

Atomistic De-novo Inhibitor Generation-Guided Drug Repurposing for SARS-CoV-2 Spike Protein with Free-Energy Validation by Well-Tempered Metadynamics

Rituparno Chowdhury^{+, [a, b]} Venkata Sai Sreyas Adury^{+, [a]} Amal Vijay^{+, [a]} Reman K. Singh^{+, [a]} and Arnab Mukherjee^{*, [a]}

Abstract: Computational drug design is increasingly becoming important with new and unforeseen diseases like COVID-19. In this study, we present a new computational de novo drug design and repurposing method and applied it to find plausible drug candidates for the receptor binding domain (RBD) of SARS-CoV-2 (COVID-19). Our study comprises three steps: atom-by-atom generation of new molecules around a receptor, structural similarity mapping to existing approved and investigational drugs, and validation of their binding strengths to the viral spike proteins based on rigorous all-

atom, explicit-water well-tempered metadynamics free energy calculations. By choosing the receptor binding domain of the viral spike protein, we showed that some of our new molecules and some of the repurposable drugs have stronger binding to RBD than hACE2. To validate our approach, we also calculated the free energy of hACE2 and RBD, and found it to be in an excellent agreement with experiments. These pool of drugs will allow strategic repurposing against COVID-19 for a particular prevailing conditions.

Introduction

SARS-CoV-2, the cause of the COVID-19 pandemic, is a positive strand RNA beta-coronavirus with large sequence similarities to the SARS-CoV and BatCoV RATG13 RNA viruses.^[1] Despite being known from at least early 2019 to as far back as 2006 as found in a report by Tang *et al.*,^[1] the virus, seemingly incurable and unstoppable, has grown to an enormous scale all across the world.

Therefore, it is essential to find a cure for the virus. Coronavirus is encapsulated by a membrane full of trimeric spike proteins. This spike protein interacts with the peptidase domain (PD) of the human angiotensin-converting enzyme 2 (hACE2),^[2] and hence has been alluded to as a potential target to design preventive and curative therapeutics.^[3] Alternate common targets for viruses are the Main protease (^MPro) and the Non-Structured Proteins (NSPs). For SARS-CoV-2 (Cov2), the high resolution crystal structure of CL3 protease, also known as ^MPro, was resolved recently.^[4]

However, simulation studies have indicated that the enzymatic active site of ^MPro is highly flexible making it less

prone to be inhibited by common viral protease inhibitors.^[6] Zhou *et al.*^[7] mapped protein-protein interaction in a network for proteins involved in Cov2 and came up with many other targets to be used for drug repurposing.

The detailed structural elucidation of Cov2 and hACE2 (hACE2) interface by Shang *et al.*^[8] shows that the spike protein's RBD region remains to be the most important target for drug design for Cov2. The study identifies residues responsible for the interaction and tagged it as receptor binding motif (RBM). RBM is a part of a binding domain called RBD, which again is a part of the S1 region of the spike protein, whose entire structure was solved recently by Wrapp *et al.*^[5] This protein is composed of three subunits. Each of these subunits contains a receptor binding domain (RBD). In two of the subunits, the RBD is in the so-called 'down' configuration and in one it is in the 'up' configuration. However, the exact structure of the RBD complexed to hACE2 and the interactions involved in the site were redetermined very recently by Yan *et al.* by X-ray diffraction.^[9]

Although the spike protein is trimeric^[5] and exists in both open and closed forms,^[10] it is the RBD of the monomer that is responsible for cellular recognition. Its movement is also independent of the rest of the protein.^[11] Figure 1a shows the structure of the spike protein, where RBD is in the 'up' configuration, in contact with hACE2. Figure 1b shows the RBD separately to highlight the interaction hotspot, i.e., RBM^[8] that serves as the hotspot for a potential inhibitor of interaction between the spike protein and hACE2 peptidase domain.

The recent havoc created by the virus resulted in an upsurge of studies towards a remedy.^[12] Most of the computational studies employed massive-docking of the existing molecules, sometimes with an implicit solvent-based MM-PBSA free energy calculations.^[13]

[a] R. Chowdhury,⁺ V. Sai Sreyas Adury,⁺ A. Vijay,⁺ R. K. Singh,⁺ A. Mukherjee
Department of Chemistry
Indian Institute of Science Education and Research Pune, Maharashtra
(India)
E-mail: arnab.mukherjee@iiserpune.ac.in

[b] R. Chowdhury⁺
Department of Chemical Sciences
Indian Institute of Science Education and Research Kolkata, West Bengal
(India)

[⁺] These authors contributed equally.

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/asia.202100268>

This manuscript is part of a special collection celebrating the 15th Anniversary of IISER Inception.

