Nanoscale

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

View Article Online
View Journal

Cite this: DOI: 10.1039/d0nr03969a

Received 23rd May 2020,
Accepted 22nd July 2020

DOI: 10.1039/d0nr03969a

rsc.li/nanoscale

Quantitative determination of mechanical stability in the novel coronavirus spike protein

Rodrigo A. Moreira, *[a] Mateusz Chwastyk, [b] Joseph L. Baker, [c] Horacio V. Guzman [d] and Adolfo B. Poma *[a]

We report on the novel observation about the gain in nanomechanical stability of the SARS-CoV-2 (CoV2) spike (S) protein in comparison with SARS-CoV from 2002 (CoV1). Our findings have several biological implications in the subfamily of coronaviruses, as they suggest that the receptor binding domain (RBD) (∼200 amino acids) plays a fundamental role as a damping element of the massive viral particle's motion prior to cell-recognition, while also facilitating viral attachment, fusion and entry. The mechanical stability *via* pulling of the RBD is 250 pN and 200 pN for CoV2 and CoV1 respectively, and the additional stability observed for CoV2 (∼50 pN) might play a role in the increasing spread of COVID-19.

Introduction

Since the recent outbreak of the 2019 novel coronavirus (CoV2) and the fast spread of the disease (COVID-19) around the globe, a rapid and very well-coordinated scientific research machinery has been put in place all over the world. In the past 5 months several scientific groups have pursued a comprehensive structural characterization of the main protein components of CoV2 to fight against the disease. Importantly, the four main structural proteins are well-resolved at atomic scale, including the spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins.[1–3] Each component plays a crucial role in cell-recognition (S), generating ionic channel (E), defining the shape of the viral envelope (M), and binding the positive-stranded RNA that is made up of about 30 000 nucleotides (N). These components are either in contact with the viral membrane (S, E and M) or in the inner part (N). In order to fight against CoV2 we need to determine the molecular weaknesses in the structure or along the processes that involve CoV2 proteins. In a recent article by Wang *et al.*,[4] the high affinity of CoV2 to the human angiotensin-converting enzyme (ACE2), which is an enzyme attached to the outer membrane surface, compared to CoV1 was shown *in vitro* (94.6 nM *vs.* 408.7 nM for CoV2 and CoV1, respectively). Such a difference in K_D has been suggested to be the reason for high spreading of the disease. However, Walls *et al.*[3] *via in vivo* studies has shown there is not a substantial change in K_D and the same conclusion has been reached by *in silico* studies.[5] It has also been suggested that the S1/S2 furin-like cleavage in the sequence Q_{677}TNSPRRAR↓SV_{687} could enhance its transmissibility and enable fusion machinery in CoV2, and that the cleavage event may lead to a destabilized structure which can facilitate viral entry.[6] Recently it was shown the possibility of a combined effect given by the lying-down conformation of RBD and the high ACE binding affinity as a mechanism to evade immune surveillance.[7] These features are the cornerstone of traditional antiviral development for inhibition of CoV2 against cell recognition,[8] while also enhancing the biophysical understanding of another component contained in the complete virus theory and simulation at different scales.[9,10] In this study, we use a combination of tools from structural biology and molecular dynamics simulation to unveil the mechanical forces that the spike proteins may withstand before losing their function. It has been shown in bacteriophage T4 long tail fibers that thermal Brownian fluctuations of the virus can exert large forces[11] such as 190 ± 70 pN at the level of the viral receptor. Here we determine a substantial mechanical stability of CoV2 and CoV1, which is about that of the Ig27 domain of titin,[12,13] F_{max}(Ig27) = 200 pN. This information allows us to envision the molecular-scale scenario during the diffusion of

[a]*Biosystems and Soft Matter divison, Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawińskiego 5B, 02-106 Warsaw, Poland. E-mail: rams@ippt.pan.pl, apoma@ippt.pan.pl*
[b]*Institute of Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland*
[c]*Department of Chemistry, The College of New Jersey, 2000 Pennington Road, Ewing, NJ 08628, USA*
[d]*Department of Theoretical Physics, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*

