THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Data Driven Computational Design and Experimental Validation of Drugs for Accelerated Mitigation of Pandemic-like Scenarios

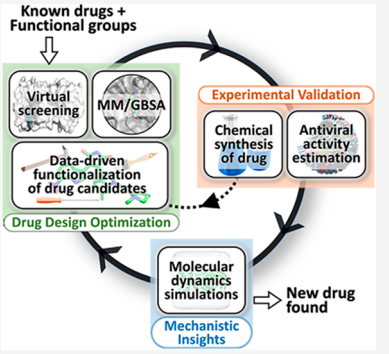
Samrendra K. Singh,[▽] Kelsie King,[▽] Cole Gannett,[▽] Christina Chuong,[▽] Soumil Y. Joshi,[▽] Charles Plate, Parisa Farzeen, Emily M. Webb, Lakshmi Kumar Kunche, James Weger-Lucarelli, Andrew N. Lowell, Anne M. Brown,* and Sanket A. Deshmukh*

Cite This: *J. Phys. Chem. Lett.* 2023, 14, 9490–9499

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Emerging pathogens are a historic threat to public health and economic stability. Current trial-and-error approaches to identify new therapeutics are often ineffective due to their inefficient exploration of the enormous small molecule design space. Here, we present a data-driven computational framework composed of hybrid evolutionary algorithms for evolving functional groups on existing drugs to improve their binding affinity toward the main protease (M^{pro}) of SARS-CoV-2. We show that combinations of functional groups and sites are critical to design drugs with improved binding affinity, which can be easily achieved using our framework by exploring a fraction of the available search space. Atomistic simulations and experimental validation elucidate that enhanced and prolonged interactions between functionalized drugs and M^{pro} residues result in their improved therapeutic value over that of the parental compound. Overall, this novel framework is extremely flexible and has the potential to rapidly design inhibitors for any protein with available crystal structures.

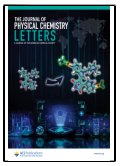


Functionalization of known small molecules and materials is a favored methodology to alter their properties and performance in catalysis, energy, environmental, biomedical, and biological applications.[1−4] For example, surface functionalization of metal nanoparticles and 2D materials has shown improved biocompatibility, solubility, and tunable thermal and electrical properties,[5,6] while functionalized polymers have shown alterations in crystallinity, wettability, and aggregation.[7] Similarly, functionalization of drug molecules has demonstrated improvements in their affinity to the target proteins.[8] However, identifying known drugs for functionalization and selecting high performing functional groups as well as sites on these drugs is a challenging and time-consuming process that is both inefficient and expensive.[9] For example, considering the functionalization of 10 drugs at 25 sites with 40 functional groups, a staggering 10^{41} distinct functionalized drugs would result, an impossibility considering current experimental approaches.

A powerful solution to search such a vast design space is applying evolutionary algorithms such as the genetic algorithm (GA)[10,11] and particle swarm optimization (PSO) algorithm.[12] GA is a metaheuristic optimization algorithm inspired by Darwinian evolution, which relies on crossover, mutation, and selection operations, to progress a population of evolving candidate solutions.[13] PSO, on the other hand, is inspired by the motion of a flock of birds searching for food, integrating the concept of particles (solutions) with swarms (groups of candidates).[14] Both GA and PSO have been successfully applied for applications such as the design of metallic alloys, coatings, porous materials, polymers, and biomaterials.[13,15,16] Furthermore, hybrid algorithms—which combine the capabilities of these algorithms with others or with machine learning approaches—have shown promise in designing materials with targeted sets of properties, for example, polymer matrix composites,[17] solid catalytic materials,[18] and electromagnets.[19] In the field of drug discovery, evolutionary algorithms and their hybrids have been used for docking of small molecules to protein targets, conformational analyses, drug dosing strategies, and crystal structure predictions.[20−22] However, their applicability in designing new drug molecules by the direct structural modification of existing drugs has not been explored.