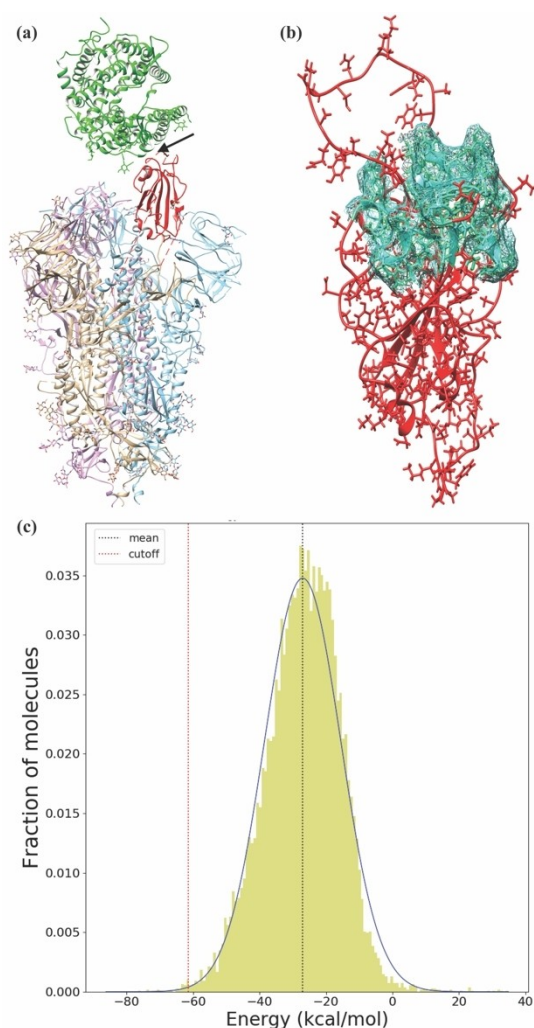


Figure 1. Molecule generation in the RBM. (a) The trimeric structure of the S1 region of the spike protein obtained from the protein data bank with ID 6VSB;^[5] RBD is shown in red and the hACE2 shown in green. The arrow indicates the target for a new inhibitor design. (b) The magnified version of the modelled RBD (in red) with the RBM (hotspot) shown as a mesh surface in cyan. (c) Distribution of interaction energies with RBD for 13516 DeNovo generated molecules. The mean and standard deviations of the fitted Gaussian distribution is -27.0 kcal/mol and 11.5 kcal/mol, respectively. The cut-off (redline) is set at three standard deviations lower than the mean, i.e., at -61.5 kcal/mol, the accepted range.

Although docking is a quick method to sieve plausible candidates for binding, it often fails to ascertain the correct binding constant due to lack of water, ions, and entropic effects resulting from the fluctuations of proteins, ligands and water molecules.^[14] Often docking studies provide a wide range of possible lead molecules, making it difficult to choose the best from them. Despite these pitfalls, docking remains a quick method for finding possible repurposable drugs.^[15] The most reliable and accurate approach to estimate the binding constant, computationally, is to calculate it from an all-atom, explicit water simulation.^[16] To the best of our knowledge, no attempt has been made to find out the binding constants of any of the ligands through such rigorous methods for any of

the target proteins for COVID-19. It has been, however, recently attempted for pseudokinase domain of the JAK2 protein.^[17]

The competitive inhibition by the drug will work if the RBD binds to the drug stronger than it does to hACE2. Therefore, estimating the binding constant accurately is essential. Using surface plasmon resonance experiments, Shang *et al.* reported a value of 44.2 nM (-10.2 kcal/mol at 300 K) binding affinity for the HIS-tagged RBD by covalently immobilizing the protein to the sample substrate.^[8] However, noncovalent association methods, which underestimates the K_d , provide a lower value of 2.9 nM^[10] and 1.2 nM.^[10] The K_d measured by Wrapp *et al.* using biolayer interferometry is 34.6 nM.^[5] The above values when converted to free energy at 300 K temperature yield a range between -10.3 and -12.3 kcal/mol. We will show later that our computational estimate of the above is very close to the above experimental result.

Therefore, our primary aim for this study is to use a reliable computational method to find a molecule that can bind to the hotspot of RBD with binding free energy lower than -10.3 kcal/mol. For that, we have developed a de novo molecule generation program, called 'DeNovo', that creates molecules atom-by-atom in the protein's hotspot (a defined structural region of a biomolecule) to optimize the interaction energy between the two. The idea of such a de novo generation stems from the fact that the chemical space is infinite^[18] and there are molecules in our chemical space that would strongly bind to any given receptor, if only we can find them. Although the program is quite general and can grow molecules for any hotspot, we apply it for the first time here to grow inhibitors for the RBD by targeting the RBM region (Figure 1b). Our approach will be particularly helpful for creating inhibitors where existing drugs have already started facing resistance and a completely new design for a drug is essential.^[19] The current problem of COVID-19, however, demands a different solution, where we need to provide molecules worthy of immediate clinical trials.

Hence, from the several de novo molecules we generated, we selected the top 35 and using them we found analogous drugs from the DrugBank^[20] that could be repurposed for COVID-19 using a similarity-based mapping. We have validated our approach by performing the computationally rigorous all-atom, explicit water well-tempered metadynamics^[21] free energy calculations for all these molecules and show that 9 molecules (4 de novo and 5 repurposed drugs) have free energy of binding to the RBD lower than -10.3 (the cut-off set above), implying their highly promising potential to inhibit the viral attack on human protein. As the benchmark, we calculated the free energy of binding between RBD-hACE2 which is in good agreement with experiments. Also, performing rigorous well-tempered metadynamics simulations with multiple coordinates helped us identify the binding mechanism of all these molecules.