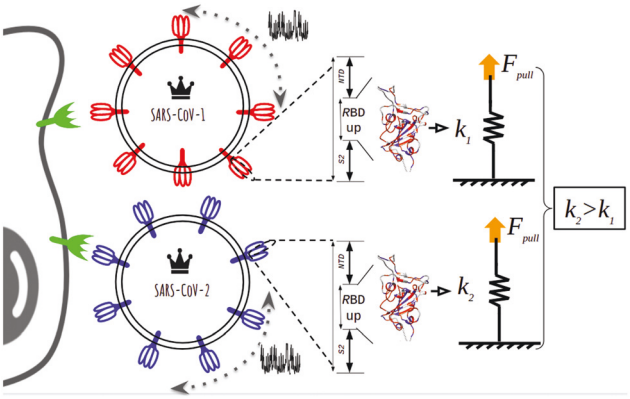

Fig. 1 Schematic diagram of the nanomechanical stability that renders CoV2 more stable than CoV1 ($k_2 > k_1$). The spike protein for CoV2 and the human ACE2 are represented in blue and green color respectively.

viral particles under Brownian motion and upon their first encounter with the cell membrane receptor (ACE2). This mechanical picture has emerged from extensive molecular dynamics and coarse-grained simulations, and it enables the description of a dynamical process of recognition and confirms that the RBD makes a significant contribution to the mechanical stability of the full spike timer (see Fig. 1). We also interrogated the contribution of the current mutations on the dynamics of the CoV2 spike protein.

Methods

COVID-19 spike glycoprotein + ACE2 models

We modeled the spike (S) protein from CoV2 and CoV1, both bound to ACE2 at the prefusion geometry and related structures (PDB IDs 6ACG, 6M0J, 6VSB, 6CRV and 2XY9), with one RBD up and two RBD at down positions and in complex with ACE2. The sequences that describe the spike proteins come from QIQ50172.1 and AAR86775.1 stored at GenBank database for CoV2 and CoV1, respectively. The standard Needleman–Wunsch algorithm was used as implemented by Chimera visualization software to align the sequences and the missing loops were modeled by homology using MODELLER[14] and its energy minimization algorithms. The disulfide bonds were the ones described by the PDB files, including 3 for ACE2 and 14 and 15 per chain of the spike homotrimer for CoV2 and CoV1, respectively.

Contacts maps

The OV + rCSU contact maps used in this work was successfully used before to describe proteins.[15,16] Here we have employed this methodology to scan through the MD trajectory and determine contacts between amino acids in a dynamics form. The overlap of enlarged spheres was used to define the OV contact map. The rCSU approach places the chemical character of each atom, and respective bonds, into categories and counts the number of stabilizing and destabilizing contacts per residue, defining a contact when both residues have a net stabilizing character. We implemented our own contact map software, as detailed by Wołek *et al.*[17] The data set of all-atom MD trajectories, contact map methods, and scripts of CoV_2 and CoV_1 spikes can be accessed *via* the Zenodo repository.[18]

Molecular dynamics

All-atom simulations were carried out with Amber18,[19] and system components (protein, ions, water) were modeled with the included FF14SB[20] and TIP3P[21] parameter sets. Energy minimization used CPU pmemd, while later simulation stages used GPU pmemd. CoV2 and CoV1 systems with one RBD up (with/without ACE2) were solvated in 12 Å octahedral water shells. Cysteine residues identified in the initial models as having a disulfide bond (DB) were bonded using tLeap. All simulations used 0.150 M NaCl. Hydrogen mass repartitioning was applied only to the protein to enable a 4 fs timestep.[22] The SHAKE algorithm was applied to hydrogens, and a real-space cutoff of 8 Å was used. Periodic boundary conditions were applied and PME was used for long-range electrostatics. Minimization was by steepest descent (2000 steps) followed by conjugate gradient (3000 steps). Heating used two stages: (1) NVT heating from 0 K to 100 K (50 ps), and (2) NPT heating from 100 K to 300 K (100 ps). Restraints of 10 kcal mol^{-1} $Å^{-2}$ were applied during minimization and heating to C_α atoms. During 6 ns of equilibration at 300 K C_α restraints were gradually reduced from 10 kcal mol^{-1} $Å^{-2}$ to 0.1 kcal mol^{-1} $Å^{-2}$. Finally, restraints were released and 320 ns unrestrained production simulations were carried out for CoV2 and CoV1 systems. Production simulations began from the final equilibrated snapshots, and five copies of each system were simulated. As unrestrained systems can freely rotate we monitored simulations for any close contacts and found that in one copy of the CoV1 simulation without ACE2 and one RBD up that a few contacts close to 8 Å occur near the end of the 320 ns between the RBD and a different subdomain of the spike complex in a periodic image. However this did not influence analyzed structural properties which is verified by comparing results across simulations. The Monte Carlo barostat was used to maintain pressure (1 atm), and the Langevin thermostat was used to maintain 300 K temperature (collision frequency 1 ps^{-1}), as implemented in Amber18.[19] In aggregate, 6.4 μs of all-atom simulation of systems ranging from 396 147 to 879 100 atoms was carried out for this work. Snapshots from the MD simulations can be found in the Zenodo repository.[18]

