Journal of Computer-Aided Molecular Design (2021) 35:963–971
https://doi.org/10.1007/s10822-021-00412-7

Quantum simulations of SARS-CoV-2 main protease M^{pro} enable high-quality scoring of diverse ligands

Yuhang Wang[1] · **Sruthi Murlidaran**[1] · **David A. Pearlman**[1]

Received: 18 February 2021 / Accepted: 12 July 2021 / Published online: 30 July 2021

Abstract

The COVID-19 pandemic has led to unprecedented efforts to identify drugs that can reduce its associated morbidity/mortality rate. Computational chemistry approaches hold the potential for triaging potential candidates far more quickly than their experimental counterparts. These methods have been widely used to search for small molecules that can inhibit critical proteins involved in the SARS-CoV-2 replication cycle. An important target is the SARS-CoV-2 main protease Mpro, an enzyme that cleaves the viral polyproteins into individual proteins required for viral replication and transcription. Unfortunately, standard computational screening methods face difficulties in ranking diverse ligands to a receptor due to disparate ligand scaffolds and varying charge states. Here, we describe full density functional quantum mechanical (DFT) simulations of Mpro in complex with various ligands to obtain absolute ligand binding energies. Our calculations are enabled by a new cloud-native parallel DFT implementation running on computational resources from Amazon Web Services (AWS). The results we obtain are promising: the approach is quite capable of scoring a very diverse set of existing drug compounds for their affinities to M pro and suggest the DFT approach is potentially more broadly applicable to repurpose screening against this target. In addition, each DFT simulation required only ~ 1 h (wall clock time) per ligand. The fast turnaround time raises the practical possibility of a broad application of large-scale quantum mechanics in the drug discovery pipeline at stages where ligand diversity is essential.

Keywords COVID-19 · Quantum mechanics · Discovery · Drug triage · DensityFunctional Theory · SARS-CoV-2 M^{pro}

Introduction

Computational chemistry has made significant progress in the past several decades, addressing bottlenecks in the drug discovery process. The improvement is particularly visible in the ligand triage step during the initial virtual screening phases (e.g., via molecular docking). The increased use of computational chemistry techniques is also seen in binary decision making near the end of a drug discovery project when the molecular scaffold has been established, and one seeks only to compare congeneric compounds (e.g., via free energy calculations). However, there is a significant computational gap in the middle of the discovery process, where more diverse compounds are encountered. There is an urgent need for computational approaches that can rank order dozens, or hundreds, of unrelated compounds, with sufficiently high accuracy [1].

The technical requirements of such an approach are that it should be able to (1) score ligands with diverse scaffolds; (2) deal with variances in formal charge and polarization; (3) be applicable to realistic models of ligand/protein interactions; and (4) perform these calculations sufficiently quickly to be compatible with modern drug discovery—all while retaining good accuracy. Existing, widely-used methods, such as those based on free energy perturbation and classical force fields [2, 3], usually satisfy criteria 3–4 but fail 1–2. In principle, high-level quantum mechanical calculations, at the level of modern density functional theory (DFT), can address both 1–2, but, until recently, could not be performed on large enough systems with sufficient throughput to address points 3–4 [4, 5]. In a recent publication [6], we described an implementation of a new algorithm for quantum calculations ("high-efficiency distributed QM" (hedQM)). This implementation allows quantum determinations at the DFT level to be performed with reasonable throughput (~ 1 h) on

✉ David A. Pearlman
pearlman@qsimulate.com

[1] Quantum Simulation Technologies, Inc, 625 Massachusetts Ave, Floor 2, Cambridge, MA 02139, USA

systems much larger than ever before possible—for example, full proteins—enabled by easily accessible commercial cloud compute resources, such as those offered by Amazon Web Services (AWS).

Here, we apply this method to a data set germane to identifying new drugs that might help battle the COVID-19 virus. The dataset originates from a recent publication [7], and we briefly describe its construction here. A set of more than 2500 drug molecules previously reported for various applications were subjected to a computational screen against M^{pro}, the SARS-CoV-2 main protease [8]. This enzyme cleaves viral polyproteins into individual proteins required for viral replications and transcription, and it is hypothesized that inhibiting this protein would inhibit replication of the COVID-19 virus [9]. From this computational screen, 100 molecules were identified as having the potential to bind to M^{pro} using a combination of molecular docking and absolute binding free energy calculations. Subsequently, this set of ligands was screened experimentally, leading to a set of 16 molecules with measurable binding to the M^{pro} receptor. With experimentally determined binding affinity values for M^{pro}, this set of ligands serves as the validation set for this study. The chemical structures of these drug molecules (except for a covalent ligand, disulfram) are shown in Fig. 1.

Two observations can immediately be made about this set of molecules. First, they are extremely diverse and reflect a highly divergent set of scaffold classes. Second, they also reflect a diversity of charge states across the potential isomers. These challenges are precisely those raised in points 1–2 above. In particular, the diversity of charge states makes it an extremely challenging data set to model by standard force-fields and associated molecular mechanics methods [10]. The multiple scaffold classes render the set difficult or impossible to address with relative difference methods like FEP, which typically require that the ligands being studied be fairly similar to one another [11]. For these reasons, this is a data set that has the potential to demonstrate the value that a high-level quantum approach—capable of determining absolute energies of binding—can bring to such an investigation.

