have been performed in implicit solvent using the standard GBSA AMBER99SB parameters. We measured the affinity to a specific binding site using the number of counts N in which the hexapeptide resides below a contact threshold of the $C\alpha$ atoms of 0.65 nm in relation to the total number of frames in the trajectory N_t. We define the relative affinity by the fraction of the affinity η divided by the maximal affinity η_{max} measured for the specific system. The total affinity ϵ is given by the logarithm of the relative affinity :

$$\epsilon = \log\left(\frac{\eta}{\eta_{max}}\right) . \tag{13}$$

We define the free energy ΔF by the propability P along 2 order parameters (i.e. the distances d_1 and d_2 between pairs of atoms) :

$$\Delta F = k_B T \log\left(\frac{P}{P_{min}}\right) , \tag{14}$$

where P_{min} stands for the minimal probability by which the histogram is populated. We assessed the convergence of the 20 NPT-MD simulations and CORE-MD simulation of the hexapeptide-ACE2 system using the average root mean square deviation of the individual distances d_{ij} and the final value :

$$\langle d_{ij} - d_{ij_{final}} \rangle = \frac{1}{N} \sum_{i<j} \sqrt{(d_{ij} - d_{ij_{final}})^2} , \tag{15}$$

over N distances. We used the GROMACS version 4.6 package for the equilibrium MD simulations and a modified GROMACS version 4.5.5 for the CORE-MD simulations [15]. We identified the preferential binding site

of the hexapeptide at ACE2 through a distance based clustering. We then used the conformation of the hexapeptide at ACE2 for the calculation of protein interaction energies using the PRODIGY program [36]. We modeled the individual hexapeptide conformers using the PyMOL modeling program [32].

Results and Discussion

Simulations of hexapeptide binding to ACE2

We tested the affinity of the hexapeptide to the ACE2 protein in 20 independent equilibrium MD simulations over a total simulation time of 50 ns. For this first set of simulations, we modeled 20 different starting conformations of the hexapeptide at an approximate distance of 1 nm away from the potential binding site between ACE2 and CoV-2 RBD. In order to examine the specific affinity of the hexapeptide for binding sites at ACE2, we determined the complete $C\alpha - C\alpha$ distance matrix and averaged over all 20 trajectories. We find that the hexapeptide binds with 55 % of the total affinity to the interface between the N-terminal helix and a β-sheet located at Glu329 and a lowered contact propensity of 43-50 % to the residues Asn330 and Trp328 (see region (4) in Figure 2 a, c). We observe another contact cluster at Asp382 and Met383, where the affinity reaches values of 13 and 15 %. A third contact cluster is located at Thr55, where the affinity of the hexapeptide to ACE2 reaches a value of 21 % (see Figure 2 a, c).

In an enhanced MD simulation, we simulated the binding of the hexapeptide to ACE2 in implicit solvent. We used this simulation to cross validate our equilibrium MD simulations. In the enhanced MD simulation over 200 ns, we observe approximately identical binding patterns of the hexapeptide to the surface of ACE2. We observe a first cluster of contacts at Gly354 with an affinity of 27 %. We find a second contact cluster with a propensity ranging from 80 to 100 % is located at Trp328 and Glu329 (see region (4) in Figure 2 b, d). We observe another contact pattern at Gln325 with a propensity of 96 %. A fourth contact cluster is located at Leu132 with a relative propensity equal 56 %. Finally, we observe a last cluster of contacts at the N-terminus of ACE2 between the residues Ser124 and Gly130 with affinities ranging from 1 to 71 % (see regions (2) and (3) in Figure 2 b, d). An additional minor contact formation with a propensity of 2 % resides at Glu57 (see region (1) in Figure 2 b, d). In order to assess the convergence of the simulations, we measured the average root mean square deviation of the distances from their final value (see Figure 2 e, f). In the set of 20 NPT-MD simulations, we observe that the average root mean square deviation decreases to a value below 0.5 nm at approximately 30 ns, which shows that the simulation length of 50 ns is sufficient for a convergent sampling (see Figure 2 e). The CORE-MD simulation over 200 ns in implicit solvent