Therefore, our study provides a new, viable, and successful approach for finding novel and repurposed drugs for COVID-19. Our approach is general, and it has the potential to be used for any other receptor-mediated drug design as well.

Results and Discussion

A. Generation of specific binders for a receptor

Our DeNovo program (see Method) is schematically shown in Figure S1 of the supporting information (SI). Using configurational bias Monte Carlo (CBMC) approach, schematically shown in Figure S2 of SI, we generated 13516 molecules having 10 to 50 heavy atoms to cover a broad spectrum of molecular scaffolds. The distribution of interaction energy between the molecule and the RBD is shown in Figure 1c. We have chosen ~0.3% of the molecules by selecting those that have interaction energy less than a cut-off value of -61.5 kcal/mol (Figure 1c), set at three standard deviations lower than the mean energy -27 kcal/mol.

Thus, we gathered 35 molecules (see Figure S3 of SI). The names of the molecules are chosen based on an internal criteria of numbering. At this point, we have a set of good molecules to work with. However, our molecule generation had a few caveats: (1) our hotspot was rigid, (2) there were no water molecules (the screening effect of solvent in the energy calculation of the DeNovo method was accounted for by using dielectric constant of 25.0), (3) and the entropy contributions could not be taken into account.

Therefore, to calculate the exact binding free energy, to the extent that a classical force field can provide, we have subjected the aforementioned molecules to a state-of-the-art rigorous well-tempered^[21] version of metadynamics^[22] simulation as described later.

B. Route to repurposing

As we will see later, the DeNovo generated molecules are, from a computational point-of-view, indeed strong binders and some of which could be candidate drugs. However, synthesis and toxicology study for these molecules may take much longer. Relying on the quality of binding of our de novo molecules (as shown later via free energy calculations) and building on the concept of similar molecules having similar chemistry,^[23] we looked for similar molecules in the DrugBank^[20] using Tanimoto (or Jaccard)^[24] similarity search which uses bitwise fragment comparison to accurately match structures based on chemical fragments.^[25] This similarity algorithm generally preserves relative positions of functional groups. We have chosen drugs for each of our 35 molecules with the Tanimoto coefficient ≥ 0.4 , following the recommendation of Baldi *et al.* who showed that a Tanimoto score of 0.4 is significant for a database of over 10,000 molecules.^[26] This similarity search led to several drug molecules for each of the 35 de novo molecules, listed in Table S1 with their similarity score in brackets. After removing the irrelevant drugs (shown in red) (see the full list and explanation in SI), we ended up with 123 unique drugs. Given the difficulty and time required to perform free energy calculations for all these molecules, we docked all these drugs to the same hotspot (Figure S4 of SI), using docking score purely as a sieving criterion, and selected molecules with a

docking score < -8.0 kcal/mol. This narrowed down our list to 20 potentially repurposable drugs. Figure S5 of SI shows the chemical structures of the chosen drug molecules and their docking scores are given in Table S2. As docking score is empirical and varies between different softwares, we carried out the free energy simulations of these drug molecules with well-tempered^[21] metadynamics method, as well.

Note that, several other molecules in the DrugBank database could have had docking scores meeting our cut-offs had we not applied the similarity criteria to narrow down the chemical space of molecules similar to the ones generated by our DeNovo program. Docking, therefore, helps us to narrow down the number of free energy calculations by removing the potentially weak binders.

To capture the relationships between the de novo and drug molecules, we plot the similarity matrix of all the 55 molecules in Figure 2. This matrix is obtained by measuring similarity-based clustering of each pair of molecules. This matrix provides valuable information on the structural diversity of molecules that have led to favorable properties. We find that there are two separate clusters, one dominated by repurposed drugs and the other dominated by de novo molecules.

We note that the molecular similarity reflects on the similarity in binding free energy as well. The best-binding de novo molecules (as scored based on DeNovo's interaction energy) are in themselves quite similar, as are the best-binding repurposed drugs among themselves. This is particularly noticeable once the molecules have been mapped to their free energies. Even though it may seem at the first glance that the cross-similarity is less, there are marked, and sporadically distributed regions of cross-similarity observed between the de novo molecules and repurposed drugs, which can be attributed to appreciable similarities in their maximal common substructures.

The left panel of Figure 2 shows the structures of some of the best molecules along with their free energy (see discussion below). It is reassuring that we obtained several molecules, both from our de novo generation and repurposing strategy, that are comparable or stronger in binding to RBD than hACE2.