Methods

The set of 16 drug compounds with experimentally measured binding to M^{pro} is taken from a recent study [7], as are the experimental ligand binding free energies. One compound (disulfiram) reported in that publication is omitted because it is believed to be covalently bound [12]. For the remaining 15 compounds, we applied our cloud-native parallel hedQM approach to determine the absolute energy of binding at the DFT level. We used the revPBE functional [13] with the D3(BJ) dispersion correction [14]—as has been demonstrated to perform well for calculating non-covalent interactions [15]. The def2-SVP basis [16] was used within a 9.1 Å sphere around the ligand within the binding site, while a minimal basis (MINAO) [17] was used for atoms outside this sphere. Details of the DFT calculations are provided in the supplemental information.

Since experimental structures of the bound ligand/protein complexes were not available, it was necessary to generate them using structure-based docking. A crystallographic structure of the M^{pro} protein (Mpro- x 3080) was obtained from the Diamond Light Source (UK) synchrotron facility's Fragalysis web application [18]. This crystal structure of M^{pro} was determined as part of the COVID Moonshot project [19]. AutoDock Bias method [20] was then used for docking. The M^{pro} protein structure consists of a domain including the binding site and a second alpha-helical domain located far from the binding site. In solution, the protein forms a homodimer. To optimize computational cost, we truncated the beginning of the unstructured N-terminal region (SER1 to LYS5) and the alpha-helical domain (residue ASP197 to THR304). The new terminal residues (MET6 and THR196) were capped with ACE and NME terminal groups, respectively. The protein structure is shown in Fig. 2. The truncated protein, which retains the active site, contains 2900 atoms. The atoms shown in grey are those truncated for the calculations.

For each ligand, we included several isomers reasonable for physiological pH, and each isomer was processed independently during the docking process. Next, all poses for the same parent molecule were aggregated for the subsequent scoring process. A total of 100 docked complexes were generated for each ligand. To rank the docked poses, we first evaluated the total energies of the docked structures at the molecular mechanics (MM) level. Then we selected the top-50 docked poses and tightly minimized the structures using molecular mechanics (see supplemental information). From the resulting set of 50 MM-minimized docked structures, the ten lowest energy ligand/protein poses were further optimized using the semi-empirical GFN1-xTB method with Generalized Born/solvent accessible surface area (GBSA) implicit solvent [21]. The two lowest-energy GFN1-xTB ligand/protein poses for each ligand were then selected for full DFT calculations with the C-PCM implicit solvent model [22]. The post-docking classical mechanics calculations were carried out using AmberTools20 [23], using the Generalized Born implicit solvent model [24] (igb = 5), the Amber 14 force field [25] for the protein, and the GAFF force field [26] for ligands assigned using Antechamber from AmberTools.

To determine the lowest energy conformation of the unbound ligand, we used a combination of conformers generated from the classical mechanics search method RDKit