Nanomechanics of proteins

The nanomechanical simulations are based on the Gō-like model[23,24] that has been used to sample conformational changes in proteins and calculate the elastic parameters under force deformation in single proteins, protein filaments, cellulose, and protein–protein, protein–polysaccharide and protein–lipid interfaces.[15,25–28] At first we pull each chain of the trimer to identify the mechanostable protein domains. Our results for CoV2 show the RBD to be last domain to unfold. The S2 and NTD unfold first than RBD (F_{maxS2} = 160 pN, F_{maxNTD} = 125 pN and F_{maxRBD} = 250 pN). In order to quantify

the difference in mechanical stability of the RBD between CoV2 and CoV1 we perform pulling simulations of the RBD which is pulled along the end-to-end vector connecting the C_α-atoms from the N- and C-terminus and the reaction coordinate is the displacement of the pulling spring. Moreover, additional beads have been attached to those C_α-atoms with the spring constant being 0.1 kcal mol^{-1} Å^{-2}, which is a typical value of the AFM cantilever stiffness in protein stretching studies. The type of coarse-grained methodology employed here offers the flexibility to use much lower pulling speeds than in all-atom MD simulations and keep it applied for a longer time. Each system was pulled over the course of 25×10^7 ps with a velocity of 10^{-6} Å ps^{-1} which equals a total simulation time of 250 μs per pulling trajectory. Although this value is still far from the experimental values of cantilever velocities (10^{-8} Å ps^{-1} (ref. 29)), it represents a significant computational improvement to access the experimental time scale. In experiments, multiple proteins are linked sequentially, and one can observe a number of corresponding peaks, which signal the full unfolding of one protein module. Because of the space resolution, intermediate unfolding states are not detected in AFM experiments. However, in the case of our model we can access these intermediate states with a better resolution and assign to each of them a force peak. The largest of these force peaks, F_{max}, defines the characteristic largest rupture force in the system before unfolding. We construct a set of 50 independent trajectories to assess the mechanostability of the RBD component.

Results and discussion

Our current understanding of both CoV1 and CoV2 has been centered in the structural analysis of the spike (S) glycoprotein. Recently, the structure of the CoV2 S glycoprotein was determined by cryo-EM at high resolution (in the range of 2.8–3.5 Å) and it has been characterized in up (PDB ID: 6VYB[3] and 6VSB[1] entry) and down (PDB ID: 6VXX[3]) conformations. The community is still debating the link between the high spread of COVID-19 and the 30% difference (due to mutations) in CoV2 S protein compared to CoV1, and also whether the concurrent existence of both up and down conformations *in vivo* could be a mechanism for enhancing receptor recognition and subsequently facilitating infection. Only the RBD in the up conformation has been found in close contact with ACE2, indicating the important role of this conformation. In this work we employ extensive molecular dynamics simulations to characterize the molecular features which can differentiate both the CoV1 and CoV2 spikes, and in particular we examine the conformational space of the RBD which is responsible for binding ACE2. A process mostly dominated by the interplay between elastic, electrostatic interactions and environmental properties.[5,30] Here, we highlight structural differences at the level of the most conserved part of the protein. In this regard we define the *consensus residues* as the set residues common in CoV2 and CoV1 after a sequence alignment and full reconstruction. Then we investigate the role of each single conserved residue in terms of the 'native and created non-native contacts' during the dynamics using a contact map determination for proteins.[17] Our analysis indicates that there is a differential change in the set of contacts between *consensus* residues present most of the time in the MD simulations of the full spike that is primarily related to the RBD and its receptor binding module (RBM) subunit (see Fig. 2(a)). In this case the frequency f_{res} of contacts was defined by $f_{res} = C_{res}/N$, where N is the total number of frames used in the analysis (7500), and C_{res} counts the number of contacts computed per residue. We show in Fig. 2 the per residue signed difference between the frequencies from CoV2 and CoV1, $df_{res} = (f_{res}^{CoV2} - f_{res}^{CoV1})$, such that a positive number indicates that CoV2 has effectively more contacts than CoV1, while a negative number indicates