C. Free Energy Surfaces of Protein-Protein and Protein-Drug Complexes: An induced fit Mechanism

Free energy calculation with an all-atom explicit water system is by far the most accurate, albeit expensive, estimate of binding among various computational methods.^[16] Metadynamics is an extremely popular state-of-the-art method to calculate free energy surface for complex systems^[27] and it has been shown to reproduce experimental observations closely.^[28] Recently, attempts to estimate free energy of binding was done for the protease, albeit with an approximate method called MM/PBSA,^[29] which takes water as a continuum, thus leading to an inaccurate estimation of entropy, and furthermore due to the lack of accurate entropy estimation, it is also not ideally suited for providing mechanistic insights. All-atom with explicit water free energy calculation not only estimates the free energy more

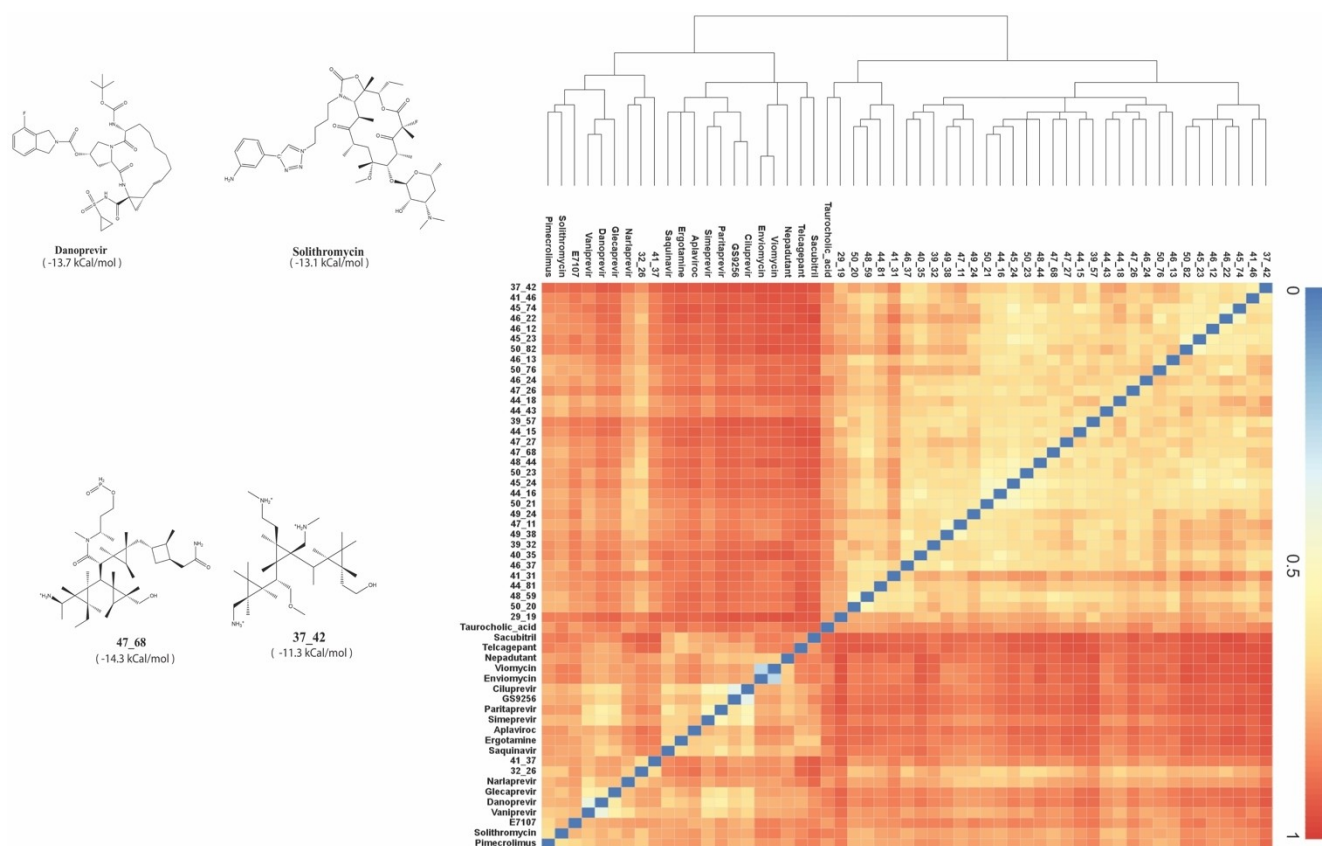


Figure 2. Select drug molecules and similarity heatmap. (left) Two of the best DeNovo generated molecules(bottom) and two repurposed drug molecules(top) with their calculated binding free energies (in brackets underneath each molecule). (right): the heat map of the Tanimoto similarity amongst the chosen list of de novo and repurposed drug molecules. The hierarchical diagram also depicts the similarity between different molecules. Two clusters of de novo molecules and drugs are clearly visible. However, cross similarities are also present.

accurately, it also captures the mechanism of the binding. In this study, we have performed the free energy calculations for the huge list of 55 ligand molecules (35 are de novo and 20 are known drug molecules). To the best of our knowledge, this study is the first to perform the free energy calculations for the spike protein's interaction with hACE2 and other ligands.