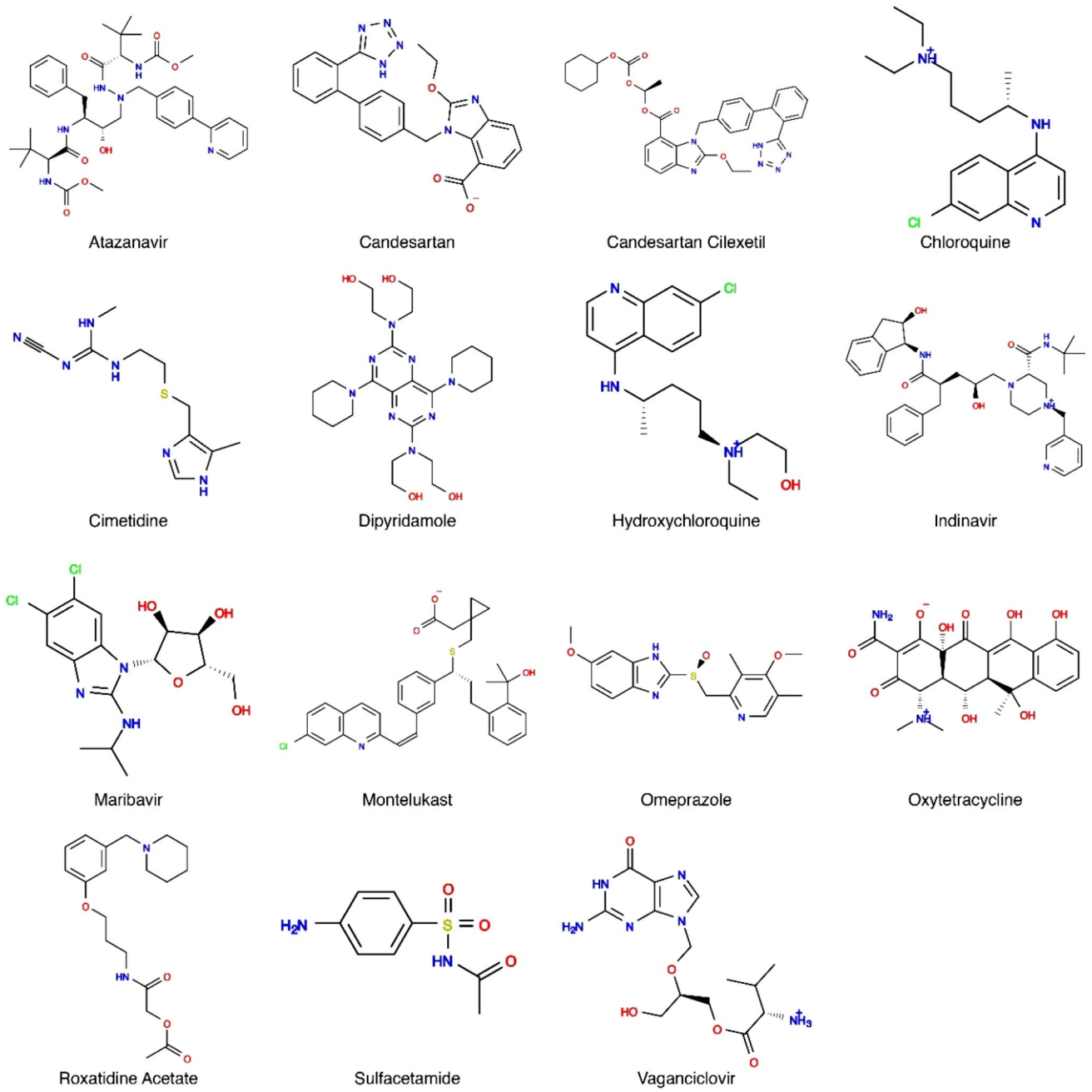

Fig. 1 Drug molecules examined in this study. Each has been experimentally determined [7] to bind with measurable affinity to SARS-CoV-2 M^{pro}. For each ligand, several isomers were considered, leading to a range of charge states

[28] and the semi-empirical conformational search protocol in CREST [29]. From the set of resultant conformers, the energies of the ten lowest energy structures were recalculated using DFT using the C-PCM implicit solvent model [22].

The net energy of binding is determined from the relationship:

$$\Delta E\,(P + L \rightarrow P \cdot L) = E(P \cdot L) - E\left(P_{complex}\right) - E\left(L_{min}\right) \quad (1)$$

where E (P·L) is the energy of the complex, $E(P_{complex})$ is the energy of the protein alone, in the same conformation as the complex, and $E(L_{min})$ is the minimum energy of unbound ligand conformer, determined using the search approach described above. To reflect conformational sampling, we used Boltzmann averaging for the binding energies of the

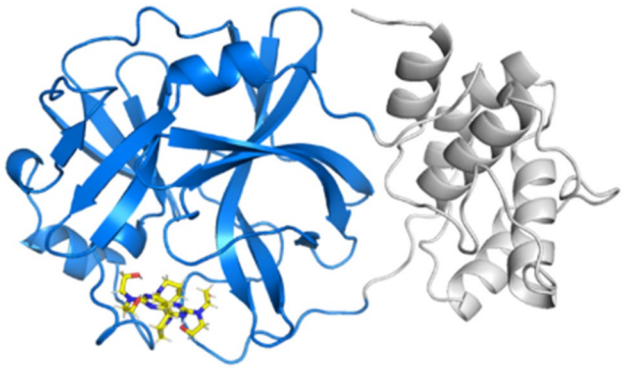

Fig. 2 M^{pro} protein with a ligand (dipyridamole) bound to its active site. The region in gray was excluded from all calculations. The ligand is shown in a licorice representation (image generated using PyMOL [27])

two most favorable docked poses for each ligand isomer. To calculate the ranking score, we performed linear averaging for the Boltzmann-averaged binding energies of all isomers for each ligand (this avoids difficulties in reweighting energies of isomers with different numbers of atoms).

Results

The net binding energies calculated using full DFT are presented in Fig. 3. We use this as a ranking score and plot it against the experimental binding free energies in a correlation plot. For comparison, the same plot is presented for the semi-empirical GFN1-xTB quantum method in Fig. 4, and for MM/GBSA in Figure S1.

As can be seen, the correlation obtained using DFT calculations of the M^{pro} binding domain is significant, with an R^2 value of 0.58 and a Predictive Index [30] (a weighted

