



Benchmarking ANI potentials as a rescoring function and screening FDA drugs for SARS-CoV-2 M^{pro}

Irem N. Zengin¹ · M. Serdar Koca^{2,3} · Omer Tayfuroglu¹ · Muslum Yildiz² · Abdulkadir Kocak¹

Received: 3 January 2024 / Accepted: 26 February 2024
© The Author(s) 2024

Abstract

Here, we introduce the use of ANI-ML potentials as a rescoring function in the host–guest interaction in molecular docking. Our results show that the “docking power” of ANI potentials can compete with the current scoring functions at the same level of computational cost. Benchmarking studies on CASF-2016 dataset showed that ANI is ranked in the top 5 scoring functions among the other 34 tested. In particular, the ANI predicted interaction energies when used in conjunction with GOLD-PLP scoring function can boost the top ranked solution to be the closest to the x-ray structure. Rapid and accurate calculation of interaction energies between ligand and protein also enables screening of millions of drug candidates/docking poses. Using a unique protocol in which docking by GOLD-PLP, rescoring by ANI-ML potentials and extensive MD simulations along with end state free energy methods are combined, we have screened FDA approved drugs against the SARS-CoV-2 main protease (M^{pro}). The top six drug molecules suggested by the consensus of these free energy methods have already been in clinical trials or proposed as potential drug molecules in previous theoretical and experimental studies, approving the validity and the power of accuracy in our screening method.

Keywords Molecular docking · Scoring function · Interaction energy · Binding mode · Rescoring

Introduction

Designing a therapeutic molecule to treat a disease can be quite expensive and time consuming. Preclinical processes such as identification of drug targets, High Throughput Screening (HTS) experiments on the target, generating a library of hit compounds, optimizing potency of bioactive compounds and preceding clinical trials require tremendous amount of resources. Typically, a drug can cost more than 2 billion US dollars and 10–15 years to be approved and enter the market [1–3]. Since repurposing an FDA approved drug is a much faster and less expensive strategy it has become one of the most popular approaches in recent

drug development endeavors. Even testing all FDA approved drugs experimentally may not be affordable in terms of economic burden and time. To help ease this problem, computational methods have been utilized in rational drug design.

In searching for drug candidates as potential inhibitors, binding mode and binding affinity are two critical questions that ought to be addressed. Docking methods are capable of screening ultra large libraries of compounds by aforementioned approaches (billions of molecules) with much reduced cost and time. Docking methods rely on a search algorithm and scoring function that are physics based, empirical, knowledge-based and Machine Learning based which compromise on accuracy to increase the speed [1, 4, 5]. Despite the great success of the docking methods, there is still a need for improving scoring functions so as to reproduce experimental results.

Machine Learning (ML) methods have just started to gain attention to advance the docking methodology recently similar to other computational techniques. The main objective of the ML techniques is to find an accurate and a fast solution to an existing problem by learning from previous experimental data. This is achieved by supervised, unsupervised and reinforcement methods [6]. Supervised methods are based

✉ Abdulkadir Kocak
kocak@gtu.edu.tr

¹ Department of Chemistry, Gebze Technical University, 41400 Gebze, Kocaeli, Turkey

² Department of Molecular Biology and Genetics, Gebze Technical University, 41400 Gebze, Kocaeli, Turkey

³ Present Address: Pfizer - Universidad de Granada - Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), 18016 Granada, Spain

on training an algorithm on a set of inputs to find an output while unsupervised learning algorithms find a pattern in data set predict results from this arrangement. On the other hand, reinforcement learning, the input progress in an environment and uses the data learned from the experience [1]. Recent studies have shown the outperformance of ML based scoring functions over classical ones, which directly use a known mathematical function [7–9]. ML based scoring functions are usually used by means of rescoring [10] due to the dependence on the training dataset [11, 12].

ANI-1, ANI-1x and ANI-1cxx have been trained to calculate DFT and CCSD(T) energies of small organic compounds containing C, H, O and N atoms in non-equilibrium conformations [13–15]. The extended version, ANI-2x, has been shown to predict DFT energies of equilibrium or non-equilibrium conformations of molecules containing C, H, O, N, S, F and Cl atoms. It has been shown to reproduce the energies at the accuracy of the ω b97x/6-31G* level with millions of times faster than the actual QM calculations [16].

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the coronavirus disease pandemic in 2019 [17]. Since the first case of COVID-19 in 2019, more than 6.8 million deaths have been reported worldwide (World Health Organization. Weekly Epidemiological Update on COVID-19. [18]). The coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has destroyed health systems, societies, and economies. After the first identification of the SARS-CoV-2 strain coming from Wuhan, several variants of concern (VOCs) have been identified. Clinical reports and epidemiological features of the infection indicate a relatively mild disease flow and increased human-to-human virus spread [19–21]. Improved COVID-19 treatment and prevention techniques are urgently required given the increasing SARS-CoV-2 infection rate and lack of efficient treatment options. Researchers around the world are working to develop treatments and vaccines to struggle the disease, and several drugs have been approved for emergency use by regulatory agencies. For now, only two SARS-CoV-2 oral drugs (Paxlovid and Molnupiravir) have been clinically approved [22, 23]. Nevertheless, the antiviral potency of these compounds is not adequate to deal with the pandemic. Paxlovid (Nirmatrelvir with Ritonavir) has been reported to permit the SARS-CoV-2 replication upon completion of 5 day long oral take [24]. In addition to limited number of vaccines and drugs, the emergence of drug resistant variants of SARS-CoV-2 brings additional concerns [24, 25] about the use of those drugs [26]. Therefore, developing anti-viral oral therapeutics for COVID-19 is still highly demanded.

Similar to other viral infections, once entering the host cell cytoplasm, the SARS-CoV-2 viral genome is

translated into approximately 30 proteins. At first, 16 of these proteins are translated as two polyproteins, and in order to continue infection, these polyproteins must be split apart into the two proteins that virally encoded proteases; the major viral protease known as 3CL^{pro} (or M^{pro}), and the protease papain-like protease (PL^{pro}), facilitate the cleavage of these polyproteins into structural and non-structural proteins (NSPs) [27], which play a crucial function in the transcription/replication during the infection. Since M^{pro} is a critical enzyme in the life cycle of the virus [28], it has been a validated high-profile antiviral drug target, and its inhibitors have been shown to have strong antiviral activity in cell cultures and animal models [22, 29–37]. Other coronavirus enzymes, like the helicase and the RNA-dependent RNA polymerase, has potential to be target for the development of antiviral drugs, such efforts are currently limited because these enzymes don't have crystal structures [37, 38]

Main protease (M^{pro}) is a three-domain (domains I–III) cysteine protease consisting of 306 amino acids. Antiparallel β -barrel structures takes place domain I and domain II (residues 8–101/residues 102–184, respectively), and domain III (residues 201–303) is connected to domain II through a long loop region (residues 185–200) that contains 5 α -helices arranged in one substantially antiparallel globular cluster [28]. The glutamine residue in the P1 position of the substrate is cleaved by M^{pro} using the protease Cys145-His41 dyad, where the cysteine thiol serves as the nucleophile in the proteolytic reaction.

Designing a therapeutic molecule from the ground up to treat an illness can become extremely costly and a time taking process. In particular, for pandemics with such high transmission rates like SARS-CoV-2, requires prompt actions for controlling the disease. Instead, drug repurposing strategy might offer inexpensive and faster solutions to the problem. Repurposed drugs that have existing clinical data on the effective dose, treatment duration, side effects, and toxicity could be rapidly translated into the treatment of patients.

High throughput virtual screening by means of drug repurposing have been used to identify safe-in-human drugs with potential anti-SARS-CoV-2 properties. Most of these studies are based on molecular docking. Although it is the gold standard method to find the binding mode, docking is a very coarse method and almost never predicts the correct experimental binding affinity trend among the inhibitors. One strategy to overcome this limitation is to use consensus scoring [39]. More sophisticated methods to calculate the potential binding free energy of inhibitor candidate to the protein ranges from post molecular dynamics simulations such as Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) to perturbation methods such as Bennett acceptance ratio (BAR), the latter being much more accurate

yet quite costly. Although there have been numerous attempts that combine docking and MD based free energy methods in virtual screening, the number of candidates is mostly reduced to tens of hit compounds prior to MD simulations in these studies due to computational cost. Therefore, the top pose from docking is usually used in MD simulations and free energy calculations. Although different docking algorithms are successful in prediction of correct binding mode in the top three poses [40], MD simulations towards screening are performed only on the first pose. This limits the success of the binding free energy calculations (BFE) when the ligand is totally mis-oriented (such as flipped) among the top poses in docking since one cannot expect the MD simulations' correcting these drastic changes no matter how long the simulations performed. Therefore, running separate MD simulations for all the top poses may become a necessity to correctly predict BFEs.

