Coarse-grained molecular simulations of the binding of the SARS-CoV 2 spike protein RBD to the ACE2 cell receptor

David De Sancho

University of the Basque Country, Faculty of Chemistry,

*Paseo Manuel Lardizabal, 3, 20018 Donostia-San Sebastián, Spain**

José A. Gavira

Laboratory of Crystallographic Studies,

IACT (CSIC-UGR), Granada, 18100, Spain

Raul Pérez-Jiménez

Ikerbasque, Basque Foundation for Science and CIC nanoGUNE Tolosa Hiribidea,

76, 20018, Donostia-San Sebastian, Spain

Abstract

Since it was first observed in late 2019, the COVID-19 pandemic has created a global emergency for national health systems due to millions of confirmed cases and hundreds of thousands of deaths. At a molecular level, the bottleneck for the infection is the binding of the receptor binding domain (RBD) of the viral spike protein to ACE2, an enzyme exposed on human cell membranes. Several experimental structures of the ACE2:RBD complex have been made available, however they offer only a static description of the arrangements of the molecules in either the free or bound states. In order to gain a dynamic description of the binding process that is key to infection, we use molecular simulations with a coarse grained model of the RBD and ACE2. We find that binding occurs in an all-or-none way, without intermediates, and that even in the bound state, the RBD exhibits a highly dynamic behaviour. From short equilibrium simulations started in the unbound state we provide snapshots that result in a tentative mechanism of binding. Our findings may be important for the development of drug discovery strategies that target the RBD.

* david.desancho@ehu.eus

The spillover of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to humans has resulted in the coronavirus disesease 2019 (COVID-19) pandemic [1]. The number of cases around the planet has skyrocketed in multiple countries, causing major disruptions and hundreds of thousands of deaths around the world. This crisis has spurred titanic efforts from the scientific community in order to develop therapies to combat the disease and prevent future outbreaks. In particular, structural biologists have been quick to resolve the 3D coordinates of a few of the viral proteins, including among others its main protease [2], the Nsp15 endoribonuclease [3], and the spike protein [4–8]. These efforts provide invaluable detail of the molecular machinery involved in the viral infection cycle, which is essential for the development of effective drugs.

Among the therapeutic strategies aiming to stop the pandemic, much work is being devoted to target the interaction between the trimeric spike protein (or protein S) to the host receptor, the angiotensin-converting enzyme 2 (ACE2) [9, 10]. The binding between both proteins precedes membrane fusion and may be the Achilles heel of the virus. The spike first engages with the host cell when it interacts with ACE2 via the very mobile receptor binding domain (RBD) of one of its monomers in its "up" state [4]. The interaction between the RBD and ACE2 buries an extended surface of ~1700 $Å^2$ involving a coiled coil formed by two long helices of ACE2 and a loosely packed binding motif in the RBD (see Figure 1). The structure for the SARS-CoV-2 RBD bound to ACE2 bears close resemblance to that of the same complex for the previous SARS-CoV virus [11]. However, there is evidence that the greater infectivity of the new virus may be determined at least in part by the ~10-20-fold higher binding affinity for ACE2 [4] (although estimates differ among groups, see Supplementary Information Fig. S1).

In order to better understand viral infection and design ways to hijack it, it is important to understand the mechanism of binding between the spike protein and ACE2. This information is not readily available from the static X-ray and cryo-EM experimental structures. Molecular dynamics (MD) simulations are ideally posed to provide this picture. However, simulating protein-protein binding using atomistic MD even for small proteins is exceedingly time-consuming even for much smaller systems than the RBD:ACE2 complex [12, 13]. Alternatively, one may resort to coarse-grained structure-based (Gō) models, whose potential energy minimum is located by construction in the reference experimental structure [14]. In this way, these models are consistent with the central prescription of a