Before calculating the free energy for all the 55 drugs molecules, we validated our approach by calculating the free energy surface for RBD and hACE2 for comparison. Starting with the available crystal structure of the spike protein and hACE2 (PDB id 6VSB^[5]), we first selected only the RBD region, modelled the loop region (see method), solvated the system with water and physiological concentrations of ions and performed multiple, long (~230 ns) well-tempered metadynamics simulations against multiple collective variables such as DISTVEC (displacement along a body-fixed vector) (Figure S6 of SI) and native contact (Figure S7 SI) to study the binding free energy surface of these two proteins. While native contact helps to untangle the interactions between two proteins, DISTVEC is a vectorial displacement coordinate that helps move the proteins apart. These coordinates were used successfully in our previous studies of drug intercalation^[28a,30] and protein-protein interactions.^[31]

Figure 3 shows the free energy surface of RBD and hACE2, which verifies the crystal structure configuration as the global free energy minimum. A few snapshots of the configuration of both proteins are shown along the unbinding pathway which depicts that the major reason for stability is due to direct protein-protein interaction. As soon as the proteins detach

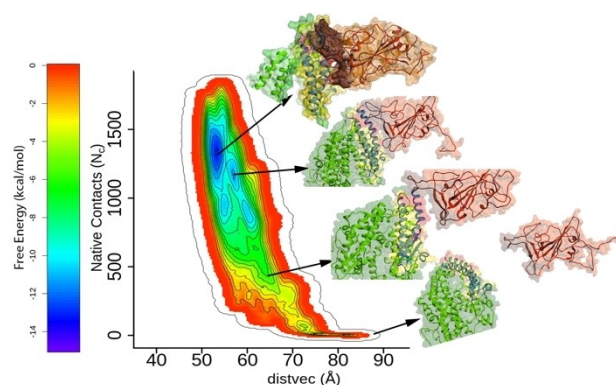


Figure 3. Free Energy of hACE2-RBD. binding Free energy surface of RBD (brown) binding to hACE2 (green) against two collective variables, native contact (Nc) and DISTVEC (Å). The structures along the path of dissociation are shown along the free energy bar diagram.

themselves, the free energy increases. The proteins thus move apart from each other, as reflected in the consequent increase in DISTVEC and a significant decrease in the number of native contacts. From this free energy surface, we estimate the binding free energy to be -13.3 kcal/mol, which is in close agreement with the experimental results. A second metadynamics simulation of the same system yielded free energy estimate of -12.7 kcal/mol making the average -13.0 kcal/mol.

Using the same set of collective variables and starting with either the docked configuration for repurposable drug molecules or the DeNovo generated configurations for the novel molecules, we performed metadynamics simulations for all the 55 molecules and calculated their free energy surfaces.

Figure S8 to S11 in SI show the two-dimensional free energy surfaces (FES) of binding of all the molecules. In Figure 4, we show the FES of three top de novo molecules and 3 top drugs. A representative structure of the most stable configuration obtained from the metadynamics simulation is shown either above or below the FES of the respective molecule. Figure 4 shows that the molecules are nicely packed inside the hotspot region.

Since each molecule behaves differently, it is not possible to find a unifying trend in their binding mechanism. However, most of the strong binding molecules have a narrow free energy profile along the native contact. Once the contacts (the short distance between the ligand and the protein) are broken,