One-trajectory approach end-state BFE calculations are quite attractive since they require only one MD simulation for each protein–ligand (PL) complex system with less computational cost. Although relatively better than docking scores, the accuracy in end-state BFE methods is still low due to over-simplifications such as implicit solvent definition in the case of MM-P(G)BSA and molecular mechanics (MM) definition of the Hamiltonian of the system. On the other hand, several new implementations have been introduced to end-state methods to improve accuracy. We have recently implemented the use of ML based potentials as a post MD simulation to improve the accuracy of BFE calculations [41, 42]. Herein, we introduce a unique virtual screening protocol to overcome aforementioned drawbacks.

Here, we first investigate and unravel the capability of ANI potentials as a rescoring function in molecular docking. Our results show that the docking power of ANI potentials can compete with the current scoring functions at the same level of computational cost. We then screened a library of 2500 clinically used drugs, either approved for human use or with extensive safety data in humans (phase 2 or 3 clinical trials), for their ability to bind SARS-CoV-2 Mpro using consensus docking scores of ANI/GOLD and free energies by MMPBSA/ANI_LIE.

Computational methods

CASF-2016 dataset

In order to test ANI's performance as a rescoring function, we have used CASF-2016 dataset, which has been created for benchmarking purposes in docking algorithms. The dataset is composed of 285 protein–ligand complexes and each complex has a native binding pose as well as maximum 100 decoy ligand binding poses, selected from

a normal distribution of 1000 poses generated by three different docking software with $\text{RMSD} \leq 10 \text{ \AA}$ from the crystal structures [40, 43, 44]. In the study, 34 different docking algorithms/software were tested in terms of “ranking power”, “scoring power”, “docking power” and “screening power”. Docking power is defined as whether a scoring function can properly differentiate the native binding pose from all the decoys within the top 1, 2, or 3 scores. In addition, when the different decoys create a potential energy surface in the binding region, the native pose should correspond to the minimum energy structure. This has been defined as “binding funnel analysis” and evaluated based on Spearman's rank correlation between the 10 bins of different RMSD windows and average scores for these bins [40]. We used the docking poses provided in this CASF-2016 dataset directly in assessment of the performance of ANI-2x. We have calculated the binding affinities by ligand interacting with only residues in docking region (grid box/sphere) by saturating the discontinuing atoms with hydrogens using Pdbfixer [45]. Since the ANI-2x has been trained for molecules containing only H, C, O, N, S, F, and Cl atoms, the total the dataset reduced to 254 from the original set of 285 proteins. In addition, we also tested the ANI's performance on SARS-CoV-2 main protease inhibitors by applying our own docking protocol using GOLD software.

GOLD docking protocol

The inhibitors of SARS-CoV-2 main protease that are reported in Protein Data Bank (PDB) were retrieved (complex PDB IDs: 7N44, 7L10, 7L11, 7L12, 7L14, 7M8M, 7M8P, 7M8O, 7M8N, 7M8X, 7M8Y, 7M90, 7M8Z, 7M91) and docked to the protein structure with PDB ID: 7L14. Docking was performed using the GOLD Suite v.5.3 software by Cambridge Structure Database (CSD) [46]. For GOLD, the ChemPLP scoring function was used since it is known to give better results in prediction of binding mode than other scoring functions implemented in the software [26]. The genetic algorithm with a minimum of 1,000,000 and maximum 1,250,000 iterations was used. A grid sphere of 10 Å radius was defined. The center of the sphere was defined by the reference ligand in the structure of 7L14. All rotatable bonds in protein were frozen while the ligand was defined as flexible. The bond orders, hydrogens, atom types, and partial charges were produced by the Hermes software's edit utilities. GOLD software was constructed to produce 25 poses for each ligand without early termination.

ANI-2x as rescoring

ANI scoring relies on the interaction energy between the receptor and the ligand similar to force field based scoring functions. The interaction energy is calculated by ANI-ML potentials rather than non-covalent coulombic and van der Waals interactions of conventional force field based scoring. The details of the calculation can be found on <https://github.com/otayfuroglu/deepQM> and the related ANI_LIE work by Akkus et al. [41], as will be discussed in the next section. The main difference is to use docking pose instead of calculating frames from an MD simulation trajectory. Apart from the input pdb file with protein and ligand complex, the user is required to give an index file containing the index groups to be calculated. The index file is the one produced by Gromacs (i.e., listing comma separated residue numbers followed by the group name in the square bracket). The ligand and the residues in the docking region (or entire protein) are defined as separate groups in the index file. Thus, ANI scores can be readily calculated as the difference in energy between the index groups corresponding to complex and free components (protein and ligand) so that $\Delta E_{ANI} = E_{complex} - (E_{protein} + E_{ligand})$. The interaction energy ΔE_{ANI} is then translated to free energy using a scaling factor, as discussed in the next section. Using this approach, each pose reported on CASF-2016 dataset or generated by our docking with GOLD was recalculated by ANI scoring function.

MD simulations

A similar protocol to our previous works has been used in MD simulations [41, 42, 47–53]. Briefly, ligands were first optimized at B3LYP/6-31++G(d,p), and Merz–Kollman (MK) electrostatic potential (ESP) charges were computed at HF/6-31G* level using Gaussian 16 software. GAFF2 parameters and restricted electrostatic potential charges (RESP) were generated using the antechamber utility in AmberTools 22. The amber99sb-ildn force field was used for the protein's topology. Solvation of protein–ligand complexes employed the TIP3P water model in a dodecahedron box with 10 Å dimensions for each axis. System neutralization occurred at a salt concentration of 0.15 M Na⁺ and Cl[−] ions. Energy minimization, utilizing the steepest descent algorithm with a Verlet cutoff scheme, was performed to a maximum force of 100 kJ mol^{−1} nm^{−1}. Electrostatic interactions were computed using the particle-mesh Ewald (PME) method, while bonds involving hydrogen atoms were constrained with a harmonic potential. Prior to the final simulations, a 5 ns NVT-MD equilibration at 310 K was conducted using a Langevin thermostat followed by 200 ps and 1 ns NPT-MD equilibrations at 1 atm using Berendsen and Parrinello-Rahman barostats, respectively.

Free energy calculations

The single-trajectory approach MMGBSA calculations were calculated using the gmx_MMPBSA [54]. The default parameters with the internal dielectric constant, $\epsilon_{int} = 2$ for polar solvation terms. The SASA-only model was applied with parameters of $\gamma = 0.0072$ kcal/mol·Å² and $b = 0$ for nonpolar terms.

In addition to MM-PBSA calculations, we also used ANI_LIE, a recent end-state binding free energy method introduced by Akkus et al. [41]. The details of the method has been discussed elsewhere [41]. Briefly, the method uses the linear interaction energy (LIE) approach utilizing the potential energies predicted by ANI-ML potentials as a replacement to molecular mechanics (MM) energy terms in LIE formalism. The free energy is calculated by:

$$\Delta G_{bind} = \beta \langle \Delta E_{ANI}^{L-P} \rangle_{PLS} + \gamma$$

where $\langle \Delta E_{ANI}^{L-P} \rangle_{PLS}$ is the average interaction energy between the protein and ligand produced by extracting the MD frames for energy groups of protein (P), ligand (L) and protein–ligand (PL).

For ANI_LIE calculations were performed using the simplest form neglecting the D3 term and solvation effects (i.e., Eq. 4 in Ref. [41]) with default β and γ empirical parameters of 0.127 and -5.11 , respectively. As reported by Akkus et al. [41], these parameters were optimized SARS-CoV-2 M^{pro} and have been successfully used in other studies [42, 47, 48, 55, 56].

Results and discussion

Benchmarking with other docking software on CASF-2016 dataset

In the first part, we tested the scoring produced by ANI predicted interaction energies on a standard dataset containing diverse protein–ligand complexes. Using ANI, we have rescored all the poses belonging to 254 protein–ligand complexes and compared to the results of the original publication of CASF-2016 dataset, which reports the performance of 34 different software/algorithm for the same poses. Figure 1 shows the success rate of predicting the poses with less than 2 Å in the top ranked solutions. The ANI based scoring has been applied when only residues in docking grid are considered as the host in the host–guest interaction, outperforming most of the other 30 scoring functions.