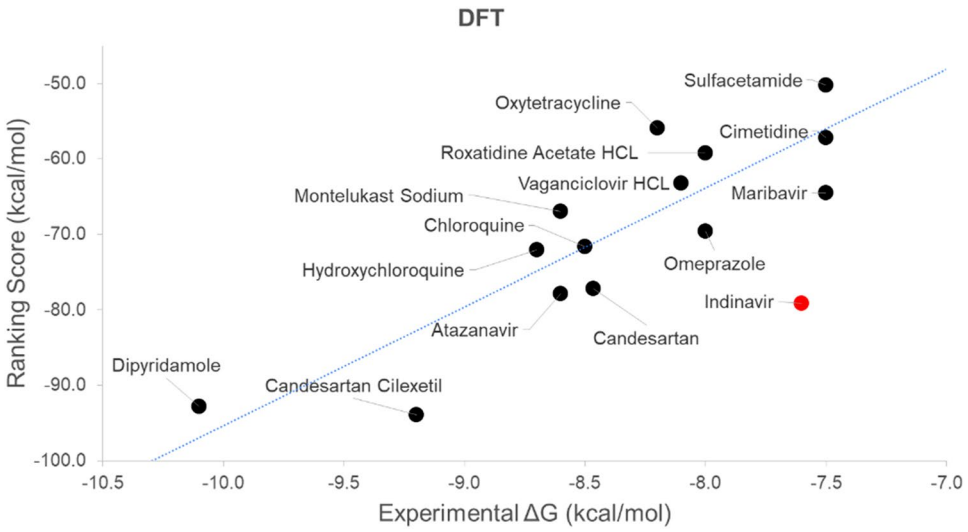

Fig. 3 Binding energies predicted using DFT. The trend line is calculated excluding the outlier indinavir. The overall R^2 value for all points (including indinavir) is 0.58. The R^2 value excluding indinavir is 0.75. The Predictive Indices with and without indinavir are 0.71 and 0.86, respectively

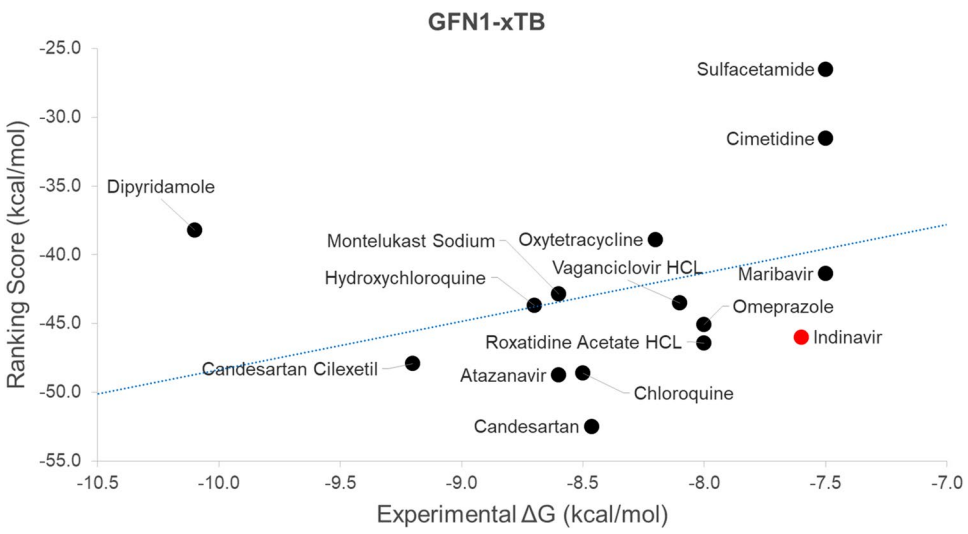

Fig. 4 Binding energies predicted using GFN1-xTB. The plotted trend line is calculated excluding the outlier indinavir. The overall R^2 value for all points (including indinavir) is 0.09. The R^2 value excluding indinavir is 0.13. The Predictive Indices with and without indinavir are 0.22 and 0.26, respectively

measure of the ability of a predictor to properly rank order) of 0.71. Only one poorly-binding ligand (indinavir) falls far off the correlation line for reasons that are not clear. It is particularly satisfying to observe that DFT very clearly differentiates the two ligands that bind best experimentally (dipyridamole and candesartan cilexetil) from the remaining ligands. Looking at correlation just among the weaker binders, by removing dipyridamole and candesartan cilexetil (and the outlier indinavir) from the set, retains an R^2 of 0.54. In contrast, the simpler semi-empirical QM approximation fails to capture the correlation for this ligand set, with an R^2 of 0.09 and a Predictive Index of 0.22. MM/GBSA showed better performance than GFN1-xTB, but appreciably worse than DFT, with an R^2 of 0.30 and a Predictive Index of 0.61.

Evaluation of the DFT results using ROC analysis [31] further corroborates this analysis. If we take the best half of the binders (top 7) and designate them as the "hits" (equivalent to designating all binders that bind better than 800 nM as hits), we get the curves shown in Fig. 5. The area under the curve for DFT is 0.89, reflecting an excellent ability to differentiate the better binders from the remaining ligands in this set.

In addition to the QM-based approaches we have applied, the publication that identified this data set [7] described the application of a classical mechanics FEP-based approach to the determination of absolute binding free energies for this set. They obtained some signal with their approach, with an R^2 of 0.29 and a Predictive Index of 0.57, although the results we obtained with DFT significantly improve on this (0.58 and 0.71, respectively). Given the diversity in formal charges among the ligands, which is better addressed with QM, this is not entirely surprising.

To further understand the origins of the differing predictions using semi-empirical GFN1-xTB and full DFT, in Fig. 6, we plot the binding energies computed using the two methods against each other. We color code the ligands by their charges, and we plot the energies of the different isomers considered for each ligand separately (as they may have different charges). Points in the plot are colored black for neutral, orange for positively charged, and green for negatively charged. As can be seen from the plot, the correlation between the semi-empirical and DFT ranking scores is acceptable when limited to positively charged ligands ($R^2 = 0.74$), but is considerably poorer for neutral ($R^2 = 0.38$) or negatively ($R^2 = 0.40$) charged ligands. Compared to our previous study of the Mcl-1 system where all ligands bore the same net charge [6], the poor performance of GFN1-xTB is most likely due to the inherent limitation of this semi-empirical method for handling a set of mixed ligands of various net charges. Charged systems, especially anions, are more challenging for semi-empirical methods because the corresponding wavefunctions and charge densities are parametrized by minimal atomic bases that cannot fully respond to large polarization effects.