Since its emergence in December 2019, Coronavirus Disease 2019 (COVID-19) has explosively spread, infecting over 765 million people and causing over 6.9 million fatalities around the world.[23] Fueled by an urgency to mitigate this viral illness and limited by the time and cost associated with the development of novel therapeutics, antiviral therapies employing repurposed noncovalent and covalently bound drugs, their derivatives and combinations, such as molnupiravir, remdesivir, lopinavir/ritonavir, ivermectin and paxlovid, have been

Received: June 26, 2023
Accepted: September 21, 2023
Published: October 18, 2023



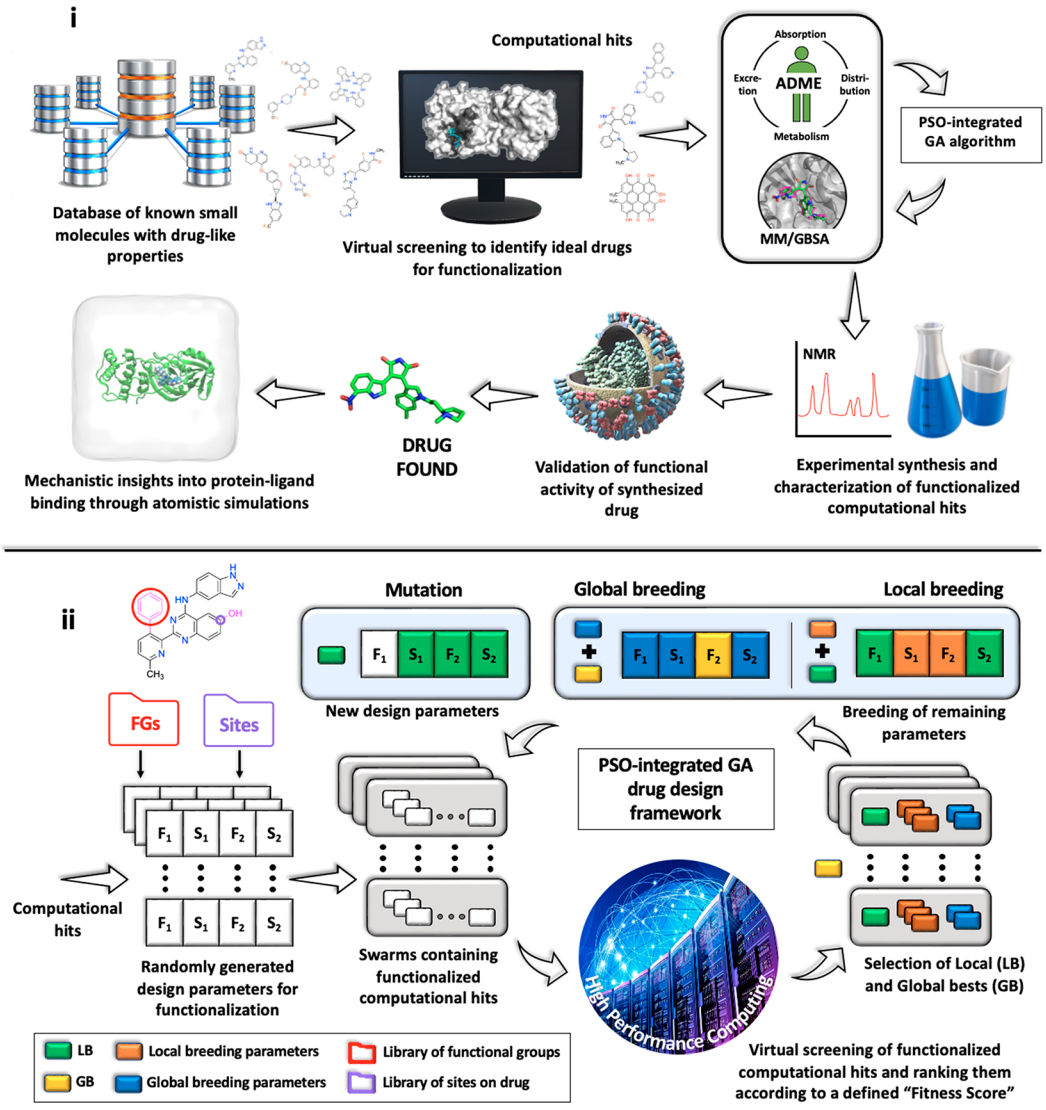


Figure 1. (i) Schematic representation of research workflow, (ii) PSO-integrated GA Framework. A detailed schematic of the PSO-integrated GA approach is shown in the SI.