In addition, we have also compared the success rate of finding the exact crystal structure in the first ranked solutions (Fig. 2). ANI can still outperform most of the methods

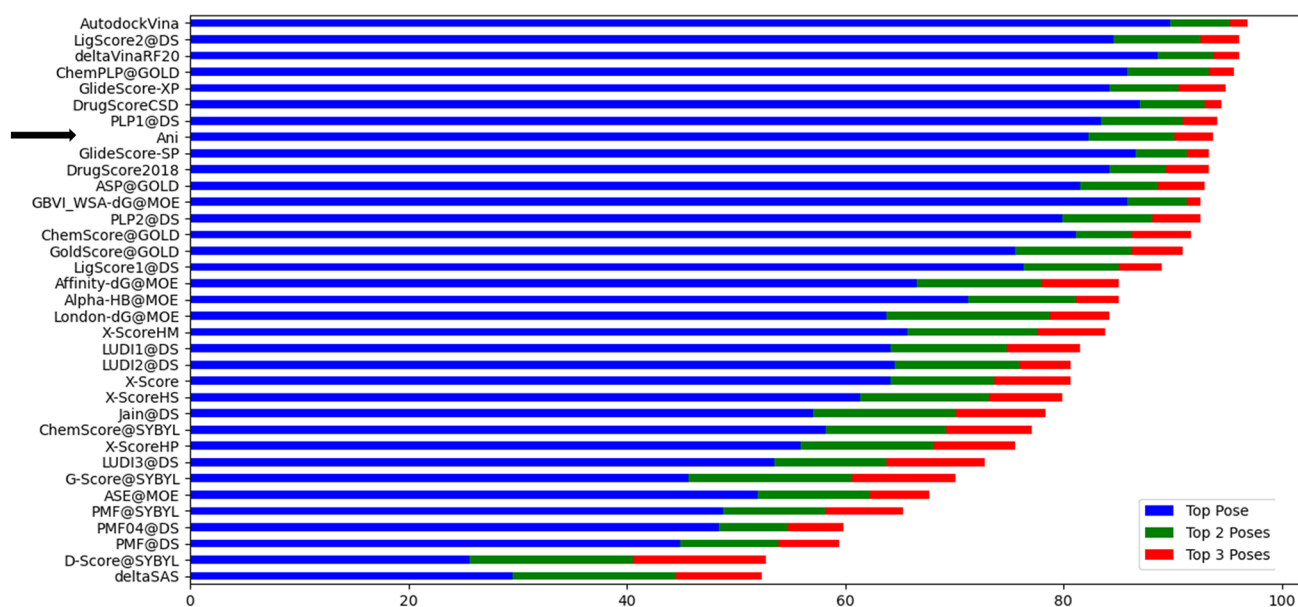


Fig. 1 The success of different scoring functions on predicting the first poses as the closest structure to the crystal with RMSD less than 2.0 Å. ANI is ranked 8th among 34 functions

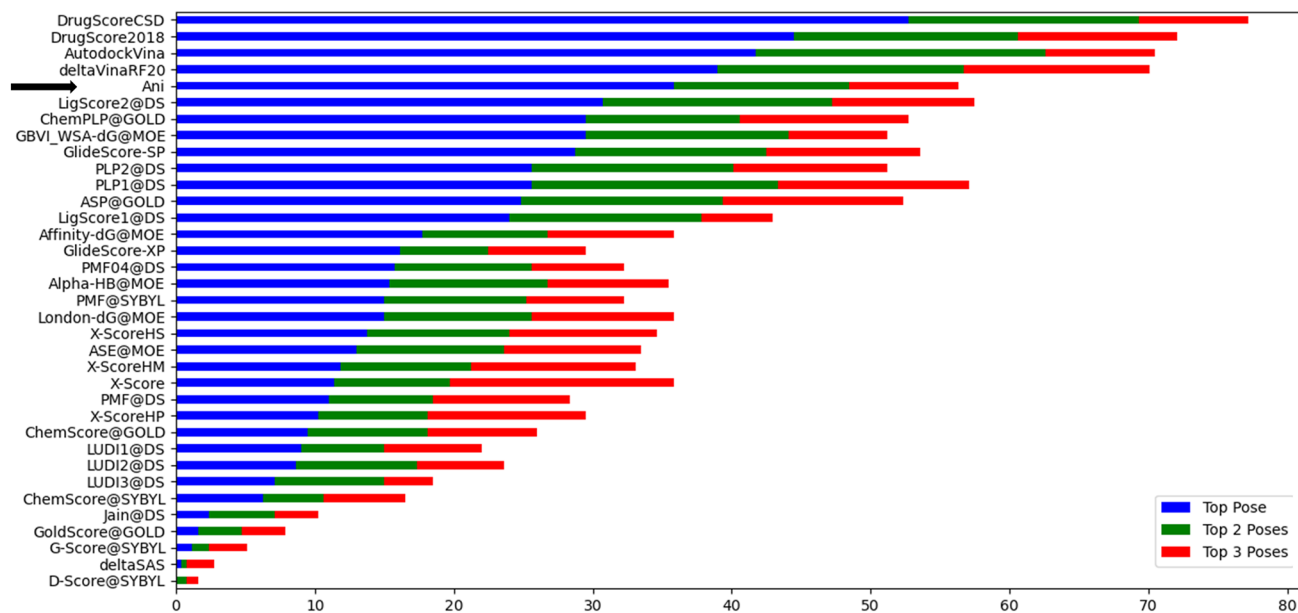


Fig. 2 The success of different scoring functions on predicting the first poses as the exact crystal structure. ANI is ranked 5th among 34 functions

finding the true crystal pose among the given decoy poses. It is clear that when a structure is its crystal orientation, ANI can distinguish it much better than most of the methods.

In addition to docking power, Yang et al. also defined as “binding funnel analysis” [40], which refers to the docking efficiency for a scoring function. The idea is that the poses that are closer to the crystal structure and those that are far

from the crystal structure create a potential energy surface that will look like a funnel. The poses that are closer to the crystal structure will be in the lower energy region of this funnel while the poses that are far from x-ray structure will lie on the higher energy regions. We have performed the same analysis for ANI as well as borrowing the rest of the methods from literature [40] for our set of 254 proteins

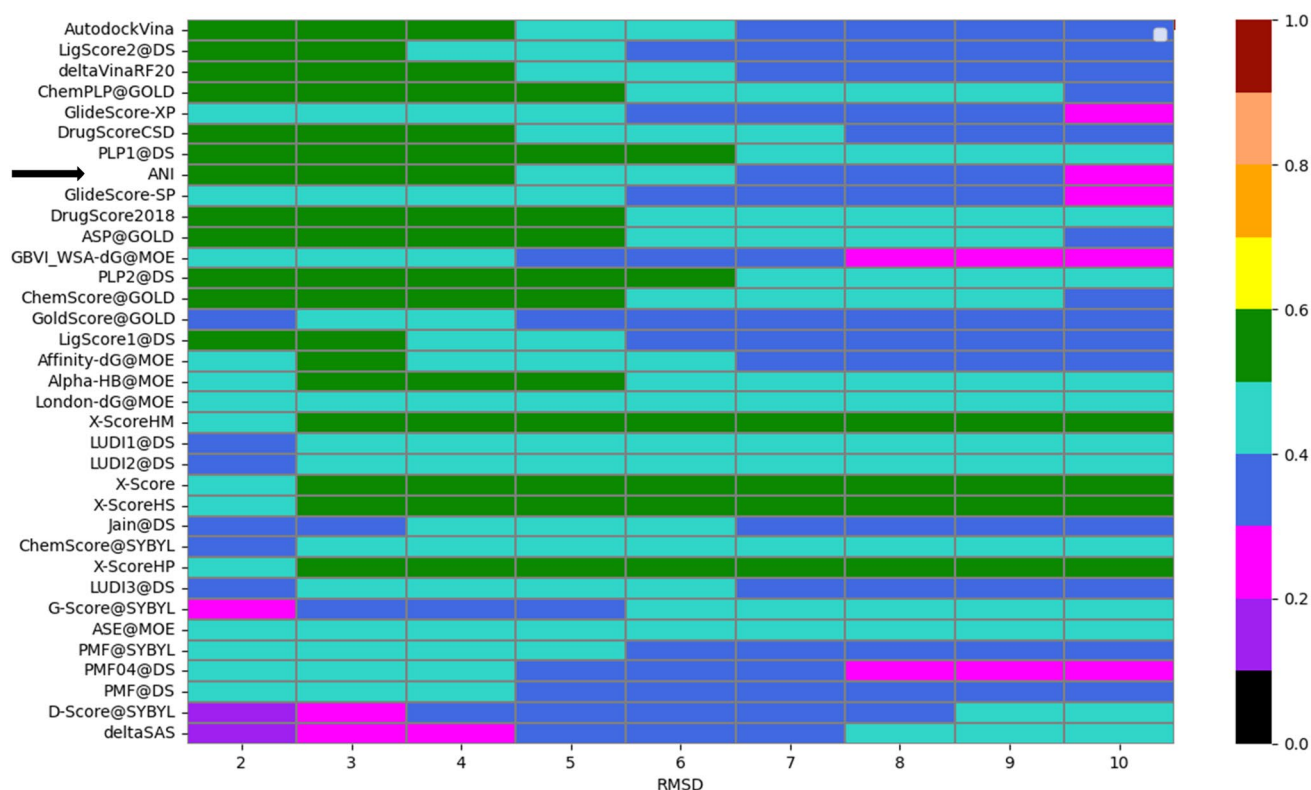


Fig. 3 The Spearman correlation coefficients between RMSD and scores for each scoring function being tested. The x-axis indicates the RMSD bins (≤ 2 , ≤ 3 , etc.) Heat map ranging between 0 and 1 indicates low/high correlations

each with 5 ligands and 100 poses. We have created bins of RMSD windows with 0–2, 0–4, .. 0–10 Å and grouped poses according to these bins. For each RMSD window, the Spearman's correlation coefficient was calculated Fig. 3. The analysis show that the docking efficiency of ANI based scores are one of the most accurate methods, giving high Spearman's correlation coefficients in the lower RMSD bins.