As noted earlier, comparing multiple ligands with varying net charges is well-known to be challenging in comparative analysis. It is thus reassuring to observe that the charge of the ligand does not bias the DFT-based binding affinity predictions. This is, of course, a fundamental advantage of QM when compared to force-fields, and of full DFT calculations that use realistic bases when compared to semi-empirical quantum approximations. The predictive power of DFT on this set is particularly noteworthy because of the uncertainties associated with the conformations for both the bound ligand/complex and the unbound ligand and because the energy determinations are single-point energies, effectively at 0° K with no entropic contribution.

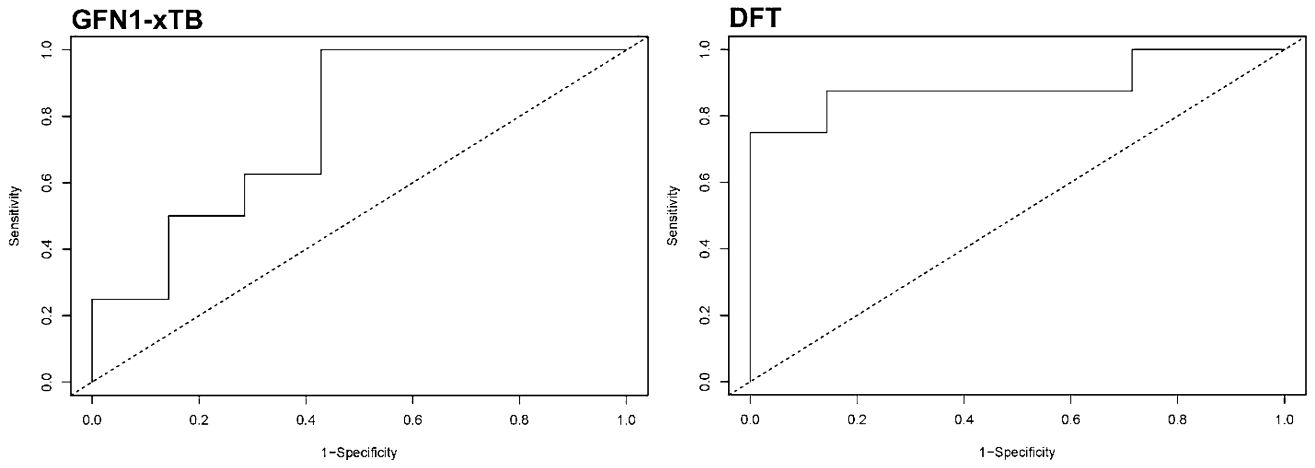

Fig. 5 The computed ROC curves for the GFN1-xTB and DFT data are shown here, with seven ligands that experimentally best bind (Ki < 800 nM) designated as true positives. The AUC for GFN1-xTB (left) is 0.77, while the AUC for DFT (right) is 0.89

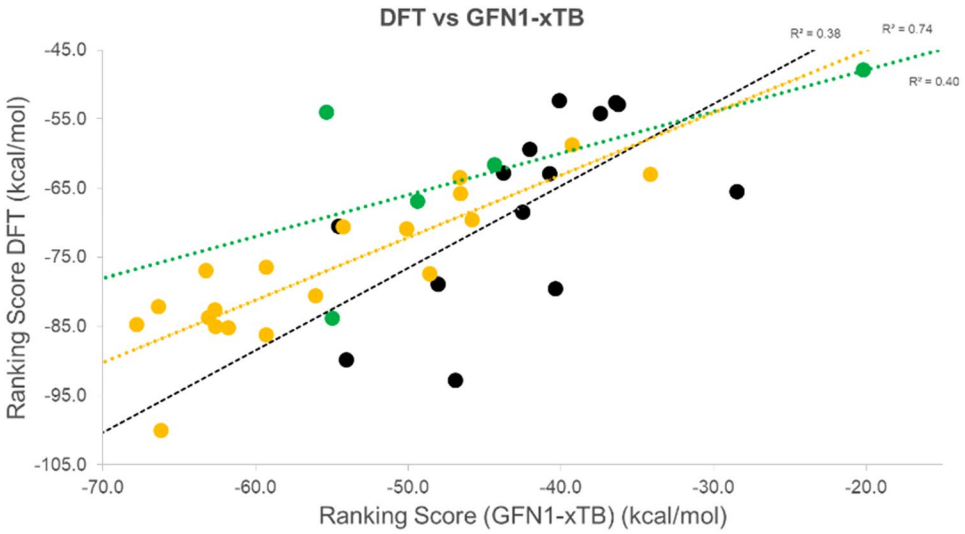

Fig. 6 A comparison of the scores calculated using the semi-empirical approach (GFN1-xTB) and DFT. Many of the 15 ligands are represented by multiple data points, corresponding to multiple isomers of that ligand. Black: neutral ligand isomers. Orange: positively charged ligand isomers. Green: negatively charged ligand isomers