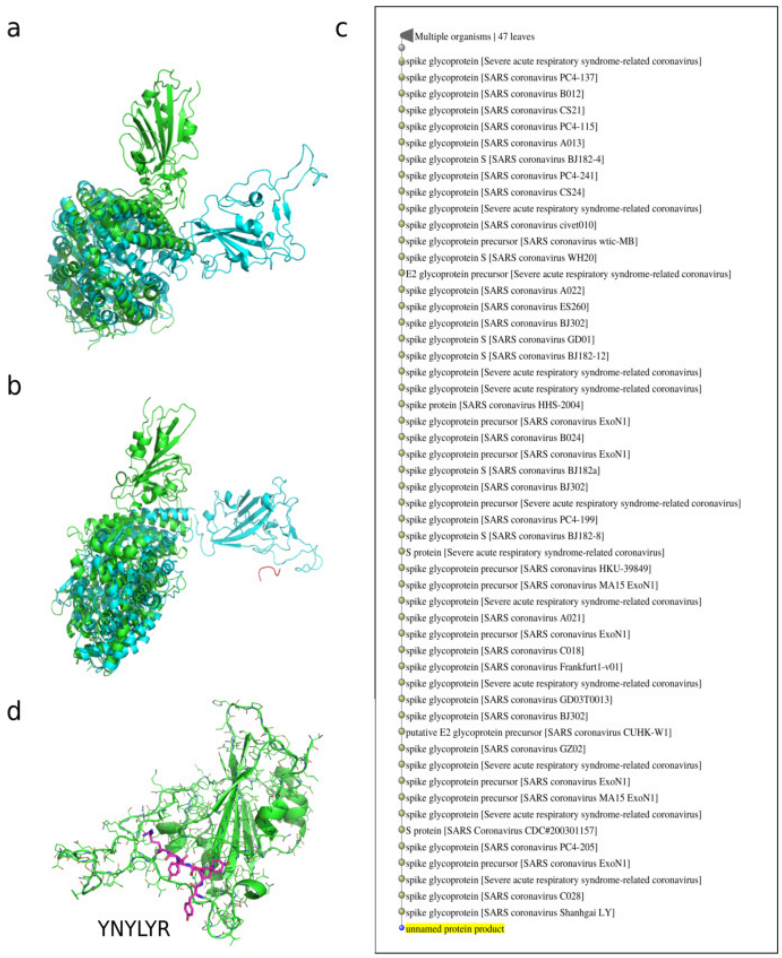

Figure 5: Structural overlays of the PDB structure (PDB: 6M0J) of CoV-2 RBD in complex with ACE2 (green) and the 2 final structures from the enhanced MD simulations without (a) and in the presence of the hexapeptide (b) (cyan). (c) List of SARS CoV viruses as a result from a BLAST search over the protein sequence space of all organisms. Surprisingly, the hexapeptide fragment preferentially occurs in SARS CoV viruses, which makes it suitable as potential drug, due to its dissimilarity with human proteins. (d) Hexapeptide fragment YNYLYR in SARS CoV-2 RBD (PDB : 6M0J, chain E), indicating that the Tyrosine repeat at the positions 1,3 and 5 might be important for the design strategy of a potential peptide-mimetic for the treatment of SARS-CoV-2 infections.

shows that the hexapeptide accesses 5 different states(see Figure 2 f).

Our set of 20 equilibrium MD simulations agrees with the enhanced MD simulation on the general contact patterns of the hexapeptide at ACE2. In both sets of simulations, we find a major binding pattern in the vicinity of the N-terminal region of ACE2, where specifically Arg4 binds to Asn322, which is in agreement with the experimental study [34]. The residues Tyr3 and Tyr5 are interacting with Ala387 in a hydrophobic binding mode (see Figure 3 a, b). We conclude that the hexapeptide binds preferentially to the N-terminal helix and the helical interface close to the N-terminus, which indeed blocks the binding interface between CoV-2 RBD and ACE2 [26, 25]. Based on our findings, the hexapeptide shows a high affinity for the ACE2 binding region, which has the potential to inhibit CoV-2 RBD activation, membrane fusion and the viral entry into the human cell [34]. Subsequently, we used the conformation of the hexapeptide at ACE2 to model five further peptide variants and used an protein binding energy predictor [36] to determine the interaction strength of the models with ACE2 (see Figure 3 c, d). We find that the hexapeptide YKYRYL binds with -7.4 kcal/mol and $K_d = 3.4{\times}10^{-6}$ M. A modification of the C-terminal residue to arginine and a hydrophobic modification at the position 4 to Leu leads to a lower interaction energy, as we find for the models YNYLYL and YNYLYR (ΔG =-6.8 and -7.1 kcal/mol). A mutation at the position two to Leu has only a moderate effect and leads to an interaction energy equal -7.3 kcal/mol. In contrast to the other variants, we found that Lys at position two should remain conserved, while a replacement of Arg at the position 4 with Asn increases the affinity of the hexapeptide to energies equal to -7.6 kcal/mol. The conformation with the lowest dissociation constant K_d is the variant YKYNYI, where the C-terminal Ile stabilizes the interaction leading to a value $K_d = 2.8 \times 10^{-6}$ M. Finally, we state that the hexapeptide variants YKYNYI and YKYNYL contain potential alternative sequences for the binding to ACE2 and the inhibition of CoV-2 RBD activation.