examined through *in vitro* studies.[23,24] Significant work has also been carried out to effectively utilize computational approaches like virtual screening and molecular dynamics (MD) simulations to identify probable drug leads from existing databases for SARS-CoV-2 proteins, including its main protease (M^{pro}).[25−27] However, none of these studies identified new compounds tailored to the target, M^{pro}, which could improve the binding affinity and specificity. Several computational studies have attempted to design novel drugs through pharmacophore modeling, other structure-based drug design softwares or through machine learning models for SARS-CoV-2 M^{pro}.[28,29] Yet, few of these drugs have been experimentally validated to confirm the computational predictions.[30,31] A recent study by Huang et al. experimentally showcased the potential of functionalizing existing compounds to obtain a new-generation selective inhibitor for M^{pro}.[32] In their work, they focused on specific regions of the parent drug to design new functionalized drugs. However, their initial two attempts achieved limited success due to trial-and-error methods.[32] These challenges emphasize the need for fast, target-specific drug design to tackle similar pandemic-like scenarios in the future. Computational screening of existing small-molecule inhibitors, coupled with the data-driven optimization of its functionalization strategies, can enable a faster and more extensive exploration of potential inhibitors for target specific action.

A schematic representation of our integrated computational/experimental workflow is presented in Figure 1(i). Specifically, our novel approach uses high-throughput virtual screening and detailed computational analysis to identify known drugs that can be functionalized using a hybrid evolutionary algorithm, where PSO assists multiple GAs (PSO-integrated GA) as shown in Figure 1(ii). Our initial assumption was that PSO-integrated GA would rapidly modify existing drugs by introducing functional groups to improve their binding affinity toward a representative targeted protein, SARS-Cov-2 M^{pro}, while retaining their synthetic viability, drug likeness, toxicity, and oral druggability. Based on their protein−ligand binding free energies and ease of synthesizability, we identified, synthesized, and tested a functionalized derivative of a selected drug. Our derivative showed greater therapeutic value over the parental compound, thus validating our computational predictions. Furthermore, comparison of docked configurations for functionalized and unfunctionalized compounds,