Given the agreement between prediction and experiment for this dataset, we can analyse the predicted binding poses for the ligands to gain insights into the importance of protein residues that line the binding site. Looking at all the predicted bound ligand poses (including multiple poses in the case where multiple isomers/protomers were used for a particular ligand), we enumerated the propensity to make hydrogen bonds with residues of the protein. This leads to a simplistic pharmacophore map that may help identify critical interactions that should be maintained in a discovery campaign. Figure 7 presents the results of this analysis. For simplicity, we have chosen to only present data for the five protein residues observed to have the highest probability of interacting with the ligands included in this study (all ligands in the binding site were included in the analysis). These residues are GLU166, ASN142, GLN189, SER46, and GLY143. In some cases (e.g., atazanavir), the pharmacophore residues can form multiple hydrogen bonds with the ligands via side-chain and/or backbone hydrogen bond donors/acceptors (Fig. 7 right). Also, the strength of hydrogen bonds and the presence of other factors (e.g., hydrophobic effects) means the number

	E166	N142	Q189	S46	G143
Atazanavir	0.5	1.5	0.5	0.5	
Candesartan		0.5	1.0		1.0
Candesartan cilexetil	1.5	0.5	0.5		
Chloroquine	1.0		0.5		
Hydroxychloroquine	2.2	0.1	0.4	0.2	
Cimetidine	1.2	0.2	0.8		
Dipyridamole		3.0	1.0		
Indinavir	0.4	1.6	1.0	0.5	
Maribavir	1.5	1.0	1.0		
Montelukas	1.0			1.0	
Omeprazole	0.5		1.0		
Oxytetracycline	2.7	2.0	0.7	0.3	0.3
Roxatidine acetate	0.5	1.5			
Sulfacetamide	1.5	1.3	0.5		1.0
Vaganciclovir	1.8	0.8	1.0	0.2	0.2
mean	1.1	0.9	0.7	0.2	0.2

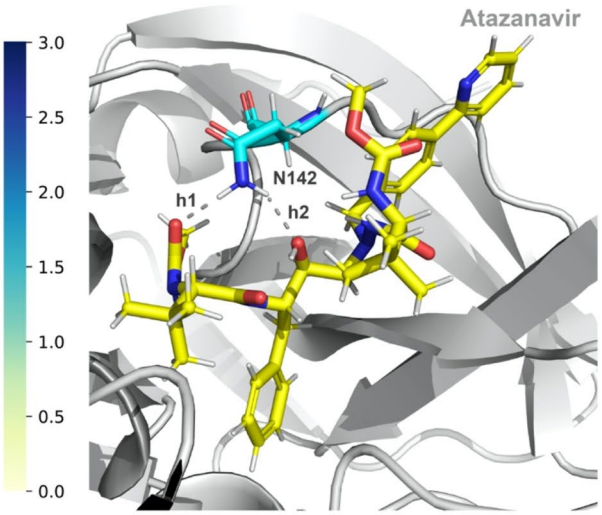

Fig. 7 (Left) The number of hydrogen bonds between a ligand and various M^{pro} residues that line the binding site. Only the five residues found most likely to form hydrogen bonds with the ligands are shown. Each number represents the average number of hydrogen bonds formed between a ligand-residue pair. The averaging was done first by averaging over different binding poses of the same isomer weighted by a Boltzmann factor, $exp\left(-E_i/(RT)\right)/Z$, where E_i is the DFT binding affinity, T is the temperature (298 K), R is the gas constant, and Z is $\sum_{i=1}^{2} exp(-E_i/(RT))$. Then, a linear averaging was calculated over different isomers of the same ligand. The number of hydrogen bonds for each residue averaged over all ligands is shown in the bottom row ("mean"). (Right) A snapshot showing hydrogen bonds (h1 and h2) formed between atazanavir and geometry-optimized M^{pro}

of hydrogen bonds alone is not a good indicator of binding affinity, as shown in Fig. 7 (left). For example, oxytetracycline makes the largest number of hydrogen bonds, but it is one of the weaker binders as determined experimentally (and as predicted).

Discussion

We have demonstrated that high-level quantum mechanics (density functional theory with dispersion corrections, using a realistic basis) can be successfully applied to rank order a scaffold-diverse set of ligands to a realistic protein receptor model—focusing on a set of existing drugs that are known to bind to the COVID-19 relevant protein M^{pro}. The full density functional treatment provides results that are substantially better than those obtained using a semi-empirical quantum approximation (which show almost no correlation for this dataset). Despite a quantum mechanical domain of nearly 3000 atoms, these calculations were carried out with realistic turnaround times and modest, accessible cloud-based computational resources using our recently described parallel implementation of DFT quantum mechanics. Each ligand isomer calculation required ~ 1 h in wall clock time with 14 AWS r5.24xlarge instances, with a compute cost of less than $90 (On-Demand instances) or $15 (Spot Instances).

It is worth noting that the first-principle physics-based nature of quantum mechanical calculations means that no target-dependent parameter fitting is required when applying this method to a specific molecular system. Methods that can be applied to a diverse ligand set like the one we have evaluated herein often incorporate system-specific learning or fitting. In such a case, the possibility of overfitting for a moderately-sized data set can be called into question. In contrast, the QM approach we have used is entirely prospective and the non-QM elements of the workflow (ligand conformer generation and molecular docking) have been applied agnostically and identically for all the ligands.