Assessment on SARS-CoV-2 with GOLD

In the second part, we assessed the performance of ANI scorings on a specific protein family of SARS-CoV-2 Mpro rather than diverse set of proteins since our focused study is to screen the FDA drugs against this protein. A representative structure of SARS-CoV-2 main protease, used in the docking was shown in Fig. 4.

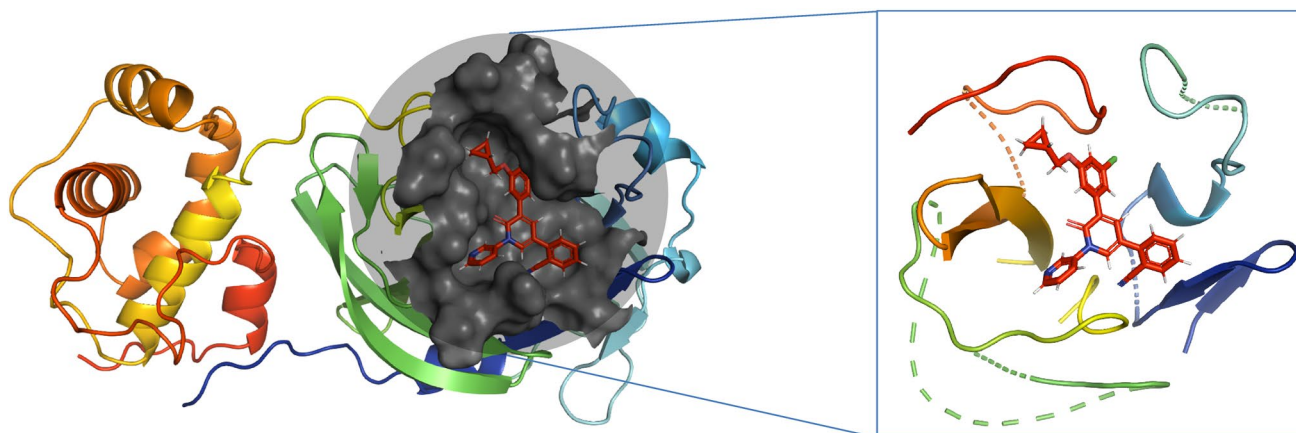


Fig. 4 The structure of the SARS-CoV-2 main protease (PDB ID:7L14) used in GOLD docking. The nearby residues around ligand by 10 Å are highlighted. ANI scores were calculated using $\Delta G_{ANI} = E_{complex} - (E_{protein} + E_{ligand})$

In order to assess the ANI's performance using our own docking protocol in GOLD for SARS-CoV-2 main protease-inhibitor complexes, self-docking is a necessity. Nevertheless, some of the complexes have missing atoms/residues. To overcome this problem, docking all the ligands into a representative receptor structure that is fully resolved (i.e. 7L14) can be a better approach. This would also eliminate possible errors due to using different receptors in comparison of the ligand affinities. However, the complexes available in the PDB show structural differences and residues in the active site show side chain movements. Therefore, selecting only one receptor as a representation for all crystals has also complications to find correlation between docking scores and native binding pose. In order to avoid these error sources, we have selected the most common conformation as a representative structure (i.e. 7L14) for docking all the inhibitors. We have specifically selected 13 other protein-inhibitor complexes available in PDB that have similar scaffold in the binding region. The list of residues in the grid sphere and their RMSD values to the reference structure 7L14 is given in Table S1. In all of these crystal structures, the positions of the residues in the active site is protected. This allows to compare the docking of the native ligands of these proteins to a reference protein (7L14).

The top ranked solutions in GOLD and ANI mostly agree and are quite successful in predicting crystal structures. For half of the docked ligands, the top solutions for GOLD and ANI is the same. In addition, the top 3 ranked solutions for GOLD for each of the ligands were also within the first three solutions of ANI.

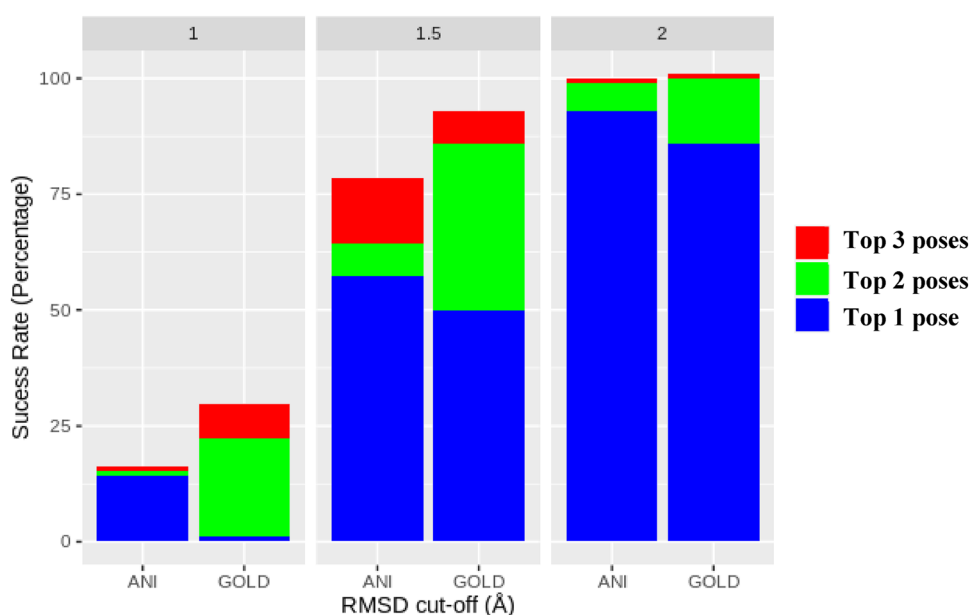
The success rate of finding the lowest RMSD structure in the top ranked solution was 35.7% for GOLD and 28.6%

for ANI. In addition, the average RMSD values of the first poses in GOLD and ANI were 1.48 Å and 1.43 Å, respectively.

When we analyzed the top 1 pose, we observed that ANI can find 14.3% of the poses with $\text{RMSD} \leq 1.0$ Å while GOLD does not find any solution within this cutoff (Fig. 5). For $\text{RMSD} \leq 1.5$ Å, these values are 57.1% and 50.0% ≤ 1.5 Å, respectively; and 92.9% and 85.7%, respectively for $\text{RMSD} \leq 2.0$ Å. ANI's success in the top 1 pose is almost 7% better than GOLD's success. In general, the top 1 score is considered in the high throughput virtual screening (HTVS) studies aiming to find the putative binders for more advanced computational analysis such as molecular dynamics (MD). Given the fact that the ANI scores are very effective in finding the correct binding conformation at the first pose, it demonstrates the capacity of the ANI based rescoring in virtual screening studies.