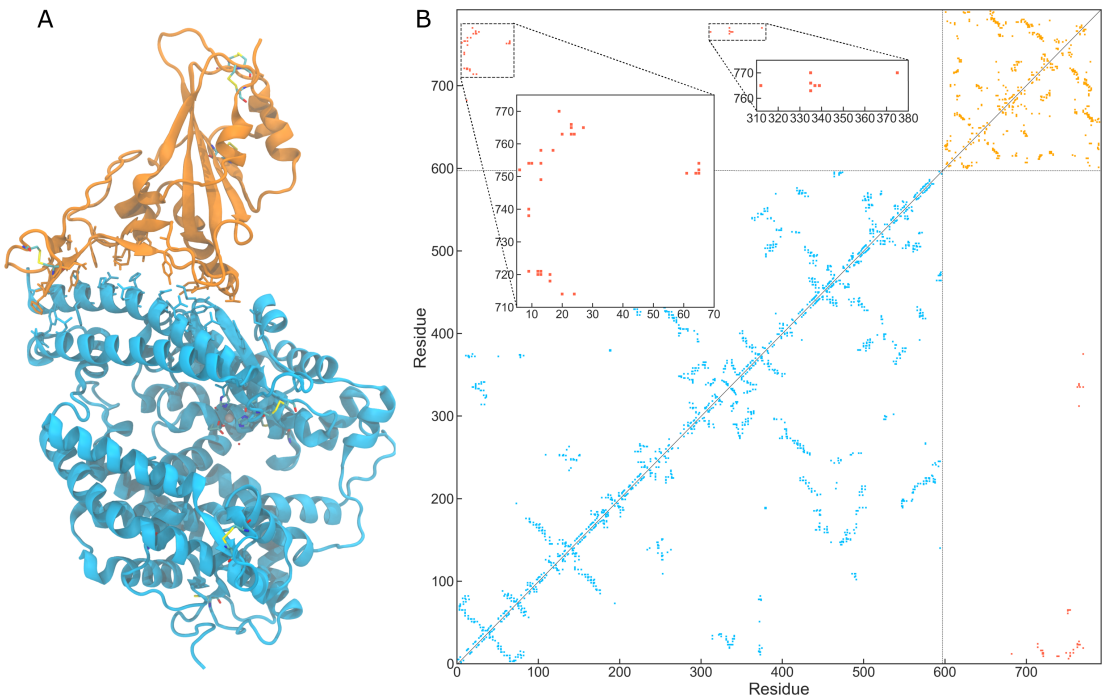

FIG. 1. Three dimensional structure (A) and contact map (B) of the ACE2:RBD complex. ACE2 is shown in cyan and the RBD in orange. We show atomic detail for residues forming intermolecular contacts, cysteines and Zn coordinating residues in ACE2. Intramolecular contacts for each protein in (B) are shown in the same colours as in (A), while intermolecular contacts are shown in red. The insets highlight the regions with most intermolecular contacts.

minimally frustrated, funnelled energy landscape of protein folding theory [15]. In the past, Gō models have been applied successfully to the study of protein-protein binding for complexes involving both folded and disordered proteins [16, 17].

In this work, we use the structure-based simulation model by Karanicolas and Brooks, which is coarse-grained to the Cα level [18] (see details in the Supplementary Methods section). We built models using a recently determined 2.45 Å resolution X-ray structure of the ACE2:RBD complex (with PDB ID: 6m0j [5], see Figure 1A). Although we take into consideration both the disulfide bonds and Zn(II) interactions in our simulations we do not consider glycosilation effects. Among the 22 predicted N-glycosylation sites in the spike protein of SARS-CoV-2 [19], only the highly conserved glycosylation site N343 is in the RBD, and this is located far from the interaction region [11]. It is also relevant to note that a second glycosylation site in the RBD of the SARS-CoV spike (N357) is lost in SARS-