The focus of this study is a diverse set of experimentally determined binders to M^{pro} that appeared in the literature in the early stages of COVID-19-related research. While these experimental measurements were performed in a single lab, and are therefore expected to be consistent and useful for validation of scoring approaches, experimental determinations of their bound conformations have not been reported. Our work, therefore, incorporates both molecular docking for bound pose generation, and an assumption that these ligands all bind to the same site on the M^{pro} protein. We have striven to be systematic in how we applied both docking and QM-based scoring to avoid bias. The clear signal we obtain suggests our methods and assumptions are good ones, but the unresolved uncertainty in how/where these drugs actually bind means that the correlation we obtained may well be a lower bound on how well this approach could do in a case where the binding conformations were experimentally validated.

In light of the performance, scope of applicability, and throughput we have detailed, one could envision running a fully quantum-based screening campaign on hundreds, or even thousands, of compounds, with diverse scaffolds, charges, and chemical structures—an endeavour that would be extremely difficult or impossible using current methods based on force-fields. The flexibility of the quantum mechanical approach thus offers potentially new ways to use computation to advance drug discovery.

The idea that QM can be applied to drug discovery is not a new one. But earlier efforts have had to make a variety of compromises, e.g., via semi-empirical energy functions, fragment or linear-scaling approximations that introduce substantial cutoffs [32] or else have restricted QM to a small nucleus in QM/MM treatments [5]. In addition, the observed turnaround time using these approaches has typically been unrealistically long, on the timescale of days or weeks. These compromises have been an obstacle to realizing the predictive potential of QM in the context of drug discovery. For example, we see that semi-empirical parameterizations, even in their modern incarnations such as GFN1-xTB, lead to substantial errors in evaluating interactions such as charge transfer that are required to suitably assess diverse ligands. Similarly, QM/MM or fragmentation methods introduce errors caused by artificial boundaries and inaccurate treatment of long-range charge polarization [32–34]. What we have now demonstrated is that it is practical to treat a substantial region of a ligand/protein system – several thousand atoms—with full DFT. This can be accomplished without introducing compromises and on a realistic computational timescale, a significant practical advance over previous applications of QM to drug discovery.

It is also important to note that although the DFT calculations described herein performed quite well, these calculations only predict the enthalpy of binding at 0° K. Entropic contributions, including desolvation of the binding pocket, as well as the entropic changes arising from conformational variability of the ligand and protein, have not been included. The quality of results for this set suggests that these contributions may be of minor importance for this protein target and ligand set. However, looking more broadly, there will assuredly be systems where that is not the case. To address these issues, we are currently working on integrating corrections to the approach we have used to account for the desolvation entropy and the entropic contributions of the ligand and protein. We will report on these improvements in a future publication.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10822-021-00412-7) contains supplementary material, which is available to authorized users.

Acknowledgements We thank Garnet Chan for his continual and extensive advice and discussion as this work was carried out and Toru Shiozaki for his assistance with the quantum calculations and his constant encouragement of this work. We also thank Jasenka Fejzo of the University of Massachusetts, Amherst, for contributions related to isomer selection. Finally, we thank Satish Gandhi at Amazon Web Services (AWS), Simone Severini, Eric Kessler, and other members of the AWS Quantum Technologies team, and the AWS organization for the very generous computational support that made this work possible.

Declarations

Conflict of interest The authors served as employees of Quantum Simulation Technologies, Inc. during this project.