Simulations of CoV-2 RBD ACE2 assembly formation

We tested the effect of the hexapeptide on the binding process of CoV-2 RBD on ACE2 (see Figure 4 and 5). Therefor, we used a starting structure in which CoV-2 RBD is separated by a distance of 2.2 nm away from the surface of ACE2 (see Figure 1 c, d). We simulated the system in and without the presence of the hexapeptide using CORE-MD enhanced sampling. In the simulation of binding of CoV-2 RBD to ACE2, the hydrophilic interface formed by the loop region between the β-sheets of CoV-2 RBD rotates away from the surface of ACE2, mainly due to a electrostatic repulsion. In a comparatively fast translatory process, CoV-2 RBD binds to the N-terminal helix (blue) of ACE2 (see Figure 4 a, c). The interface between CoV-2 RBD and ACE2 is mainly stabilized by hydrophobic interactions between

residues in the range between Leu535 and Thr351 (CoV-2 RBD) and Gln60 to Glu75. Especially a hydrophobic interaction between Leu63 and His537 plays a major role in the stabilization of CoV-2 RBD at ACE2 (see Figure 4 e). The relaxed final conformation of CoV-2 RBD at ACE2 is rotated by approximately 90^o in its attached orientation, when we compare the structure in an overlay with the experimental X-ray structure (PDB: 6M0J [25]) (see Figure 5 a). Additionally, the global contact pattern changed from a hydrophilic interface to a hydrophobic interaction at a different region of CoV-2 RBD, which is initialized by a rotatory motion in the beginning of the simulation that is induced by an electrostatic driving force.

In the simulation of CoV-2 RBD binding to ACE2 in the presence of the hexapeptide, we observe a fast binding process of the hexapeptide to CoV-2 RBD at Glu465 (CoV-2 RBD) and Lys2 (Hexapeptide). We find a secondary contact between the hexapeptide and CoV-2 RBD between Ser349 (CoV-2 RBD) and Arg4 (hexapeptide) in the initial stage of the simulation. The hexapeptide then diffuses along the surface of CoV-2 RBD till its binds strongly to Glu465 (see Figure 4 d, f). An initial rotatory process of CoV-2 RBD is highly analogous to the simulation without the hexapeptide, in which the binding motif of CoV-2 RBD rotates away from the surface of ACE2 due to an electrostatic driving force. In the implicit solvent environment, the formation of the contact interface is driven by the surface charge of CoV-2 RBD and ACE2. Therefore, the formation of a contact interface between polar residues as shown in the PDB structure is less probable. This behavior can potentially change at high salt concentrations. In contrast to the simulation without the hexapeptide, we find that CoV-2 RBD attaches only weakly at the peripheral region of the N-terminal helix 4 nm away from the conformation without the hexapeptide (see Figure 4 b, d and Figure 5 b). In this case, we again emphasize that the hexapeptide changes the electrostatic patterns leading to a change in the structure of the assembly. The hexapeptide binds to CoV-2 RBD leading to the weak attachment of CoV-2 RBD to ACE2. We anticipate that this conformation corresponds to an inhibited state, in which CoV-2 RBD does not become activated and the process of membrane fusion might get inhibited due to the low affinity of CoV-2 RBD for ACE2. Due to thermal fluctuations and on longer timescales, CoV-2 RBD might dissociate away from ACE2, which would inhibit CoV-2 activation.

We were surprised by the high specificity with which the hexapeptide bound to ACE2 and CoV-2 RBD. Since we observed that the hexapeptide inhibits the binding process of CoV-2 RBD, we performed a BLAST search over all organisms, which contain the specific fragment in their protein sequences [2, 12]. Surprisingly, we found that 47 out of 50 hits in the sequence search returned SARS Corona Virus organisms, while only 3 hits were contained in bacteria, which shows that the hexapeptide-pattern preferentially occurs for SARS CoV viruses, but not in human proteins (see Figure 5 c).