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499

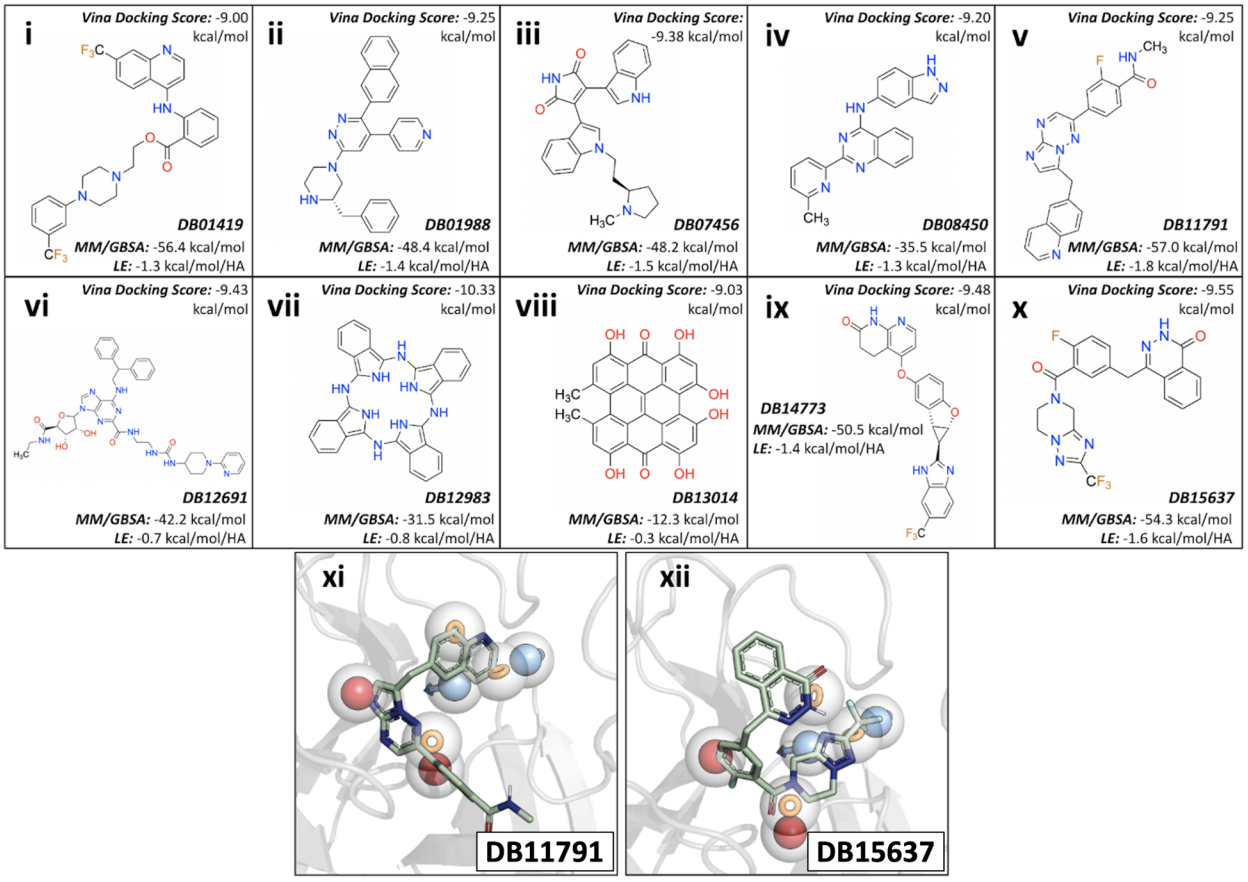


Figure 2. (i)−(x) 2D structures of the top 10 bare compounds as determined by virtual screening and lowest energy poses of (xi) DB11791 and (xii) DB15637 docked to SARS-CoV-2 M^{pro}. Structures are overlaid with pharmacophore model visualization, with aromatic groups represented as orange open circles, negative ionic groups as red spheres, and hydrogen bond donors represented as blue spheres with arrows. M^{pro} is shown as a gray cartoon.

coupled with atomistic MD simulations, showed that the addition of functional groups to drugs altered their binding poses, resulting in protein−ligand interactions with more residues for longer durations.

First, 9249 small molecules from the DrugBank database (Version 5.1.6) were virtually screened against three distinct SARS-CoV-2 M^{pro} crystal structures (PDB IDs: 6LU7, 6W63 and 6YB7) using the docking protocol reported in the Supporting Information (SI).[30,33,34] Table S3 lists the properties of top 10 computational hits that were selected based on docking scores from AutoDock Vina,[33] ADME (absorption, distribution, metabolism, and excretion) property predictions, and the highest number of normalized (normalized for compound size) interactions with key residues identified as critical to binding from residue-ligand interactions observed in various SARS-CoV-2 M^{pro} structures crystallized with ligands (Figure S1). As shown in Figure 2(i−x), all top 10 computational hits contained multiple aromatic features, as well as multiple hydrogen bond donating and accepting groups. Autodock Vina scores of bindings ranged from −10.33 to −9.0 kcal/mol. The average volume of these molecules was 1457 ± 360 $Å^3$, with molecular weights ranging from 354 to 778 g/mol and computationally calculated logP (octanol/water) values ranging from 3.2 to 7.7. Next, these top computational hits were evaluated for their protein binding affinity using the more robust MM/GBSA free energy calculations and ligand efficiency (LE)—defined as the free energy of binding divided by the number of heavy, non-hydrogen atoms (HA) in the compound.[35] Free energy of binding from MM/GBSA calculations and the ligand efficiencies, with an exception for the drug DB13014, ranged from −57.0 to −31.5 kcal/mol and from −1.8 to −0.7 kcal/mol/HA, respectively. Overall, most compounds were found to have favorable binding energies by all metrics, and exhibited interactions with key residues, indicating potential for inhibiting SARS-CoV-2 M^{pro}.

Pharmacophore modeling, a receptor-based technique, was used to approximate the ideal locations of ligand chemical features in the M^{pro} binding cavity to validate and provide a biochemical rationale for the binding of the 10 computational hits.[36] The top two candidates with the most negative ligand efficiency, DB11791 (LE = −1.8 kcal/mol/HA) and DB15637 (LE = −1.6 kcal/mol/HA), had three and one aromatic features, respectively, that were in range of the approximate locations suggested by the receptor pharmacophore model (Figure 2(xi−xii)). Aromatic interactions with HIS41 were observed for both compounds, and interaction with CYS145 was observed for DB11791 (Figures S4−S5). These data indicate that, generally, the top computational hits had chemical features suitable for binding critical M^{pro} residues and were thus chosen as leads for further functionalization.[37] Pharmacophore analysis was also used to compare the performance of the compounds identified in our work to that of a known inhibitor, 11a, which forms a covalent bond