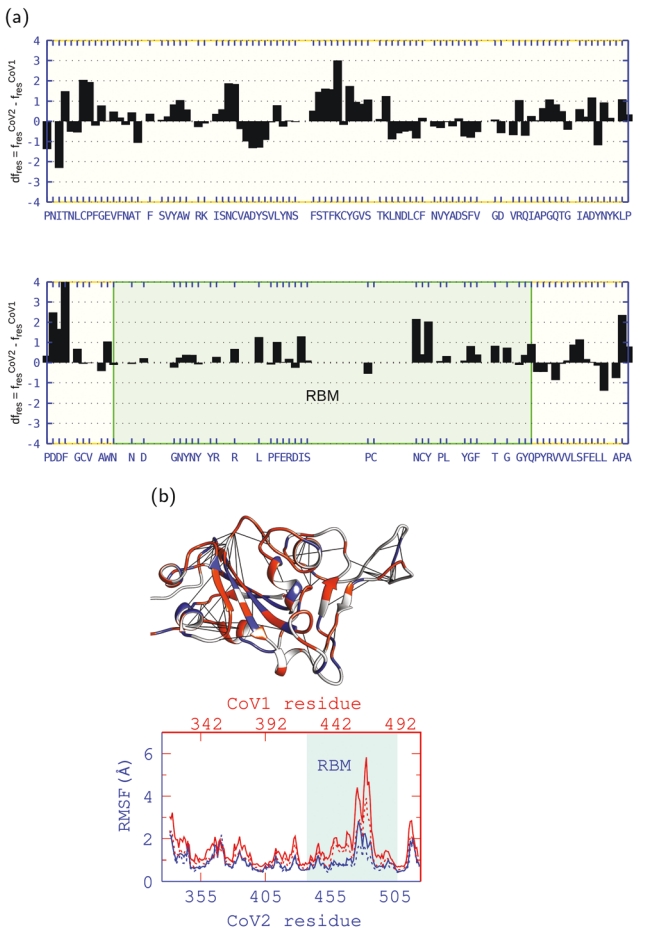

Fig. 2 Panel (a) shows for CoV2 and CoV1 with ACE2 (similar results w/o ACE2) the difference between frequencies $df_{res} = (f_{res}^{CoV2} - f_{res}^{CoV1})$ of chains from CoV2 and CoV1 in contact to ACE2 at each consensus residue, shown in blue color. Panel (b) shows the structure of the mechanostable protein domain in the spike (*i.e.* RBD) based on the computational nanomechanics, where red (blue) residues represent positive (negative) values of df_{res} at consensus residues and lines the additional set of high-frequency contacts of CoV2 not present in CoV1. On the right side is depicted the reduced flexibility as gauged by the RMSF all over the RBD sequence both in the presence of ACE2 (dashed line) and without ACE2 (solid line), which agrees with a larger number of consensus contacts, as shown between brackets in panel (a).

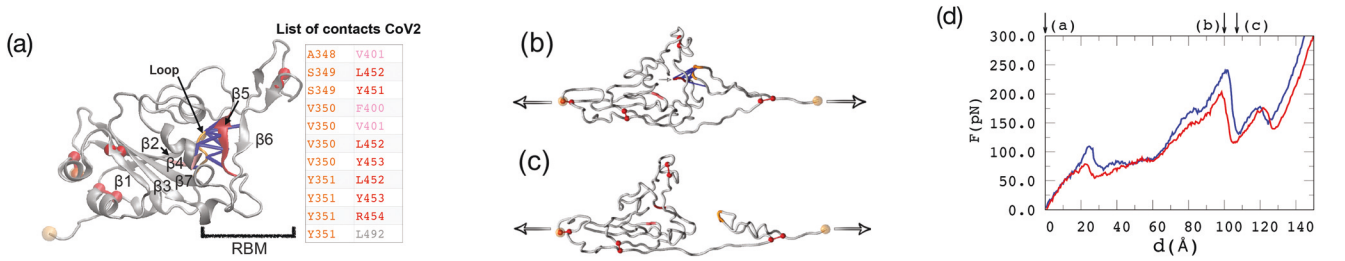

Fig. 3 Pulling simulation of the RBD in up conformation. Top panel shows snaphots of the deformation process at cantilever distance equal to: (a) d = 0, (b) d = 100 and (c) d = 105 and direction of pulling (arrows). Panel (d) shows the pulling force and cantilever distance in simulation for CoV2 (blue) and CoV1 (red). The meaning of F_{max} or rupture force is associated with three protein segments highlighted in red, blue and pink color in panel (a) which also lists 11 native contacts associated with F_{max} in CoV2. Four disulphide bonds are represented as red sticks.