That result shows that an *a priori* affinity for another function in the human organism can be excluded, which makes the hexapeptide a suitable candidate as a potential drug. When we analyzed the sequence of CoV-2 RBD, we find a hexapeptide sequence YNYLYR, which contains the same Tyr repeat at the positions 1, 3 and 5, but different residues at the positions 2, 4 and 6, which might be an indicator that Tyr at the positions 1, 3 and 5 is imminent for the specificity of CoV-2 RBD (see Figure 5 d). However, we found that this hexapeptide sequence leads to the lowest interaction energy ΔG=-6.8 kcal/mol as we found in a modeling approach using the preferential hexapeptide ACE2 binding site as a structural model (see Figure 3 c, d). We only can speculate that the aminoacids at the positions 2, 4 and 6 are affecting the relative affinity of the fragment for the ACE2 receptor, while we find that Tyr at the positions 1, 3 and 5 is essential for the binding. We assume that Tyr at the positions 1, 3, and 5 has to be conserved for the design of a peptide mimetic used as potential drug against SARS CoV-2, while the hexapeptide sequence YKYRYL inhibits the viral interaction with ACE2 as we have shown in this work. Finally, we conclude that binding of CoV-2 RBD to ACE2 is unexpectedly highly heterogeneous, which is also the case for the interaction of the hexapeptide with ACE2. We anticipate that the CoV-2 RBD - ACE2 interaction (as well as the hexapeptide - ACE2 complex) depends from the ionic strength. At high ionic strengths, a polar CoV-2 RBD - ACE2 binding interface as given in the experimental structure might be stabilized in a strong field of surrounding ions and water[25, 21].

Conclusions

In this paper, we investigated the binding process of a fragment of the SARS Coronavirus spike protein receptor domain (CoV RBD), the hexapeptide YKYRYL on the ACE2 receptor and its effect on the assembly formation and activation of CoV-2 RBD at ACE2. In agreement with an experimental study, we find a high affinity of the hexapeptide to the binding interface between CoV-2 RBD and ACE2, which we investigated using 20 independent equilibrium MD simulations over a total of 1 μs and a 200 ns enhanced MD simulation. We then evaluated the effect of the hexapeptide on the binding process of CoV-2 RBD to ACE2 in long-time enhanced MD simulations. In that set of simulations, we found that CoV-2 RBD does not bind to ACE2 with the binding motif shown in experiments, but it rotates due to an electrostatic repulsion and forms a hydrophobic interface with ACE2. Surprisingly, we observed that the hexapeptide binds to CoV-2 RBD, which has the effect that this protein only weakly attaches to ACE2, so that the activation of CoV-2 RBD might be inhibited in this case. Our results indicate that the hexapeptide might be a possible treatment option which prevents the viral activation through the inhibition of the interaction between ACE2

and CoV-2 RBD.

References

[1] A. O. Adedeji and S. G. Sarafinos, Curr. Opin. Virol., 8 (2014), pp. 45–53.

[2] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, J. Mol. Biol., 215 (1990), pp. 403–410.

[3] M. Axmann, PhD Thesis, (2007).

[4] E. S. Brielle, D. Schneidmann-Duhovny, and M. Linial, Viruses, 12 (2020), pp. 2–10.

[5] G. Bussi, D. Donaldio, and M. Parrinello, J. Chem. Phys., 126 (2007), p. 014101.

[6] J. F. Chan, S. Yuan, K. H. Kok, K. K. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C. Yip, R. W. Poon, H. W. Tsoi, S. K. Lo, K. H. Chan, V. K. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. Cheng, H. Chen, C. K. Hui, and K. Y. Yuen, Lancet, 395 (2020), pp. 30154–30159.

[7] E. de Wit, N. van Doremalen, D. Falzarano, and V. J. Munster, Nat. Rev. Microbiol., 14 (2016), pp. 523–534.

[8] L. Du, Y. He, Y. Zhou, S. Liu, B. J. Zheng, and S. Jiang, Nat. Rev. Microbiol., 7 (2009), pp. 226–236.

[9] R. Feynman and A. R. Hibbs, MacGraw Hill Companies, (1965).

[10] X. Y. Ge, J. L. Li, X. L. Yang, A. A. Chmura, G. Zhu, J. H. Epstein, J. K. Mazet, B. Hu, W. Zhang, C. Peng, Y. J. Zhang, C. M. Luo, B. Tan, N. Wang, Y. Zhu, G. Crameri, S. Y. Zhang, L. F. Wang, P. Daszak, and Z. L. Shi, Nature, 503 (2013), pp. 535–538.