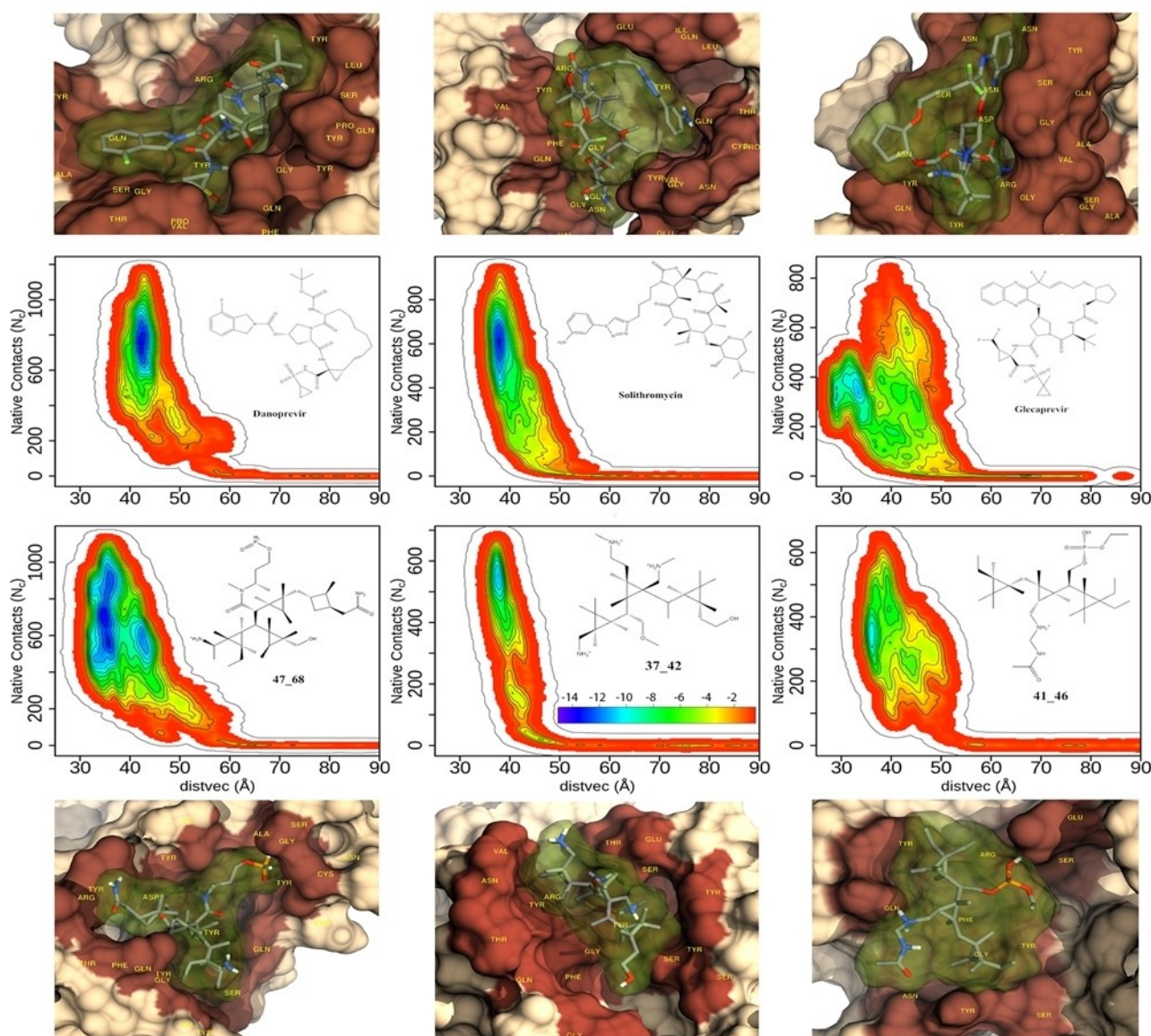


Figure 4. Free energy surfaces (FES) of three best de novo molecules and three best repurposed drug molecules. The chemical structure and name of each molecule are shown in the inset. The free energy bar used to plot FES is shown in the inset of 37_42. The structures of the free energy minimum of each molecule, made using chimera,^[32] are shown above/below the corresponding FES. The protein is drawn in surface representation while the drug is shown with both stick and surface (with 80% transparency). Note the deep cavity formed by each of these molecules in the protein.

well with experimental value -12.3 kcal/mol. The same protocol was applied to all the 55 molecules (each with more than 100 ns of simulations) to obtain their free energy of binding to RBD. Out of these, we obtained 21 molecules with free energy less than -9 kcal/mol, indicating the success of our approach. Moreover, at least four de novo molecules and five repurposed drug molecules bind to RBD at comparable or higher strength than the hACE2. This implies that these molecules could in principle inhibit the interaction of the RBD with hACE2 and prevent the viral attack on the human cell.

Methods

DeNovo: The atom-by-atom synthesis of a strong binder

Here we describe our algorithm of de novo atom-by-atom construction of molecules in the hotspot of a receptor. The algorithm is based on the fact that any molecule can be represented as a graph where nodes are atoms and the bonds are edges. The first evidence of graph representation of the molecules dates back to 1867 by Kekule. Recently, a graph-based algorithm was used to create 166 billion molecules in vacuum from only 17 atoms^[33] of C, N, O, S, and halogens, which reinforces the fact that the chemical space is infinite and therefore it is possible to get multiple strong binders for each hotspot.

We, however, grow the molecules in the protein hotspot from the similar set of atoms (C, N, O, S, etc.). The classical nonbonding interaction energy (both van der Waals and electrostatic) between the drug and the hotspot is calculated using CHARMM27^[34] force field at each stage and the generation proceeds following an algorithm similar to that of the configuration bias Monte Carlo (CBMC)^[35] (see Figure S1 and S2 and corresponding discussion in SI). Prior to molecular generation, the incomplete residues were completed using xleap of AMBERTools^[36] and the missing residues of the protein were modelled by Modeller 9.21.^[37] We targeted a prefixed atom number for the molecule and tried to improve the interaction of the molecule using CBMC criteria. This way, we obtained molecules having between 10 to 50 heavy atoms. Unlike in CBMC, our molecule is made of many different atoms. We used geometric criteria (equilibrium distance and angles) for the formation of rings. Finally, when a molecule reaches the desired number of atoms, we use CBMC criteria to accept it with reference to the previously generated molecule. We repeat this protocol several times until we have the required number of molecules. The program is fully CPU-parallelized and runs efficiently over several cores (here we used 48 cores for our molecule generation).

Similarity Measurement

We have chosen all de novo generated molecules that interact with the protein with energy lower than our cut-off -61.5 kcal/mol (see Figure 2). Subsequently, we performed a similarity search using the DrugBank^[20] search engine employing Tanimoto algorithm.^[38] All approved and investigational drug molecules with similarity above 0.4, a cutoff based on the studies of Baldi et al.,^[26] were considered. Table S1 of SI lists all the drug molecules for each of the 35 de novo molecules that match the criterion.

Docking of the matched drugs

Based on the similarity measurement, we came up with a list of 123 unique drug molecules from our 35 de novo molecules. We then

docked these drug molecules to the hotspot using AutoDock Vina version 1.1.2.^[39] A box centered around SER443 is created with dimensions $36.7 \text{ \AA} \times 26.1 \text{ \AA} \times 40.9 \text{ \AA}$ and default vina parameters. The docking setup is shown pictorially in Figure S4 of SI. The docking scores for all these molecules are shown in bracket in Table S1. The drug molecules with docking score less than -8.0 kcal/mol were considered for free energy calculations.