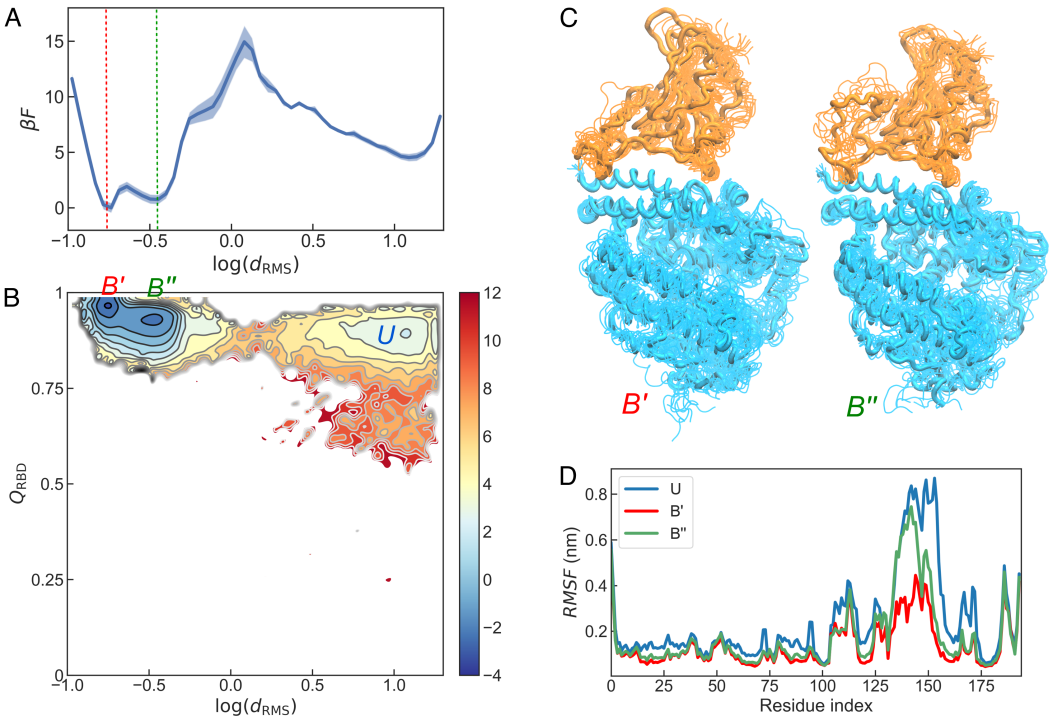

FIG. 2. (A, B) Potentials of mean force for the binding of SBD to ACE2 for the projection on $\log(d_{\mathrm{RMS}})$ and both the fraction of intra-RBD contacts Q_{RBD} and $\log(d_{\mathrm{RMS}})$. B′ and B″ are the fully bound and fluctuating sub-basins within the bound state and U marks the unbound state. Free energies are in units of kT. (C) Representative snapshots for the B′ and B″ states. (D) $RMSF$ for RBD in the different states as a function of the residue index.

CoV-2 [19]. The energy terms in the model were calibrated carefully in order to recover a value of K_{D} at 300 K in the experimental range (i.e. tens of nM, see Supplementary Information Fig. S1).

First, we run replica exchange molecular simulations (REMD) of the RBD in presence of the ACE2 receptor. We used 32 replicas for temperatures ranging between 285 and 378 K , using a Langevin dynamics integrator with a 10 fs time-step and a friction constant of 0.2 ps^{-1} using the Gromacs software package [20]. The resulting 2 μs of simulation data at each temperature were projected on an order parameter for binding, the logarithm of the d_{RMS}, that collapses the unbound state into a narrow well and expands the bound state (see Supplementary Methods). The simulations are well converged, as shown by cumulative averages that do not change after 0.8 μs at each temperature (see Supplementary Information Fig. S2).

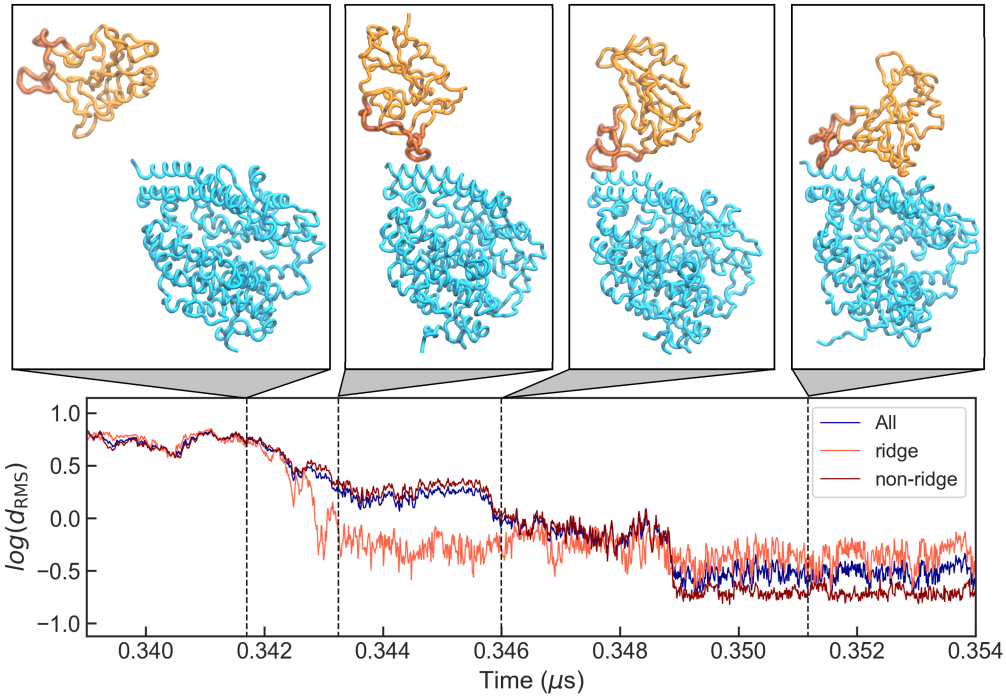

FIG. 3. Representative transition path for a binding transition between RBD and ACE2. We show the projection of the simulation data on the $\log(d_{\mathrm{RMS}})$ for all contacts (dark blue) or different subsets of interprotein contacts, either involving the ridge or not (light and dark red, respectively). Snapshots at different time-points are shown at the top. Colors as in Fig. 1. The thicker region in the RBD corresponds to the ridge.