[11] S. Gierer, S. Bertram, F. Kaup, F. Wrensch, A. Heurich, A. Krämer-Kühl, K. Welsch, M. Winkler, B. Meyer, C. Drosten, U. Dittmer, T. von Hahn, G. Simmons, H. Hofmann, and S. Pöhlmann, J. Virol., 87 (2013), pp. 5502–5511.

[12] W. Gish and D. J. States, Nature Genet., 3 (1993), pp. 266–272.

[13] I. Glowacka, S. Bertram, M. A. M. ller, P. Allen, E. Soilleux, S. Pfefferle, I. Steffen, T. S. Tsegaye, Y. He, K. Gnirss, D. Niemeyer, H. Schneider, C. Drosten, and S. Pölmann, J. Viorol., 85 (2011), pp. 4122–4134.

[14] A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, S. Perlman, L. L. M. Poon, D. V. Samborskiy, I. A. Sidorov, I. Sola, and J. Ziebuhr, Nat. Microbiol., 5 (2020), pp. 536–544.

[15] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput., 4 (2008), pp. 435–447.

[16] A. Heurich, H. Hofmann-Winkler, S. Gierer, T. Liepold, O. Jahn, and S. Pöhlmann, J. Virol., 88 (2014), pp. 1293–1307.

[17] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, and S. Pöhlmann, Cell, 181 (2020), pp. 271–280.

[18] K. V. Holmes, Science, 309 (2005), pp. 1822–1823.

[19] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins, 65 (2006), pp. 712–725.

[20] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao, Lancet, 395 (2020), pp. 497–506.

[21] P. L. Kastritis and A. M. J. J. Bonvin, J. R. Soc. Interface, 10 (2013), p. 20120835.

[22] A. Kilianski and S. C. Baker, Antivir. Res., 101 (2014), pp. 105–112.

[23] H. Kleinert, World Scientific, 5th edition (2009), pp. 1–1547.

[24] K. Kuba, Y. Imai, S. Rao, H. Gao, F. Guo, B. Guan, Y. Huan, P. Yang, Y. Zhang, W. Deng, L. Bao, B. Zhang, G. Liu, Z. Wang, M. Chappell, Y. Liu, D. Zheng, A. Leibbrandt, T. Wada, A. S. Slutsky, D. Liu, C. Qin, C. Jiang, and J. M. Penninger, Nat. Medicine, 11 (2005), pp. 875–879.

[25] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, and X. Wang, Nature, (2020), pp. https://doi.org/10.1038/s41586–020–2180–5.

[26] F. Li, W. Li, M. Farzan, and S. C. Harrison, Science, 309 (2005), pp. 1864–1867.

[27] W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan, Nature, 426 (2003), pp. 450–454.

[28] L. R. Mead and N. Papanicolaou, J. Math. Phys., 25 (1984), p. 2404.

[29] E. K. Peter, J. Chem. Phys., 147 (2017), p. 214902.

[30] E. K. Peter and J.-E. Shea, Phys. Chem. Chem. Phys., 19 (2017), pp. 17373–17382.

[31] E. K. Peter and J. Černý, Int. J. Mol. Sci., 20 (2019), p. 370.

[32] Schrödinger, LLC, *The PyMOL molecular graphics system, version 1.8*, (2015).

[33] V. Srivastava, ed., *Peptide Therapeutics*, Drug Discovery, The Royal Society of Chemistry, 2019.

[34] A.-W. Struck, M. Axmann, S. Pfefferle, C. Drosten, and B. Meyer, Antivir. Res., 94 (2012), pp. 288–296.

[35] E. Valeur, S. M. Guéret, H. Adihou, R. Gopalakrishnan, M. Lemurell, H. Waldmann, T. N. Grossmann, and A. T. Plowright, Angew. Chem. Int. Ed., 56 (2017), pp. 10294–10323.

[36] A. Vangone and A. M. J. J. Bonvin, eLife, 4 (2015), p. e074054.

[37] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, Cell, 180 (2020), pp. 281–292.

[38] World Health Organization, Coronavirusdisease (COVID-19) Situation Report–110, (May 9, 2020).

[39] S. Xia, M. Liu, C. Wang, W. Xu, Q. Lan, S. Feng, F. Qi, L. Bao, L. Du, S. Liu, C. Qin, F. Sun, Z. Shi, Y. Zhu, S. Jiang, and L. Lu, Cell Research, 30 (2020), pp. 343–355.

[40] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, and Z.-L. Shi, Nature, 579 (2020), pp. 270–273.

[41] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan, and China Novel Coronavirus Investigating and Research Team., New Engl. J. Med., 382 (2020), pp. 727–733.