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499

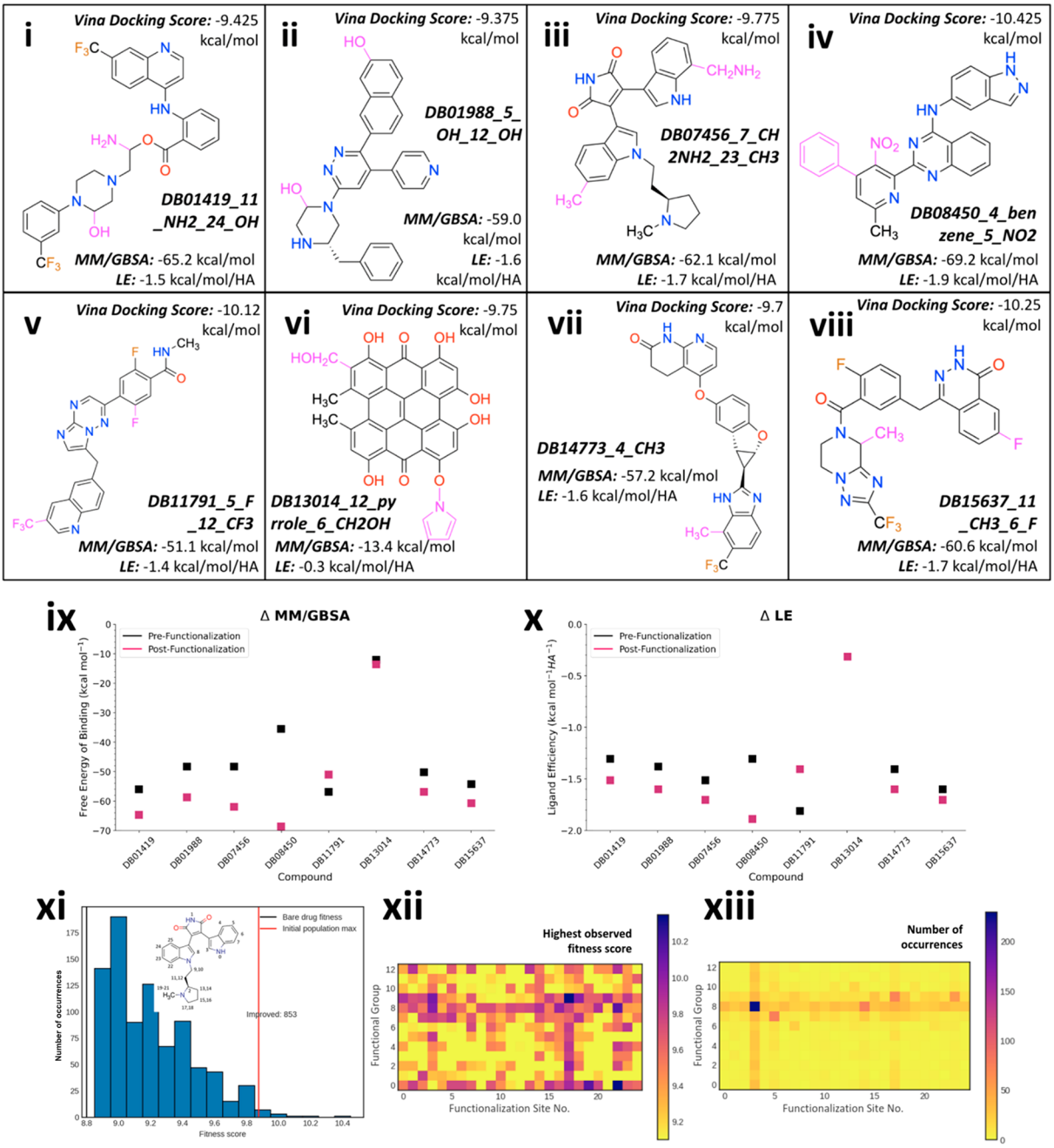


Figure 3. (i)−(viii) 2D structures of the top functionalized drugs. Pink functional groups show computational optimization. (ix)−(x) Differences in the MM/GBSA and LE values between the functionalized and unfunctionalized computational hits. (xi) Distribution of functionalized computational hits with fitness scores higher than the unfunctionalized compounds (inset: sketch for the bare drug DB07456 and the corresponding functionalization site IDs). (xii) Highest fitness scores were recorded for functional group−site pairs. Fitness scores are defined as the weighted averages of free energies for all three proteins used for docking multiplied by the fraction of Lipinski's Rule of 5 criteria met. (See SI.) (xiii) Occurrences of compounds with each combination of functional group and site IDs. Please refer to Table S2 in the SI for the names of functional groups corresponding to the given numbers.

with CYS145. A detailed account of all pharmacophore analyses for the computational hits can be found in Supplementary Text 1.