that CoV1 has more contacts. In Fig. 2(b) we depict the additional strong set of consensus contacts present in the conserved part of the protein. This result shows that the CoV2 RBD is structurally more stable compared to CoV1 and this is further supported by the smaller deviation of the root-mean-square fluctuations (RMSF) as we show in Fig. 2.

The observed differences in contacts strongly suggest a larger stability of the CoV2 RBD compared to the CoV1 RBD. To further demonstrate the mechanical stability of the RBD, we employed a validated structure-based modelling approach.[15,28,29,31,32] Our computational structure-based study (see Fig. 3) shows that the RBD is a mechanically stable component in the spike (see next section) and also identifies the structural motif (see Fig. 3) that helps to hold together the structure before the largest force of rupture of intermolecular interactions (F_{max} = 250 ± 11 pN for CoV2 and F_{max} = 200 ± 13 pN for CoV1). The observed difference of ≈50 pN is the equivalent of non-invasive AFM indentation of proteins[33,34] and it suggests that random mutations of the RBD have led to increased stability of the CoV2 RBD compared to the CoV1 RBD. Some deviation from the mean values of our result *via* Single Molecule Force Spectroscopy (SMFS) are expected, but the conclusive differential effect as this point is mostly due to the structural differences.

We also identify the amino acids in CoV2 that contribute to the largest stability in the RDB which turns out to be the same for CoV1. The protein segments involved simultaneously in the maximum force (see Fig. 3) are (A348–A352), (F400–R403) and (N450–R454) which corresponds to a hairpin loop that couples two beta strands together (β4 and β5). The total number of contacts that stabilize those regions are 11 interactions or native contacts which are mostly hydrophobic in character (see Fig. 3(a)). These structural elements are in close contact with loops that interact directly with ACE2. Based on this analysis we would suggest to target new experiments and molecular modelling in antiviral drug design that may disrupt those interactions in the RBD and as a consequence perhaps destabilize the process of cellular recognition. This result highlights the need to launch new rapid research not only in the development of antibodies, entry inhibitors, and antivirals as mechanism of action, but also in new therapies aimed at destabilizing certain key contacts responsible for the increased stability. Also, we expect to motivate nanomechanical studies *via* single molecule force spectroscopy[35,36] of the spike proteins with focus on the RBD in coronavirus system as well as other experiments which can be relevant to elucidate other weaknesses in the protein structure of CoV2 and connect this information with its biological role during the cell recognition process.

Author contribution

A. B. P and R. A. M designed the research; R. A. M., M. C., H. V. G., J. L. B., and A. B. P performed research; R. A. M., H. V. G., J. L. B., H. V. G., M. C. and A. B. P. analyzed data; and R. A. M, H. V. G, J. L. B and A. B. P wrote the paper; and A. B. P supervised the research.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

A. B. P. and R. A. M. thank the financial support from the National Science Centre, Poland, under grant No. 2017/26/D/NZ1/00466. H. V. G thanks the financial support by the Slovenian Research Agency (Funding No. P1-0055). J. L. B. acknowledges use of the ELSA high performance computing cluster at The College of New Jersey. This cluster is funded in part by the National Science Foundation under grant numbers OAC-1826915 and OAC-1828163. M. C. has received support from the National Science Centre (NCN), Poland, under grant No. 2018/31/B/NZ1/00047. The authors gratefully acknowledge the computing provided by the Jülich Supercomputing Centre on the supercomputer JURECA at Forschungszentrum Jülich and additional computer resources were supported by the PL-GRID infrastructure. R. A. M. thanks for source codes gently provided by Marek Cieplak.