The trend between the docking scores of the best poses obtained with the GOLD software and the experimental binding energies shows that the Pearson correlation value is $R^2 = 0.47$ (Fig. 6). GOLD scores are arbitrary and the highest scoring values correspond to the highest binding energy (negatively, the lowest binding energy), so the correlation should be inverse. It appears that the GOLD scores are actually in good agreement with the experimental values for the 19 compounds studied. In fact, when the D3F ligand is considered as an outlier, the success of GOLD scores reaches $R^2 = 0.72$. This ligand contains multiple nitro (-NO₂) groups and each N=O bond is assumed double bonds (causing N atom to make 5 bonds) in GOLD due to the difficulty to define the partial atomic charges for these groups as discussed in the GOLD manual [46]. This might explain too low docking scores for this ligand. The ANI interaction

Fig. 5 GOLD-PLP and ANI scoring on the poses generated by Gold docking on 14 selected inhibitors of SARS-CoV-2 main protease. ANI shows superiority to GOLD in the top one solution



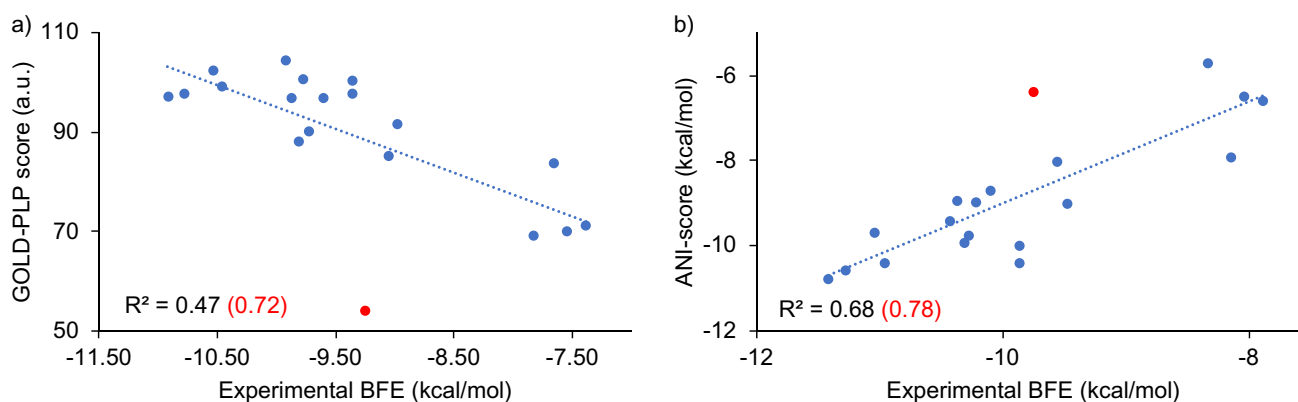


Fig. 6 Docking scores generated by **a** GOLD-PLP and **b** ANI with respect to experimental BFEs. ANI scores were produced by scaling with 0.127 as outlined in [41]

energy between the amino acids in the binding site and the ligand showed a better trend with the experimental binding energies. Pearson correlation coefficient was found to be $R^2 = 0.68$, outperforming the correlation of GOLD scores. Similarly, when the D3F ligand (red dots in the plots) is considered as an outlier, this coefficient increases up to 0.78.

Instead of assessing the performance of ANI and GOLD scores individually, we have also compared when both methods are combined in consensus scoring. A much higher trend with $R^2 = 0.84$ was observed to the experimental values when the top three poses commonly recommended for both GOLD and ANI are considered, clearly supporting the idea of combining these two different methods (GOLD + ANI) as consensus scoring.

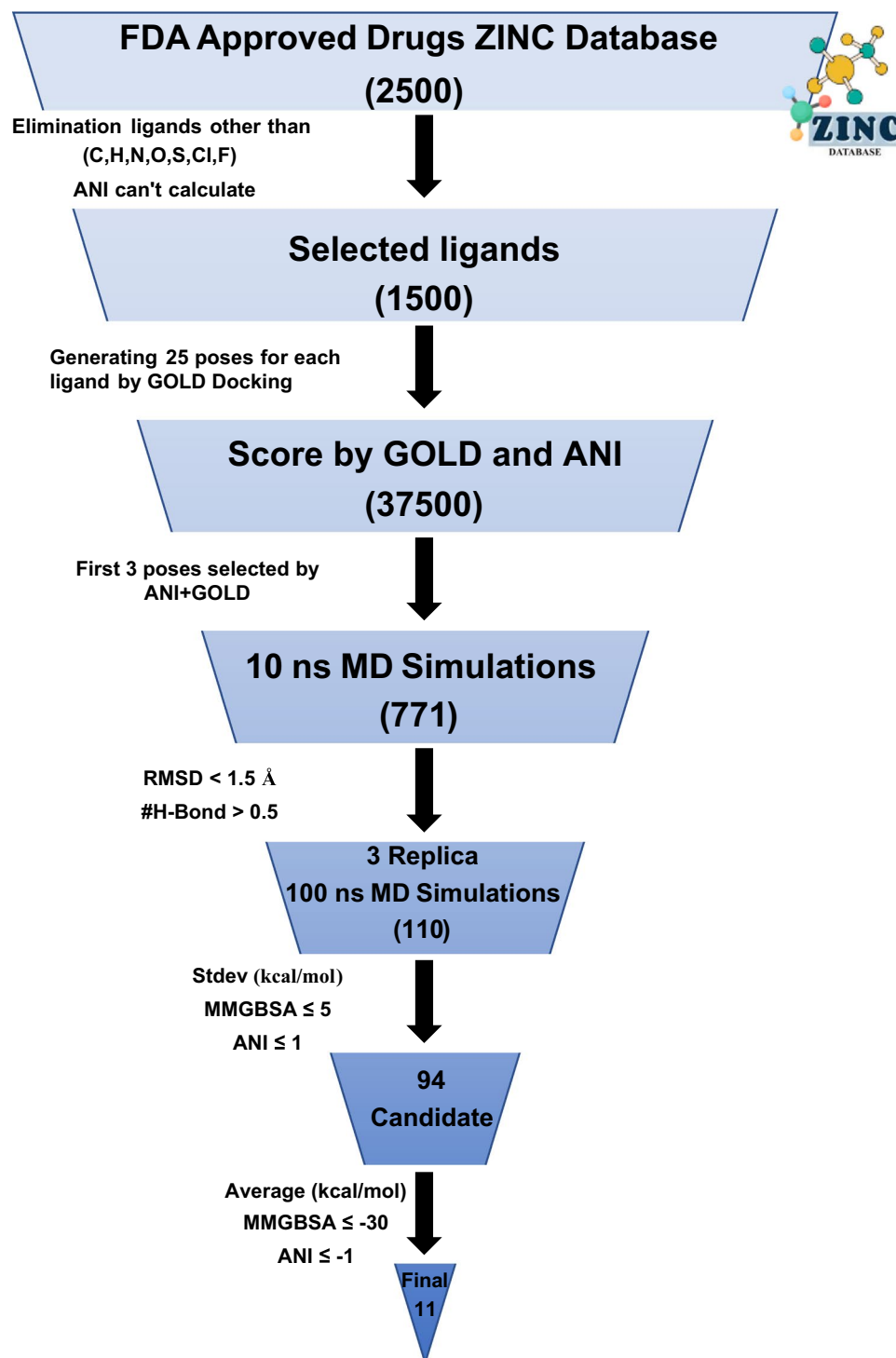
Screening FDA drugs for SARS-CoV-2 Mpro

Our unique screening protocol is given in Fig. 7. MD simulations were initiated using the top three poses obtained from docking. The docking of all FDA-approved drug (2500) molecules in the Zinc database was performed using GOLD. 25 poses for each drug molecule produced by GOLD and rescored with ANI for an additional filtering tool. Since ANI can only perform calculations for C, H, O, N, S, F, and Cl atoms, the number of drugs that are screened decreased to 1460. Our consensus scoring based on GOLD and ANI is as follows: if the top scored three poses generated by GOLD are also ranked in the top three scores with ANI, then these poses are used to MD simulation in the later stage. Otherwise, this complex is discarded. As a result of this filtering process, only 771 out of the best 4380 GOLD docking poses for 1460 drug molecules were also among the top 3 poses according to ANI. All 771 poses belonging to 669 different drug molecules were subjected to MD simulations for 10 ns in the next stage (a total of 7.71 μ s of simulations).

After the MD simulations, the resulting conformations were evaluated from several aspects to further filter possible SARS-CoV-2 main protease inhibitor candidates. Firstly, it was monitored how far the ligand moved away from the binding site during the MD simulation. If the ligand stably binds to the binding site, it will not move far away from the initial structure. We measured this by fitting the MD trajectory on proteins and calculating the RMSD values of the ligands. The cut-off for the RMSD for this stability test was 3.0 Å. Based on this evaluation, 467 drug molecules remained stable in the binding site. The second criterion for filtering out the ligands is protection of the number of H-bonds. By monitoring the number of H-bonds between the ligands and protein, we eliminated those structures that lose more than 0.5 H-bonds throughout the simulation from our hit list. Figure 8 shows the examples for these criteria. Thus, a total of 346 molecules were identified as potential candidates for binding to the protein. As a final criterion in the classical MD simulations, the stability of the protein–ligand complex was assessed by means of RMSD of back bone atoms of the protein with a cutoff of 0.15 Å (Figure S1). This eliminated further drug poses, leaving over 110 complexes.