Analyzing the oral druggability of the top 10 computational hits based on their ADME properties (Table S3), 8 hits were selected and functionalized using PSO-integrated GA (see SI), with 13 common functional groups (FGs): C_6H_5 (benzene), C_2H_5, CF_3, CH_2OH, CH_3, CH_2NH_2, CHF_2, F, NH_2, NO_2, OCH_3, OH, and C_4H_4N (pyrrole) (Table S2 in the SI).[3,8,38,39] Excluded were DB12691 because it was too large and violated Lipinski's rule of five and DB12983 because it is a promiscuous drug that lacks specificity.[40] The PSO-integrated GA framework was developed to have aspects representative of both algorithms for optimization. The optimizer consisted of four individual swarms (N_s) containing eight particles each, which represented children in genetic algorithm framing. Thus, a total of 32 children—unique drug design parameters—were identified in each epoch of optimization. The 4 swarms aided the algorithm in exploring multiple independent regions of the search space simultaneously to accelerate drug discovery.

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499

Genetic operations including crossover and mutation, were used to update the drug design parameters after each epoch. Instead of implementing the conventional crossover between a select number of high performing sets of parameters ("parents") as seen in traditional GA, we used a probability-based crossover where best parameter sets from each swarm (L_{best}) as well as the entire framework (G_{best}), spread their genes into the current generation i to create design parameters for the next generation $i + 1$, symbolic of the PSO algorithm. The hybrid PSO-integrated GA algorithm was designed to design functionalized candidates with high "fitness scores"—absolute values of weighted free energy with penalties applied based on the number of Lipinski's Rule of Five violations, to create orally administrable compounds. To ensure rapid and straightforward chemical synthesis of the computationally optimized leads, two arbitrary design rules were applied onto the PSO-integrated GA: Rule 1: Drugs may be functionalized with up to two types of functional groups, and Rule 2: Both functional groups cannot replace a hydrogen atom off the same site.

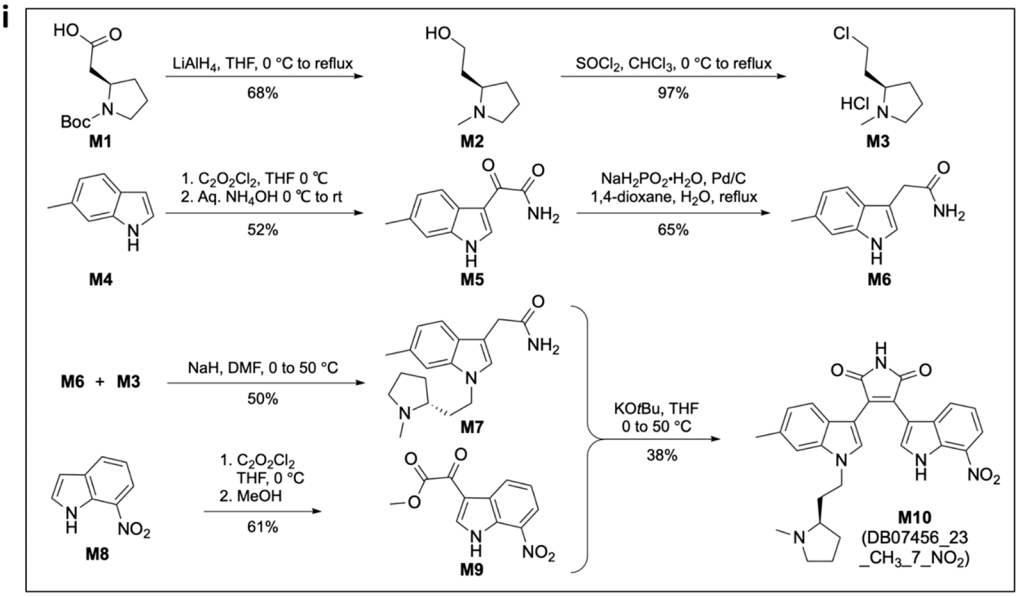
All optimizations were carried out for 100 epochs, resulting in 3200 functionalized candidates. Exhaustive tuning of the algorithm hyperparameters was carried out on DB07456 and DB11791. Furthermore, to highlight the improved performance of the PSO-integrated GA over traditional GA and random sampling, the same functionalization run of 3200 was carried out for DB07456 using all three methods as shown in Figure S7. It was observed that while the GA algorithm was able to find the overall best solution, the PSO-integrated GA yielded a superior top-10 solution set, which is a more desirable result for novel drug design. More details regarding the optimization process, the functionalized computational hits it identified, and the performance of PSO-integrated GA in different functionalization cycles can be found in Supplementary Text 2.