Force Field Generation

All drug molecules were optimized using HF (Hartree Fock) theory with 6-31G* basis set using Gaussian 09 software.^[40] Thereafter, antechamber^[41] was used for the RESP charge calculation and generation of force field for the molecules. The topology and coordinates were then converted into the GROMACS format by using a python script acpype.py (available at <https://github.com/t-acpype>).

System Setup for Simulation

The starting structure of the SARS-CoV-2 spike glycoprotein (PDB ID: 6VSB) was obtained from the Protein Data Bank. Modeller 9.21^[37] was used for modeling the missing residues, which predicted five three-dimensional structural forms using chain A of protein as the template. The best possible structure was predicted considering the DOPE (Discrete Optimized Protein Energy) score.^[42] All of the simulations were performed using molecular dynamics software GROMACS 2019.6.^[43] For the study of the protein-ligand complex, we only considered the RBD region of the protein (residues 302 to 506) and topology was prepared using AMBER99SB force field.^[44] Each complex system was solvated by ~ 23000 TIP3P^[45] water molecules in a box of dimensions $70 \times 70 \times 180 \text{ \AA}^3$. The physiological concentration (150 mM) of Na^+ and Cl^- ions along with extra Cl^- ion were used to neutralize the system.

Equilibration and Simulation

Initially each system was energy minimized using steepest descent method^[46] for 10000 steps, followed by heating it to 300 K in 200 ps using Berendsen thermostat and barostat^[47] with coupling constant of 0.6 ps. Restraints of $25 \text{ kcal/mol/\AA}^2$ were applied on heavy atoms during the heating process. Thereafter, equilibration was carried out for 2 ns at constant temperature (300 K) and pressure (1 bar) without any restraints using same thermostat and barostat with coupling constants of 0.2 ps each. The last 100 ps of NPT simulation was used to calculate the average volume the same, which was used in the final 5 ns unrestrained NVT equilibration using the Nosé-Hoover thermostat^[48] with coupling constant of 0.2 ps. During the simulation, LINCS algorithm^[49] was used to constrain all the bonds and Particle Mesh Ewald (PME) method^[50] was used for electrostatics. The distance cut-offs for the van der Waals (vdW) and electrostatic long-range interaction was kept at 10 \AA . The time step for each simulation was taken to be 2 fs.

Free energy calculation using Metadynamics

The equilibrated ligand-bound protein structure was initially simulated for 5 ns. If the ligand was found to be bound after 5 ns simulation, we subjected the system to free energy calculation. To calculate the binding free energy of drugs, well-tempered metadynamics^[21] simulations were performed after equilibration using DISTVEC and native contacts as collective variables (see SI for full description). We performed long (~ 100 ns) metadynamics simulation with a hill height of 0.2 kJ/mol, a bias factor of 10, and a

hills deposition rate of 2 ps. Gaussian widths for DISTVEC and native contacts were taken to be 0.6 Å and 5, respectively. An upper wall restraint was applied at 45° on the angle between two vectors \hat{b} and \hat{d} , as shown in Figure S6 of SI. For free-energy calculations, PLUMED 2.6^[51] was used along with GROMACS. The system size and run lengths of all the systems are provided in the Table S3 of SI.

Supporting Information

DeNovo algorithmic flow charts, schematic CBMC profile, image containing all the selected de novo molecules, table showing list of similar drugs, image showing docking region, table of docking scores for selected drugs, table for docking variations, image of all the selected drug molecules, definition and picture of the collective variables used, table showing system size and other information for all the systems, and free energy surfaces of binding to RBD for all the 55 molecules.

Acknowledgements

The work is partly funded by the department of biotechnology, India (BT/PR34215/AI/133/22/2019). Authors acknowledge the support and the resources provided by 'PARAM Brahma Facility' under the National Supercomputing Mission, Government of India at the Indian Institute of Science Education and Research, Pune. Authors also acknowledge help of system administrator Nisha Kurkure. Scientific discussions with Dr. Kausik Chakraborty from IIGB and Samarпита Sen from IISER Kolkata are gratefully acknowledged. RC (fellow id SX-1411075) and VSSA (fellow id: SX-1711011) thank KVPY and AV thanks CSIR for their fellowships.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Covid-19 · de Novo drug design · Spike protein · human ACE2 · repurposing therapeutics · docking · molecular dynamics · free energy · well-tempered metadynamics