References

1 D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham and J. S. McLellan, *Science*, 2020, **367**, 1260–1263.
2 R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo and Q. Zhou, *Science*, 2020, **367**, 1444–1448.
3 A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire and D. Veesler, *Cell*, 2020, **180**, 281–292.
4 Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K.-Y. Yuen, *et al.*, *Cell*, 2020, **181**, 894–904.
5 C. C. Giron, A. Laaksonen and F. L. B. da Silva, *Virus Res.*, 2020, 198021.
6 B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N. Seidah and E. Decroly, *Antiviral Res.*, 2020, **176**, 104742.
7 J. Shang, Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach and F. Li, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 11727–11734.
8 S. Boopathi, A. B. Poma and P. Kolandaivel, *J. Biomol. Struct. Dyn.*, 2020, 1–14.
9 M. Martínez, C. D. Cooper, A. B. Poma and H. V. Guzman, *J. Chem. Inf. Model.*, 2019, **60**, 974–981.
10 R. Zandi, B. Dragnea, A. Travesset and R. Podgornik, *Phys. Rep.*, 2020, **847**, 1–102.
11 P. Ares, C. Garcia-Doval, A. Llauro, J. Gomez-Herrero, M. Van Raaij and P. De Pablo, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2014, **89**, 052710.
12 M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 3694–3699.
13 H. Lu, B. Isralewitz, A. Krammer, V. Vogel and K. Schulten, *Biophys. J.*, 1998, **75**, 662–671.
14 A. Fiser and A. Šali, *Methods Enzymol.*, Elsevier, 2003, vol. 374, pp. 461–491.
15 A. B. Poma, M. Cieplak and P. E. Theodorakis, *J. Chem. Theory Comput.*, 2017, **13**, 1366–1374.
16 M. Chwastyk, A. P. Bernaola and M. Cieplak, *Phys. Biol.*, 2015, **12**, 046002.
17 K. Wołek, À. Gómez-Sicilia and M. Cieplak, *J. Chem. Phys.*, 2015, **143**, 243105.
18 R. A. Moreira, M. Chwastyk, J. L. Baker, H. V. Guzman and A. B. Poma, *Zenodo*, 2020, DOI: 10.5281/zenodo.3817447.
19 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.
20 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
21 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
22 C. W. Hopkins, S. Le Grand, R. C. Walker and A. E. Roitberg, *J. Chem. Theory Comput.*, 2015, **11**, 1864–1874.
23 J. I. Sułkowska and M. Cieplak, *Biophys. J.*, 2008, **95**, 3174–3191.
24 C. Clementi, H. Nymeyer and J. N. Onuchic, *J. Mol. Biol.*, 2000, **298**, 937–953.
25 M. Sikora, J. I. Sułkowska, B. S. Witkowski and M. Cieplak, *Nucleic Acids Res.*, 2010, **39**, D443–D450.
26 M. Sikora, J. I. Sułkowska and M. Cieplak, *PLoS Comput. Biol.*, 2009, **5**, e1000547.
27 A. B. Poma, M. S. Li and P. E. Theodorakis, *Phys. Chem. Chem. Phys.*, 2018, **20**, 17020–17028.
28 A. B. Poma, H. V. Guzman, M. S. Li and P. E. Theodorakis, *Beilstein J. Nanotechnol.*, 2019, **10**, 500–513.
29 A. Valbuena, J. Oroz, R. Hervás, A. M. Vera, D. Rodríguez, M. Menéndez, J. I. Sulkowska, M. Cieplak and M. Carrión-Vázquez, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13791–13796.
30 M. Hernando-Pérez, A. Cartagena-Rivera, A. L. Božič, P. J. Carrillo, C. San Martín, M. G. Mateu, A. Raman, R. Podgornik and P. De Pablo, *Nanoscale*, 2015, **7**, 17289–17298.
31 S. Senapati, A. B. Poma, M. Cieplak, S. Filipek and P. S.-H. Park, *Anal. Chem.*, 2019, **91**, 7226–7235.
32 A. B. Poma, M. Chwastyk and M. Cieplak, *Phys. Chem. Chem. Phys.*, 2017, **19**, 28195–28206.
33 H. V. Guzman, A. P. Perrino and R. Garcia, *ACS Nano*, 2013, **7**, 3198–3204.
34 H. V. Guzman, *Beilstein J. Nanotechnol.*, 2017, **8**, 968–974.
35 M. Krieg, G. Fläschner, D. Alsteens, B. M. Gaub, W. H. Roos, G. J. Wuite, H. E. Gaub, C. Gerber, Y. F. Dufrêne and D. J. Müller, *Nat. Rev. Phys.*, 2019, **1**, 41–57.
36 B. Yang, Z. Liu, H. Liu and M. A. Nash, *Front. Mol. Biosci.*, 2020, **7**, 85.