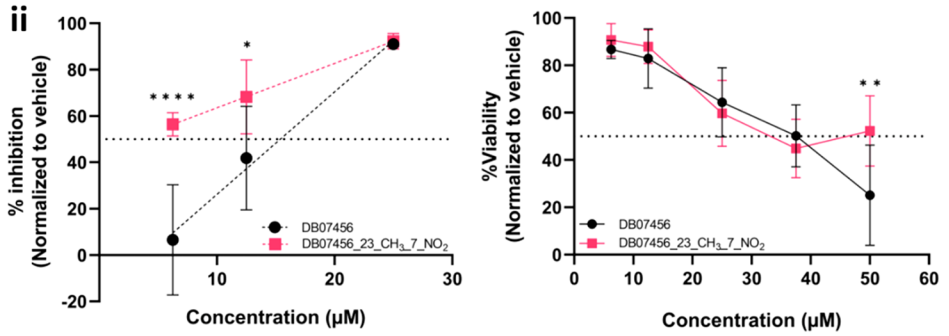
Using PSO-integrated GA, functionalized versions of the computational hits—FCHs ("functionalized computational hits"), with better binding performance than their respective unfunctionalized counterparts, UCHs ("unfunctionalized computational hits"), were identified by searching as low as 2% of the entire search space as shown in Table S4. Top FCHs were identified for all compounds using the ranking protocol described in SI. All FCHs are named according to the convention: DrugBankID_SiteID1_FG1_SiteID2_FG2. Figure 3(i−viii) and Table S5 lists the top FCHs, which were selected based on MM/GBSA free energy calculations, LE, and interactions with critical binding pocket residues. Overall, these data indicate that functionalization of UCHs with the PSO-integrated GA framework generally improved their performance, producing energetically favorable FCHs relative to the UCHs. MM/GBSA free energies of binding ranged from −69.2 to −51.1 kcal mol^{-1} (median = −59.8 kcal mol^{-1}) and ligand efficiencies ranged from −1.9 to −1.4 kcal mol^{-1} HA^{-1} (median = −1.60 kcal mol^{-1} HA^{-1}) for the FCHs (Figure 3(ix−x)) compared to those of the UCHs which were −57.0 kcal/mol to −31.5 kcal/mol, and from −1.8 kcal/mol/HA to −0.7 kcal/mol/HA. Overall, seven out of eight chosen compounds exhibited an improvement in binding energy post-functionalization. However, the drug DB13014, which originally showed relatively poor binding compared to the other candidates, did not change significantly after functionalization, showing that addition of functional groups will not necessarily result in compounds with improved binding affinities toward the target protein. This result also highlights the importance of the computational framework as an efficiency measure; it enabled us to identify candidates that likely would not improve with functionalization. Additionally, the functionalizations installed by the algorithm maintained ADME properties within acceptable ranges for known drug-like compounds. FCHs generally agreed with aromatic pharmacophore feature positioning (Figure S8). Tables and figures for the top 5 FCHs identified for every compound except DB13014 are also presented in the SI (Tables S6−S12 and Figures S9−S15).

Analysis of the data generated during the data-driven functionalization process, revealed that electron withdrawing $-CF_3$, $-NO_2$, and −F, as well as hydrophobic pyrrole and $-CH_3$ functional groups could be critical in improving binding affinity of selected compounds toward M^{pro}. Table S13 shows a list of FGs and functionalization sites on a parent drug that occurred commonly in the PSO-integrated GA algorithm, as well as those that were found in the top 10 FCHs. Furthermore, as a representative example, Figure 3(xii) shows the highest observed fitness scores for different combinations of FGs and sites for DB07456. The number of occurrences of each combination can be seen in Figure 3(xiii). FGs 7, 8, and 9 (−F, $-NH_2$, and $-NO_2$, respectively) and sites 3 and 17 from DB07456 were most favored due to more consistent high fitness scores relative to other possible functionalizations. From the total of 3200 screened compounds, the algorithm produced only 239 derivatives (7.47%) with more favorable fitness scores than the unfunctionalized DB07456 (9.375), with scores as high as 10.4 observed (Figure 3(xi)). More performance schematics for the functionalization of the different shortlisted compounds are shown in Figures S16 to S23. Table S14 shows the overall best functionalized computational hits identified from the top 8 compounds. MM/GBSA free energies of binding for these high performing compounds ranged from −69.2 to −59.0 kcal mol^{-1} and ligand efficiencies ranged from −1.9 to −1.4 kcal mol^{-1} HA^{-1}, indicating that the top functionalized computational hits performed much better than all of the unfunctionalized hits.