In the next step, we performed end-state free energy calculations on remaining 110 compounds each followed by three replica 10 ns long MD simulations (a total of 2.2 μ s additional MD simulations). For these selected 110 compounds, free energies were calculated using MMGBSA and ANI_LIE methods (Table S2). Ideally, if a drug-pose interacts with M^{pro} strongly, the free energy values would be much lower. However, if any of the BFE values of the replica simulations is not in the same order as with the other two simulations, then the value predicted is not reliable and should not be trusted. Therefore, we have applied another criterion of standard deviation among the free energies calculated by the replica simulations (5 kcal/mol MMGBSA and 1 kcal/mol ANI_LIE).

Fig. 7 Our unique screening protocol involving consensus scoring of ANI and GOLD, classical MD simulations and consensus BFEs by end-state methods



In the final step, ANI_LIE and MM-GBSA were used in combination in a way that drug molecules with the lowest binding energies relative to other drugs were suggested by both methods were selected. When the average BFE values of three replica MD simulations for a drug molecule were below -8 kcal/mol by ANI_LIE and -30 kcal/mol by MM-GBSA, the drug molecule was assumed plausible

inhibitor for the SARS-CoV-2 M^{pro}. The rest of the candidates were eliminated with the assumption that they would not be sufficient for inhibition. Thus, a total of 11 drug molecules remained for which significant binding energy was suggested by 2 different free energy methods. These drug molecules are listed in Table 1.

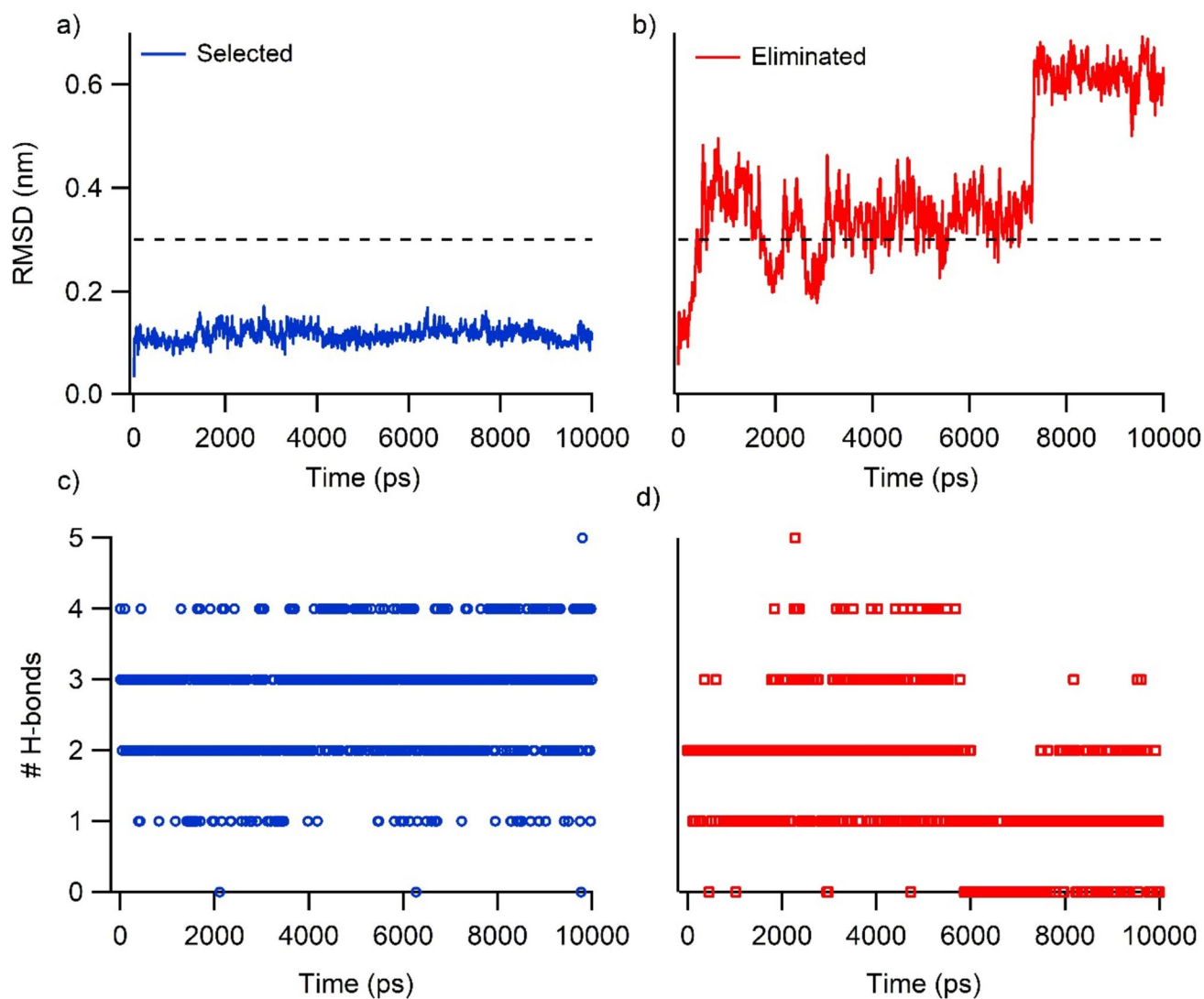


Fig. 8 Application of elimination (red)/selection (blue) criteria to MD trajectories. $\Delta(\text{RMSD}) \leq 3.0 \text{ \AA}$ for the ligand, and $\Delta(\#\text{H-bonds}) \geq -0.5$ between ligand and protein

In all three methods, structures with the lowest binding energies were selected as potential drug molecules by consensus. Among the top 5 drug molecules suggested by the common three methods with the highest binding energies, 3 had already been proposed as potential drug molecules in previous theoretical and experimental studies. This confirms the effectiveness of the screening method used here.

Drug molecules listed here are the most plausible SARS-CoV-2 Mpro inhibitors and it is worth performing further experimental analysis to reveal their potency. In our previous studies, we have shown that ANI_LIE values are much more accurate in predicting absolute binding free energies [41]. Therefore, Cabazitaxel, Rivaroxaban and Dapagliflozin might have sub-nanomolar inhibition concentrations for Mpro due to BFEs below $-10 \text{ kcal mol}^{-1}$.

Conclusions

In this study, we have shown the capability of using ANI potentials as a rescoring function in molecular docking. Our benchmarking studies showed that this method can outperform most of the conventional scoring functions. In particular, the prediction of the top 1 solution showed one of the best performances in prediction of the true binding mode. The method can be adopted in any docking software to screen the drug like molecules as potent inhibitors to proteins. As a case study, we introduce our unique screening protocol which incorporates consensus scoring of ANI and GOLD, classical MD simulations and consensus BFEs by end-state methods.

Table 1 The final drug list we recommended in the FDA drug repurposing study to target SARS-CoV-2 Mpro

Name	GOLD score (a.u.)	ANI score (kcal/mol)	MM-PBSA (kcal/mol)		ANI_LIE (kcal/mol)	
			Avg	SD	Avg	SD
Cabazitaxel	58.3	− 9.0	− 44.3	2.6	− 10.8	0.5
Rivaroxaban ^a	61.0	− 8.4	− 40.5	4.8	− 10.4	0.7
Dapagliflozin	69.6	− 10.9	− 33.5	2.3	− 10.3	0.8
Acalabrutinib*	68.3	− 7.9	− 39.3	4.2	− 9.1	0.8
Zaleplon	61.7	− 8.0	− 34.5	2.2	− 8.8	0.3
Sotorasib	62.0	− 8.3	− 31.5	1.5	− 8.7	0.2
Lopac-T-9025	62.9	− 8.5	− 38.3	2.7	− 8.6	0.3
Prucalopride	54.0	− 8.0	− 31.7	0.9	− 8.6	0.1
Lurbinctedin	57.3	− 7.7	− 30.0	1.8	− 8.3	0.6
Thiethylperazine	65.4	− 9.0	− 34.4	0.3	− 8.2	0.0
N-Desmethyl ulipristal acetate	56.6	− 7.0	− 37.3	1.6	− 8.0	0.2

^aCompounds that have been in clinical trials or suggested in literature

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10822-024-00554-4>.