- [1] X. C. Tang, J. X. Zhang, S. Y. Zhang, P. Wang, X. H. Fan, L. F. Li, G. Li, B. Q. Dong, W. Liu, C. L. Cheung, K. M. Xu, W. J. Song, D. Vijaykrishna, L. L. M. Poon, J. S. M. Peiris, G. J. D. Smith, H. Chen, Y. Guan, *J. Virol.* **2006**, *80*, 7481–7490.
- [2] X. Y. Ge, J. L. Li, X. L. Yang, A. A. Chmura, G. J. 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. Cramer, S. Y. Zhang, L. F. Wang, P. Daszak, Z. L. Shi, *Nature* **2013**, *503*, 535–+ +.
- [3] L. Y. Du, Y. X. He, Y. S. Zhou, S. W. Liu, B. J. Zheng, S. B. Jiang, *Nat. Rev. Microbiol.* **2009**, *7*, 226–236.
- [4] 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, H. Yang, *Nature* **2020**.
- [5] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, *Science* **2020**, *367*, 1260–1263.
- [6] M. Bzówka, K. Mitusińska, A. Raczynska, A. Samol, J. Tuszyński, A. Góra, *bioRxiv* **2020**, 2020.2002.2027.968008.
- [7] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, *Cell Discov* **2020**, *6*, 14.
- [8] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, *Nature* **2020**.
- [9] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, *Science* **2020**.
- [10] A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Velesler, *Cell* **2020**, *181*, 281–292 e286.
- [11] M. Letko, A. Marzi, V. Munster, *Nat. Microbiol.* **2020**, *5*, 562–+ +.
- [12] G. Colavizza, R. Costas, V. A. Taag, N. J. van Eck, T. van Leeuwen, L. Waltman, *bioRxiv* **2020**, 2020.2004.2020.046144.
- [13] a) J. C. Smith, M. Smith, **2020**; b) P. K. Panda, M. N. Arul, P. Patel, S. K. Verma, W. Luo, H.-G. Rubahn, Y. K. Mishra, M. Suar, R. Ahuja, *Sci. Adv.* **2020**, *6*, eabb8097.
- [14] T. Scior, A. Bender, G. Tresadern, J. L. Medina-Franco, K. Martinez-Mayorga, T. Langer, K. Cuanalo-Contreras, D. K. Agrafiotis, *J. Chem. Inf. Model.* **2012**, *52*, 867–881.
- [15] R. K. Guy, R. S. DiPaola, F. Romanelli, R. E. Dutch, *Science* **2020**.
- [16] L. Wang, Y. J. Wu, Y. Q. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel, *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
- [17] K. J. Cutrona, A. S. Newton, S. G. Krimmer, J. Tirado-Rives, W. L. Jorgensen, *J. Chem. Inf. Model.* **2020**.
- [18] P. Kirkpatrick, C. Ellis, *Nature* **2004**, *432*, 823–823.
- [19] S. B. Levy, B. Marshall, *Nat. Med.* **2004**, *10*, S122–S129.
- [20] D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, *Nucleic Acids Res.* **2008**, *36*, D901–906.
- [21] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*.
- [22] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.
- [23] P. Willett, *Drug Discovery Today* **2006**, *11*, 1046–1053.
- [24] P. Jaccard, *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 547–589.
- [25] D. Bajusz, A. Racz, K. Heberger, *J. Cheminformatics* **2015**, *7*.
- [26] P. Baldi, R. W. Benz, *Bioinformatics* **2008**, *24*, 1357–1365.
- [27] G. Bussi, A. Laio, *Nat. Rev. Phys.* **2020**, *2*, 200–212.
- [28] a) W. D. Sasikala, A. Mukherjee, *Phys. Chem. Chem. Phys.* **2013**, *15*, 6446–6455; b) R. Capelli, A. Bochicchio, G. Piccini, R. Casanovas, P. Carloni, M. Parrinello, *J. Chem. Theory Comput.* **2019**, *15*, 3354–3361.
- [29] J. Wang, *J. Chem. Inf. Model.* **2020**.
- [30] A. Mukherjee, W. D. Sasikala, *Adv. Protein Chem.* **2013**, *92*, 1–62.
- [31] R. K. Singh, W. D. Sasikala, A. Mukherjee, *J. Phys. Chem. B* **2015**, *119*, 11590–11596.
- [32] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- [33] L. Ruddigkeit, R. van Deursen, L. C. Blum, J. L. Reymond, *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875.
- [34] A. D. MacKerell, N. Banavali, N. Foloppe, *Biopolymers* **2001**, *56*, 257–265.
- [35] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press Inc., 6277 Sea Harbor Drive Orlando, FL United States, **1996**.
- [36] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, P. A. Kollman (2020), AMBER 2020, University of California, San Francisco.
- [37] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, A. Sali, *Curr. Protoc. Bioinforma.* **2006**, Chapter 5, Unit-5 6.
- [38] J. W. Godden, L. Xue, J. Bajorath, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 163–166.
- [39] D. S. Goodsell, G. M. Morris, A. J. Olson, *J. Mol. Recognit.* **1996**, *9*, 1–5.
- [40] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. M. Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O.

- Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision A.02, 2009.
- [41] J. M. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- [42] M. Y. Shen, A. Sali, *Protein Sci.* **2006**, *15*, 2507–2524.
- [43] M. J. Abraham, D. v. d. Spoel, E. Lindahl, B. Hess, the GROMACS development team, GROMACS version 2019, 2019.
- [44] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 712–725.
- [45] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
- [46] W. H. T. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, CAMBRIDGE UNIVERSITY PRESS, 2007.
- [47] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [48] S. Nose, *Mol. Phys.* **1984**, *52*, 255–268.
- [49] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- [50] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [51] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604–613.

Manuscript received: March 17, 2021
 Revised manuscript received: April 20, 2021
 Accepted manuscript online: May 5, 2021
 Version of record online: May 18, 2021