References

1. Cournia Z, Allen BK, Beuming T, Pearlman DA, Radak BK, Sherman W (2020) Rigorous free energy simulations in virtual screening. J Chem Inf Model 60(9):4153–4169. https://doi.org/10.1021/acs.jcim.0c00116
2. Pearlman DA, Rao BG (eds) (1998) Free energy calculations: methods and applications. Wiley, New York
3. Mobley DL, Gilson MK (2016) Predicting binding free energies. Annu Rev Biophys 46(1):1–28. https://doi.org/10.1146/annurev-biophys-070816-033654PMID-28399632
4. Merz KM (2014) Using quantum mechanical approaches to study biological systems. Acc Chem Res 47(9):2804–2811. https://doi.org/10.1021/ar5001023PMID-25099338
5. Ryde U, Söderhjelm P (2016) Ligand-binding affinity estimates supported by quantum-mechanical methods. Chem Rev 116(9):5520–5566
6. Mardirossian N, Wang Y, Pearlman DA, Chan GK-L, Shiozaki T (2020) Novel algorithms and high-performance cloud computing enable efficient fully quantum mechanical protein-ligand scoring. Arxiv.
7. Li Z, Li X, Huang YY, Wu Y, Liu R, Zhou L, Lin Y, Wu D, Zhang L, Liu H, Xu X, Yu K, Zhang Y, Cui J, Zhan CG, Wang X, Luo HB (2020) Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.2010470117
8. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. EMBO J 21(13):3213–3224. https://doi.org/10.1093/emboj/cdf327
9. 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 (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 368(6497):1331–1335. https://doi.org/10.1126/science.abb4489PMID-32321856
10. Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH (2013) Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. J Phys Chem 139:1089–7690
11. Lee T-S, Allen BK, Giese TJ, Guo Z, Li P, Lin C, McGee TD, Pearlman DA, Radak BK, Tao Y, Tsai H-C, Xu H, Sherman W, York DM (2020) Alchemical binding free energy calculations in AMBER20: advances and best practices for drug discovery. J Chem Inf Model. https://doi.org/10.1021/acs.jcim.0c00613
12. Lin M-H, Moses DC, Hsieh C-H, Cheng S-C, Chen Y-H, Sun C-Y, Chou C-Y (2018) Disulfiram can inhibit MERS and SARS coronavirus papain-like proteases via different modes. Antivir Res 150:155–163. https://doi.org/10.1016/j.antiviral.2017.12.015PMID-29289665
13. Zhang Y, Yang W (1998) Comment on "generalized gradient approximation made simple." Phys Rev Lett 80(4):890–890. https://doi.org/10.1103/physrevlett.80.890
14. Grimme S, Ehrlich S, Goerigk L (2011) Effect of the damping function in dispersion corrected density functional theory. J Comput Chem 32(7):1456–1465. https://doi.org/10.1002/jcc.21759PMID-21370243
15. Mardirossian N, Head-Gordon M (2017) Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. Mol Phys 115(19):2315–2372. https://doi.org/10.1080/00268976.2017.1333644
16. Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Phys Chem Chem Phys 7(18):3297–3305. https://doi.org/10.1039/B508541A
17. Knizia G (2013) Intrinsic atomic orbitals: an unbiased bridge between quantum theory and chemical concepts. J Chem Theory Comput 9(11):4834–4843. https://doi.org/10.1021/ct400687b
18. Fragalysis: Mpro crystal structures provided by the Fragalysis web application (Diamond Light Source/UK). (2020) https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro accessed
19. Chodera J, Lee AA, London N, Delft FV (2020) Crowdsourcing drug discovery for pandemics. Nat Chem 12(7):581–581. https://doi.org/10.1038/s41557-020-0496-2
20. Arcon JP, Modenutti CP, Avendaño D, Lopez ED, Defelipe LA, Ambrosio FA, Turjanski AG, Forli S, Marti MA (2019) AutoDock Bias: improving binding mode prediction and virtual screening using known protein–ligand interactions. Bioinformatics 35(19):3836–3838. https://doi.org/10.1093/bioinformatics/btz152
21. Grimme S, Bannwarth C, Shushkov P (2017) A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z = 1–86). J Chem Theory Comput 13(5):1989–2009. https://doi.org/10.1021/acs.jctc.7b00118
22. Mennucci B (2012) Polarizable continuum model. Wiley Interdiscip Rev Comput Mol Sci 2(3):386–404. https://doi.org/10.1002/wcms.1086
23. Case DA, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TEI, Cruzeiro VWD, Darden TA, Duke RE, Giambasu G, Gilson MK, Gohlke H, Goetz AW, Harris R., Izadi S, Izmailov SA, Kasavajhala K., Kovalenko A, Krasny R, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Man V, Merz KM, Miao Y, Mikhailovskii O, Monard G, Nguyen H, Onufriev AF, Pan, Pantano S, Qi R., Roe DR, Roitberg A, Sagui C., Schott-Verdugo S, Shen J., Simmerling CL, Skrynnikov, NR, Smith J, Swails J, Walker RC, Wang J, Wilson L, Wolf RM, Wu X., Xiong, Y., Xue, Y., York, D.M., Kollman, P.A.: Amber 20. In. University of California, San Francisco, (2020)
24. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified

generalized born model. Proteins Struct Funct Bioinformat 55(2):383–394. https://doi.org/10.1002/prot.20033
25. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 11(8):3696–3713. https://doi.org/10.1021/acs.jctc.5b00255
26. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174. https://doi.org/10.1002/jcc.20035PMID-15116359
27. DeLano WL (2020) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. In.
28. Rdkit (2020) RDKit version 2020.03.1 https://www.rdkit.org.in Accessed
29. Pracht P, Bohle F, Grimme S (2020) Automated exploration of the low-energy chemical space with fast quantum chemical methods. Phys Chem Chem Phys 22(14):7169–7192. https://doi.org/10.1039/c9cp06869d
30. Pearlman DA, Charifson PS (2001) Are free energy calculations useful in practice? a comparison with rapid scoring functions for the p38 MAP kinase protein system. J Med Chem 44(21):3417–3423
31. Goksuluk, D., Selcuk, K., Zararsiz, G., Karaagaoglu, AE (2016) easyROC: An interactive web-tool for ROC curve analysis using R language environment. R J 8:2 213–230.
32. Antony J, Grimme S (2012) Fully ab initio protein-ligand interaction energies with dispersion corrected density functional theory. J Comput Chem 33(21):1730–1739. https://doi.org/10.1002/jcc.23004
33. Fox SJ, Dziedzic J, Fox T, Tautermann CS, Skylaris CK (2014) Density functional theory calculations on entire proteins for free energies of binding: application to a model polar binding site. Proteins Struct Funct Bioinformat 82(12):3335–3346. https://doi.org/10.1002/prot.24686
34. Cole DJ, Skylaris CK, Rajendra E, Venkitaraman AR, Payne MC (2010) Protein-protein interactions from linear-scaling first-principles quantum-mechanical calculations. EPL (Europhys Lett) 91(3):37004. https://doi.org/10.1209/0295-5075/91/37004