Acknowledgements The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). This work supported by Scientific and Technological Research Council of Turkey-TÜBİTAK (project number: 120Z732)

Author contributions A. Kocak conceptualized the study and designed the methodology. I. N. Zengin performed MD simulations and free energy calculations. M. S. Koca performed the docking studies. O. Tayfuroglu prepared the scripts. A. Kocak and M. Yildiz evaluated and discussed findings. A. Kocak and M. Yildiz wrote the manuscript. A. Kocak supervised the overall study. All authors approved the manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability Data is provided within the manuscript or supplementary information files

Declarations

Conflict of interest The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants, patents received or pending, or royalties.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Kimber TB, Chen Y, Volkamer A (2021) Deep learning in virtual screening: Recent applications and developments. *Int J Mol Sci*. <https://doi.org/10.3390/ijms22094435>
- Berdigaliyev N, Aljofan M (2020) An overview of drug discovery and development. *Future Med Chem* 12:939–947. <https://doi.org/10.4155/fmc-2019-0307>
- Cavasotto CN, Di Filippo JI (2021) Artificial intelligence in the early stages of drug discovery. *Arch Biochem Biophys* 698:108730. <https://doi.org/10.1016/j.abb.2020.108730>
- Li J, Fu A, Zhang L (2019) An overview of scoring functions used for protein-ligand interactions in molecular docking. *Interdiscip Sci* 11:320–328. <https://doi.org/10.1007/s12539-019-00327-w>
- Shen C, Ding J, Wang Z, Cao D, Ding X, Hou T (2020) From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *WIREs Comput Mol Sci* 10:1429. <https://doi.org/10.1002/wcms.1429>
- Ayodele TO (2010) Types of machine learning algorithms. *New Adv Mach Learn* 3(19–48):5–1
- Khamis MA, Gomaa W, Ahmed WF (2015) Machine learning in computational docking. *Artif Intell Med* 63:135–152. <https://doi.org/10.1016/j.artmed.2015.02.002>
- Ma D-L, Chan DS-H, Leung C-H (2013) Drug repositioning by structure-based virtual screening. *Chem Soc Rev* 42:2130–2141. <https://doi.org/10.1039/C2CS35357A>
- Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH (2012) Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J* 14:133–141. <https://doi.org/10.1208/s12248-012-9322-0>
- Zhang L, Ai H-X, Li S-M, Qi M-Y, Zhao J, Zhao Q, Liu H-S (2017) Virtual screening approach to identifying influenza virus neuraminidase inhibitors using molecular docking combined with machine-learning-based scoring function. *Oncotarget* 8:47
- Zhang L, Qiao M, Gao H, Hu B, Tan H, Zhou X, Li CM (2016) Investigation of mechanism of bone regeneration in a porous

- biodegradable calcium phosphate (CaP) scaffold by a combination of a multi-scale agent-based model and experimental optimization/validation. *Nanoscale* 8:14877–14887. <https://doi.org/10.1039/C6NR01637E>
12. Zhang L, Zhang S (2017) Using game theory to investigate the epigenetic control mechanisms of embryo development: Comment on: “Epigenetic game theory: How to compute the epigenetic control of maternal-to-zygotic transition” by Qian Wang. *Phys Life Rev* 20:140–142. <https://doi.org/10.1016/j.plrev.2017.01.007>
 13. Smith JS, Isayev O, Roitberg AE (2017) Data Descriptor: ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Sci Data* 4:1–8. <https://doi.org/10.1038/sdata.2017.193>
 14. Smith JS, Roitberg AE, Isayev O (2018) Transforming computational drug discovery with machine learning and AI. *ACS Med Chem Lett* 9:1065–1069. <https://doi.org/10.1021/acsmchemlett.8b00437>
 15. Smith JS, Nebgen B, Lubbers N, Isayev O, Roitberg AE (2018) Less is more: Sampling chemical space with active learning. *J Chem Phys*. <https://doi.org/10.1063/1.5023802>
 16. Smith JS, Nebgen BT, Zubatyuk R, Lubbers N, Devereux C, Barros K, Tretiak S, Isayev S, Roitberg A (2018) Outsmarting quantum chemistry through transfer learning. <https://doi.org/10.26434/chemrxiv.6744440.v1>
 17. Kaul D (2020) An overview of coronaviruses including the SARS-2 coronavirus: molecular biology, epidemiology and clinical implications. *Curr Med Res Pract* 10:54–64. <https://doi.org/10.1016/j.cmrp.2020.04.001>
 18. World Health Organization (n.d.) Weekly epidemiological update on COVID-19.
 19. Torre-Fuentes L, Matías-Guiu J, Hernández-Lorenzo L, Montero-Escribano P, Pytel V, Porta-Etessam J, Gómez-Pinedo U, Matías-Guiu JA (2021) ACE2, TMPRSS2, and Furin variants and SARS-CoV-2 infection in Madrid, Spain. *J Med Virol* 93:863–869. <https://doi.org/10.1002/jmv.26319>
 20. Vankadari N (2020) Structure of Furin protease binding to SARS-CoV-2 spike glycoprotein and implications for potential targets and virulence. *J Phys Chem Lett* 11:6655–6663. <https://doi.org/10.1021/acs.jpcclett.0c01698>
 21. Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, Khattri J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, Vogrig R, Ward D, Watson B, Brunham RC, Krajden M, Petric M, Skowronski DM, Upton C, Roper RL (1979) The Genome sequence of the SARS-associated coronavirus. *Science* 300(2003):1399–1404. <https://doi.org/10.1126/science.1085953>
 22. Owen DR, Allerton CMN, Anderson AS, Aschenbrenner L, Avery M, Berritt S, Boras B, Cardin RD, Carlo A, Coffman KJ, Dantonio A, Di L, Eng H, Ferre R, Gajiwala KS, Gibson SA, Greasley SE, Hurst BL, Kadar EP, Kalgutkar AS, Lee JC, Lee J, Liu W, Mason SW, Noell S, Novak JJ, Obach RS, Ogilvie K, Patel NC, Pettersson M, Rai DK, Reese MR, Sammons MF, Sathish JG, Singh RSP, Steppan CM, Stewart AE, Tuttle JB, Updyke L, Verhoest PR, Wei L, Yang Q, Zhu Y (1979) An oral SARS-CoV-2 M^{pro} inhibitor clinical candidate for the treatment of COVID-19. *Science* 374(2021):1586–1593. <https://doi.org/10.1126/science.abl4784>
 23. Painter GR, Bowen RA, Bluemling GR, DeBergh J, Edpuganti V, Gruddanti PR, Guthrie DB, Hager M, Kuiper DL, Lockwood MA, Mitchell DG, Natchus MG, Sticher ZM, Kolykhalov AA (2019) The prophylactic and therapeutic activity of a broadly active ribonucleoside analog in a murine model of intranasal Venezuelan equine encephalitis virus infection. *Antiviral Res* 171:104597. <https://doi.org/10.1016/j.antiviral.2019.104597>
 24. Boucau J, Uddin R, Marino C, Regan J, Flynn JP, Choudhary MC, Chen G, Stuckwisch AM, Mathews J, Liew MY, Singh A, Reynolds Z, Iyer SL, Chamberlin GC, Vyas TD, Vyas JM, Turbett SE, Li JZ, Lemieux JE, Barczak AK, Siedner MJ (2019) Characterization of virologic rebound following nirmatrelvir-ritonavir treatment for coronavirus disease 2019 COVID-19. *Clin Infect Dis* 76(2023):e526–e529. <https://doi.org/10.1093/cid/ciac512>
 25. Hu Y, Lewandowski EM, Tan H, Zhang X, Morgan RT, Zhang X, Jacobs LMC, Butler SG, Gongora MV, Choy J, Deng X, Chen Y, Wang J (2022) Naturally occurring mutations of SARS-CoV-main protease confer drug resistance to nirmatrelvir. *BioRxiv*. <https://doi.org/10.1101/2022.06.28.497978>
 26. Mitsuya H, Maeda K, Das D, Ghosh AK (2008) Development of protease inhibitors and the fight with drug-resistant HIV-1 variants. *Adv Pharmacol* 56:169–197. [https://doi.org/10.1016/S1054-3589\(07\)56006-0](https://doi.org/10.1016/S1054-3589(07)56006-0)
 27. Hilgenfeld R (2014) From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J* 281:4085–4096. <https://doi.org/10.1111/febs.12936>
 28. Macchiagodena M, Pagliai M, Procacci P (2020) Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. *Chem Phys Lett* 750:137489. <https://doi.org/10.1016/J.CPLETT.2020.137489>
 29. Ma C, Xia Z, Sacco MD, Hu Y, Townsend JA, Meng X, Choza J, Tan H, Jang J, Gongora MV, Zhang X, Zhang F, Xiang Y, Marty MT, Chen Y, Wang J (2021) Discovery of Di- and trihaloacetamides as covalent SARS-CoV-2 main protease inhibitors with high target specificity. *J Am Chem Soc* 143:20697–20709. <https://doi.org/10.1021/jacs.1c08060>
 30. Kitamura N, Sacco MD, Ma C, Hu Y, Townsend JA, Meng X, Zhang F, Zhang X, Ba M, Szeto T, Kukuljac A, Marty MT, Schultz D, Cherry S, Xiang Y, Chen Y, Wang J (2022) Expedited approach toward the rational design of noncovalent SARS-CoV-2 main protease inhibitors. *J Med Chem* 65:2848–2865. <https://doi.org/10.1021/acs.jmedchem.1c00509>
 31. Ma C, Sacco MD, Hurst B, Townsend JA, Hu Y, Szeto T, Zhang X, Tarbet B, Marty MT, Chen Y, Wang J (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res* 30:678–692. <https://doi.org/10.1038/s41422-020-0356-z>
 32. Qiao J, Li Y-S, Zeng R, Liu F-L, Luo R-H, Huang C, Wang Y-F, Zhang J, Quan B, Shen C, Mao X, Liu X, Sun W, Yang W, Ni X, Wang K, Xu L, Duan Z-L, Zou Q-C, Zhang H-L, Qu W, Long Y-H-P, Li M-H, Yang R-C, Liu X, You J, Zhou Y, Yao R, Li W-P, Liu J-M, Chen P, Liu Y, Lin G-F, Yang X, Zou J, Li L, Hu Y, Lu G-W, Li W-M, Wei Y-Q, Zheng Y-T, Lei J, Yang S (1979) SARS-CoV-2 M^{pro} inhibitors with antiviral activity in a transgenic mouse model. *Science* 371(2021):1374–1378. <https://doi.org/10.1126/science.abf1611>
 33. Boras B, Jones RM, Anson BJ, Arenson D, Aschenbrenner L, Bakowski MA, Beutler N, Binder J, Chen E, Eng H, Hammond H, Hammond J, Haupt RE, Hoffman R, Kadar EP, Kania R, Kimoto E, Kirkpatrick MG, Lanyon L, Lendy EK, Lillis JR, Logue J, Luthra SA, Ma C, Mason SW, McGrath ME, Noell S, O’Bach RS, O’Brien MN, O’Connor R, Ogilvie K, Owen D, Pettersson M, Reese MR, Rogers TF, Rosales R, Rossulek MI, Sathish JG, Shirai N, Steppan C, Ticehurst M, Updyke LW, Weston S, Zhu Y, White KM, García-Sastre A, Wang J, Chatterjee AK, Mesecar AD, Friedman MB, Anderson AS, Allerton C (2021) Preclinical characterization of an intravenous coronavirus CL protease inhibitor for the