In Figure 2A we show potentials of mean force derived using the weighted histogram analysis method WHAM [21]. The free energy surface suggests that the RBD:ACE2 complex formation occurs in an all-or-none way, with a dominant free energy barrier between the bound and unbound states. Despite the absence of stable intermediates, the bound state has a small intra-state barrier. The presence of two sub-basins in the bound state with average values of $\log(d_{\mathrm{RMS}})$ of -0.45 and -0.75 is also apparent in the two dimensional free energy landscape we show in Figure 2B, including also a folding coordinate for the RBD, the fraction of native contacts (Q_{RBD}, see Supplementary Methods).

Hence, the free energy landscapes suggest some degree of heterogeneity in RBD even in the bound state. In Figure 2C we show snapshots of conformations within each of the bound state sub-basins, derived from the REMD simulation data at 300 K. The lowest $\log(d_{\mathrm{RMS}})$ form (which we term B′) very much matches the experimental structure, while

in the higher $\log(d_{\mathrm{RMS}})$ form (B″) there are considerable conformational fluctuations that seem mainly localized in the region termed receptor binding ridge [7, 11], despite the disulfide bond that barely holds this region folded.

To characterize the residual disorder in the bound state we calculate the root mean square fluctuations ($RMSF$) for the RBD estimated from snapshots of both substates and the unbound state (see Figure 2D). As anticipated, fluctuations are primarily localized in the ridge region, both in the unbound and in the bound form of the protein, although in the former they extend a bit longer to the receptor binding region. We see that upon binding, fluctuations decrease considerably in the RBD but they remain larger than average in the receptor binding ridge. Importantly, the ridge region could not be properly resolved in recent cryo-EM structures of the spike protein [4]. Also, structural disorder is predicted to appear around residue 150 of RBD from multiple servers [22]. Finally, the location of the fluctuations of the coarse-grained simulations is consistent with atomistic MD [23], giving further confidence to the outcome of our simulations.

In order to recover a dynamic picture of the binding process we have run tens of molecular simulations starting from randomly oriented conformations in the unbound state, resulting in a few binding events. This allows to understand the importance of the properties of the underlying energy landscape in the binding process, in particular the relevance of disorder in RBD, and propose a binding mechanism.

In Figure 3 we show the projection on relevant progress coordinates of the simulation data for one of the binding events (two more are shown in the Supplementary Information Fig. S3). Again, we monitor the $\log(d_{\mathrm{RMS}})$ as a progress coordinate for binding, but now considering either all contacts or different subsets of contacts separately. In particular, we discern between contacts formed by the ridge residues and those in the remaining part of the extensive RBM. In all the binding events we have observed binding occurs first via the disordered region, which establishes the first few contacts, resulting in a decrease of $\log(d_{\mathrm{RMS}})$. Then, the remaining part of the RBD rests on the ACE2 binding site, until the protein fully commits to the bound form. Representative snapshots help visualize how the extended, disordered part of the protein more efficiently searches for its binding partner, possibly due to a slightly increased capture radius (Figure 3). This is reminiscent of the so-called "fly-casting" mechanism that has been invoked to explain the kinetic advantage of intrinsically disordered proteins for binding [24].

The disorder of the viral protein in the RBD:ACE2 complex formation may seem surprising. However, it is consistent with the lack of resolution in the receptor binding ridge in the cryo-EM experiments [4]. More generally, structural disorder has recently been highlighted as a prevalent feature in viral proteins [25], suggesting that small regions of structural disorder may provide an evolutionary advantage to viruses.

Our results provide, to our knowledge, the first molecular description of the binding process of the RBD to the ACE2 receptor which is an essential first step in the infection by COVID-19. There are several potential limitations in our work due to the coarse-grained nature of the model and the simplicity in the description of the interactions. First, the simulation model does not consider explicitly long-range charged interactions, although pairwise energy terms are residue-specific [18]. This prevents an adequate description of electrostatic steering effects that may speed up the binding process [26]. Additionally, we do not incorporate the glycosilation effects, although eperimental structures show that glycosilated sites in ACE2 and RBD are far from the binding interface and hence likely to be small.