- potential treatment of COVID19. *Nat Commun* 12:6055. <https://doi.org/10.1038/s41467-021-26239-2>
34. Dai W, Zhang B, Jiang X-M, Su H, Li J, Zhao Y, Xie X, Jin Z, Peng J, Liu F, Li C, Li Y, Bai F, Wang H, Cheng X, Cen X, Hu S, Yang X, Wang J, Liu X, Xiao G, Jiang H, Rao Z, Zhang L-K, Xu Y, Yang H, Liu H (1979) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* 368(2020):1331–1335. <https://doi.org/10.1126/science.abb4489>
35. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R (1979) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* 368(2020):409–412. <https://doi.org/10.1126/science.abb3405>
36. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582:289–293. <https://doi.org/10.1038/s41586-020-2223-y>
37. Quan B-X, Shuai H, Xia A-J, Hou Y, Zeng R, Liu X-L, Lin G-F, Qiao J-X, Li W-P, Wang F-L, Wang K, Zhou R-J, Yuen TT-T, Chen M-X, Yoon C, Wu M, Zhang S-Y, Huang C, Wang Y-F, Yang W, Tian C, Li W-M, Wei Y-Q, Yuen K-Y, Chan JF-W, Lei J, Chu H, Yang S (2022) An orally available Mpro inhibitor is effective against wild-type SARS-CoV-2 and variants including Omicron. *Nat Microbiol* 7:716–725. <https://doi.org/10.1038/s41564-022-01119-7>
38. Subissi L, Imbert I, Ferron F, Collet A, Coutard B, Decroly E, Canard B (2014) SARS-CoV ORF1b-encoded nonstructural proteins 12–16: Replicative enzymes as antiviral targets. *Antiviral Res* 101:122–130. <https://doi.org/10.1016/j.antiviral.2013.11.006>
39. Palacio-Rodríguez K, Lans I, Cavasotto CN, Cossio P (2019) Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci Rep* 9:5142. <https://doi.org/10.1038/s41598-019-41594-3>
40. Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, Wang R (2019) Comparative assessment of scoring functions: the CASF-2016 update. *J Chem Inf Model* 59:895–913. <https://doi.org/10.1021/acs.jcim.8b00545>
41. Akkus E, Tayfuroglu O, Yildiz M, Kocak A (2022) Accurate binding free energy method from end-state MD simulations. *J Chem Inf Model* 62:4095–4106. <https://doi.org/10.1021/acs.jcim.2c00601>
42. Akkus E, Tayfuroglu O, Yildiz M, Kocak A (2023) Revisiting MMPBSA by adoption of MC-based surface area/volume, ANI-ML potentials, and two-valued interior dielectric constant. *J Phys Chem B* 127:4415–4429. <https://doi.org/10.1021/acs.jpcc.3c00834>
43. Li Y, Liu Z, Li J, Han L, Liu J, Zhao Z, Wang R (2014) Comparative assessment of scoring functions on an updated benchmark: 1 compilation of the test set. *J Chem Inf Model* 54:1700–1716. <https://doi.org/10.1021/ci500080q>
44. Li Y, Han L, Liu Z, Wang R (2014) Comparative assessment of scoring functions on an updated benchmark: 2 evaluation methods and general results. *J Chem Inf Model* 54:1717–1736. <https://doi.org/10.1021/ci500081m>
45. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, Beauchamp KA, Lane TJ, Wang L-P, Shukla D, Tye T, Houston M, Stich T, Klein C, Shirts MR, Pande VS (2013) OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theory Comput* 9:461–469. <https://doi.org/10.1021/ct300857j>
46. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. Edited by F. E. Cohen. *J Mol Biol* 267:727–748. <https://doi.org/10.1006/jmbi.1996.0897>
47. Temel M, Tayfuroglu O, Kocak A (2023) The performance of ANI-ML potentials for ligand-n(H₂O) interaction energies and estimation of hydration free energies from end-point MD simulations. *J Comput Chem* 44:559–569. <https://doi.org/10.1002/jcc.27022>
48. Tasci HS, Akkus E, Yildiz M, Kocak A (2023) Computational analysis of substrate recognition of Sars-Cov-2 Mpro main protease. *Comput Biol Chem* 107:107960. <https://doi.org/10.1016/j.compbiolchem.2023.107960>
49. Kocak A, Erol I, Yildiz M, Can H (2016) Computational insights into the protonation states of catalytic dyad in BACE1–acyl guanidine based inhibitor complex. *J Mol Graph Model* 70:226–235. <https://doi.org/10.1016/j.jmgm.2016.10.013>
50. Kocak A, Yildiz M (2019) Molecular dynamics studies of the norovirus-host cell interaction mediated by h-type 1 antigen. *Trak Univ J Nat Sci ASAP* 20(1):19–26. <https://doi.org/10.23902/trkjin.at.508120>
51. Yildiz M, Kocak A (2019) Molecular dynamics studies of histoblood group antigen blocking human immunoglobulin an antibody and escape mechanism in noroviruses upon mutation. *J Comput Biol* 26:962–974. <https://doi.org/10.1089/cmb.2018.0163>
52. Kocak A, Yildiz M (2017) Docking, molecular dynamics and free energy studies on aspartoacylase mutations involved in Canavan disease. *J Mol Graph Model* 74:44–53. <https://doi.org/10.1016/j.jmgm.2017.03.011>
53. Kocak A (2019) HBGA binding modes and selectivity in noroviruses upon mutation: a docking and molecular dynamics study. *J Mol Model* 25:369. <https://doi.org/10.1007/s00894-019-4261-7>
54. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Account Chem Res* 33:889–897. <https://doi.org/10.1021/ar000033j>
55. Kankinou SG, Yildiz M, Kocak A (2023) Exploring potential Plasmodium kinase inhibitors: a combined docking MD and QSAR studies. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2023.2249111>
56. Uzuner U, Akkus E, Kocak A, Uzuner SÇ (2023) Exploring epigenetic drugs as potential inhibitors of SARS-CoV-2 main protease: a docking and MD simulation study. *J Biomol Struct Dyn* 1:12. <https://doi.org/10.1080/07391102.2023.2236714>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.