We hope the insight from our coarse grained simulations will be helpful in future drug discovery work either for antivirals that exploit the binding properties of RBD:ACE2 or for vaccine development. One possibility would be to try to tune the disorder in the RBD, distorting the balance between conformational entropy loss and enthalpic gain. For example, by favouring a reductive environment for the virus that promotes reduction of disulfides would increase the conformational disorder of the receptor binding ridge even further. Alternatively, fragments of ACE2 involving the two main helices may be tested as potential inhibitors of the virus.

AUTHOR CONTRIBUTIONS

DDS, RPJ and JAG conceived research and wrote the paper, DDS carried out all simulations and analyzed the data.

ACKNOWLEDGMENTS

DDS receives financial support from the grants PGC2018-099321-B-I0 and RYC-2016-19590 from the Spanish Ministry of Science, Research and Universities (MINECO/FEDER) and the Basque Government through grant IT588-13. The Spanish Ministry of Science, Research and Universities also supports JAG and RPJ through grants BIO2016-74875-P and BIO2016-77390-R, respectively. DDS wishes to thank Robert B. Best for generously sharing computational tools and the staff at the DIPC Supercomputing Center for technical support. DDS also acknowledges María M. Caffarel, for helpful discussions and Carmen and Lucía de Sancho for key insights in the structure of the complex.

---

[1] 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, and W. Tan, N. Engl. J. Med. **382**, 727 (2020).

[2] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, and H. Yang, bioRxiv (2020), 10.1101/2020.02.26.964882.

[3] Y. Kim, R. Jedrzejczak, N. I. Maltseva, M. Endres, A. Godzik, K. Michalska, and A. Joachimiak, bioRxiv (2020), 10.1101/2020.03.02.968388.

[4] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, Science **367**, 1260 (2020).

[5] 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), 10.1038/s41586-020-2180-5.

[6] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou, Science **367**, 1444 (2020).

[7] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, and F. Li, Nature (2020), 10.1038/s41586-020-2179-y.

[8] Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K.-Y. Yuen, Q. Wang, H. Zhou, J. Yan, and J. Qi, Cell (2020), https://doi.org/10.1016/j.cell.2020.03.045.

[9] H. Zhang, J. M. Penninger, Y. Li, N. Zhong, and A. S. Slutsky, Intensive Care Med , 1 (2020).

[10] 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 (2020), https://doi.org/10.1016/j.cell.2020.02.052.

[11] F. Li, W. Li, M. Farzan, and S. C. Harrison, Science **309**, 1864 (2005).

[12] N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, Nat. Chem. **9**, 1005 (2017).

[13] A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich, and D. E. Shaw, Proceedings of the National Academy of Sciences **116**, 4244 (2019), https://www.pnas.org/content/116/10/4244.full.pdf.

[14] H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).

[15] J. Bryngelson, J. Onuchic, N. Socci, and P. Wolynes, Proteins **21**, 167 (1995).

[16] Y. Levy, P. G. Wolynes, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **101**, 511 (2004).

[17] A. G. Turjanski, J. S. Gutkind, R. B. Best, and G. Hummer, PLOS Computational Biology **4**, 1 (2008).

[18] J. Karanicolas and C. L. Brooks, Protein Sci **11**, 2351 (2002).

[19] Y. Chen, Y. Guo, Y. Pan, and Z. J. Zhao, Biochemical and Biophysical Research Communications **525**, 135 (2020).

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

[21] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).

[22] G. Yachdav, E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Hönigschmid, A. Schafferhans, M. Roos, M. Bernhofer, L. Richter, H. Ashkenazy, M. Punta, A. Schlessinger, Y. Bromberg, R. Schneider, G. Vriend, C. Sander, N. Ben-Tal, and B. Rost, Nucleic Acids Res. **42**, W337 (2014).

[23] J. He, H. Tao, Y. Yan, S.-Y. Huang, and Y. Xiao, Viruses **12** (2020), 10.3390/v12040428.

[24] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **97**, 8868 (2000).

[25] B. Xue, D. Blocquel, J. Habchi, A. V. Uversky, L. Kurgan, V. N. Uversky, and S. Longhi, Chem. Rev. **114**, 6880 (2014).

[26] R. C. Tan, T. N. Truong, J. A. McCammon, and J. L. Sussman, Biochemistry **32**, 401 (1993).