From the overall top 8 compounds presented in Table S14, two derivatives each were observed for DB01419, DB15637, and DB07456. However, upon closely inspecting their functionalization patterns, the two derivatives for DB07456, DB07456_23_CH_3_7_CH_2NH_2 and DB07456_23_CH_3_7_CHF_2 with decreased free energies by −15 kcal mol^{-1} (−63.2 vs −48.2 kcal mol^{-1} of the unfunctionalized DB07456) and −11.9 kcal mol^{-1} (−60.1 vs −48.2 kcal mol^{-1}), respectively, were found to differ through only one functional group (Figure S24(i−ii)). While both compounds interacted with the protein through similar interactions, a notable shift in the interaction propensity was observed for THR25 via CH_2NH_2 and CHF_2 addition (Figures S24(iii−iv) and S25). Also, increased interactions were observed for both DB07456_23_CH_3_7_CH_2NH_2 and DB07456_23_CH_3_7_CHF_2, and residues HIS41 and GLN189 in the most favorable docked pose, although the position of the indole group was variable in both functionalized and unfunctionalized docked positions (Figure S24(v−vi)). Moreover, these derivatives were also relatively insensitive to mutations in the binding cavity relative to other compounds in the tenth percentile (Figure S26). Based on these encouraging results, we further docked and calculated MM/GBSA free energies for derivatives of DB07456 by fixing $-CH_3$ at site 23

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499





Compound	EC50 (μM)	CC50 (μM)	SI
DB07456	15.38	35.81	2.33
DB07456_23_CH_3_7_NO_2	**2.94**	22.63	**7.70**

Figure 4. Experimental validation of drug performance: (i) Concise synthesis of computational lead M10 (DB07456_23_CH_3_7_NO_2). (ii) Inhibition of SARS-CoV-2 (left) and toxicity levels in mammalian cells (right) between DB07456 and DB07456_23_CH_3_7_NO_2 with the corresponding EC50, CC50, and selectivity index values.

and modifying site 7 with all remaining functional groups. It was observed that all derivatives attained similar energetically favorable poses in the binding pocket (Figure S27) that resulted in comparable MM/GBSA and LE values (Table S15). Thus, we advanced several DB07456 derivatives simultaneously with the goal of creating a derivative for antiviral testing to validate our computationally guided approach.

Of the various derivatives, the use of 7-nitroindole (**M8**, Figure 4(i)) proved most amenable for synthesis among the possible DB07456 derivatives, directly furnishing the right hemisphere of DB07456_23_CH_3_7_NO_2 (**M10**) as the 7-nitroindole-3-oxoacetate methyl ester (**M9**) via Friedel−Crafts acylation with oxalyl chloride and methanol quench.[41] The left hemisphere of **M10** was constructed using a nucleophilic displacement strategy.[42] Electrophile **M3** was derived from *N*-Boc-homoproline (**M1**)[43] via global reduction (**M2**) followed by treatment with thionyl chloride. Acylation of 6-methylindole (**M4**) with oxalyl chloride and quenching with aqueous ammonia furnished **M5**,[41] which was carefully reduced by the slow addition of sodium hypophosphate giving acetamide **M6**. Use of **M6** as a nucleophile with **M3** completed the left hemisphere (**M7**). Fragment union under basic conditions resulted in DB07456_23_CH_3_7_NO_2 (**M10**).[41] In addition to being the most readily synthesized compound (four linear steps, seven total steps, 6.4%), **M10** also had the most favorable binding energy of all of the docked derivatives (−63.2 kcal mol^{-1}) (Table S15).

To assess whether the derived compound DB07456_23_CH_3_7_NO_2 increased antiviral activity, we compared the bare and derived compounds' ability to limit SARS-CoV-2 replication in Vero E6-TMPRSS2-T2A-ACE2 cells (Figure 4(ii)). Both compounds effectively inhibited SARS-CoV-2 > 90% at 25 μM. However, cells treated with DB07456_23_CH_3_7_NO_2 had significantly decreased viral replication at 12.5 μM (p = 0.0147) and 6.25 μM (p = <0.0001) when compared to DB07456. The antiviral studies established an improved half maximal effective concentration

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499

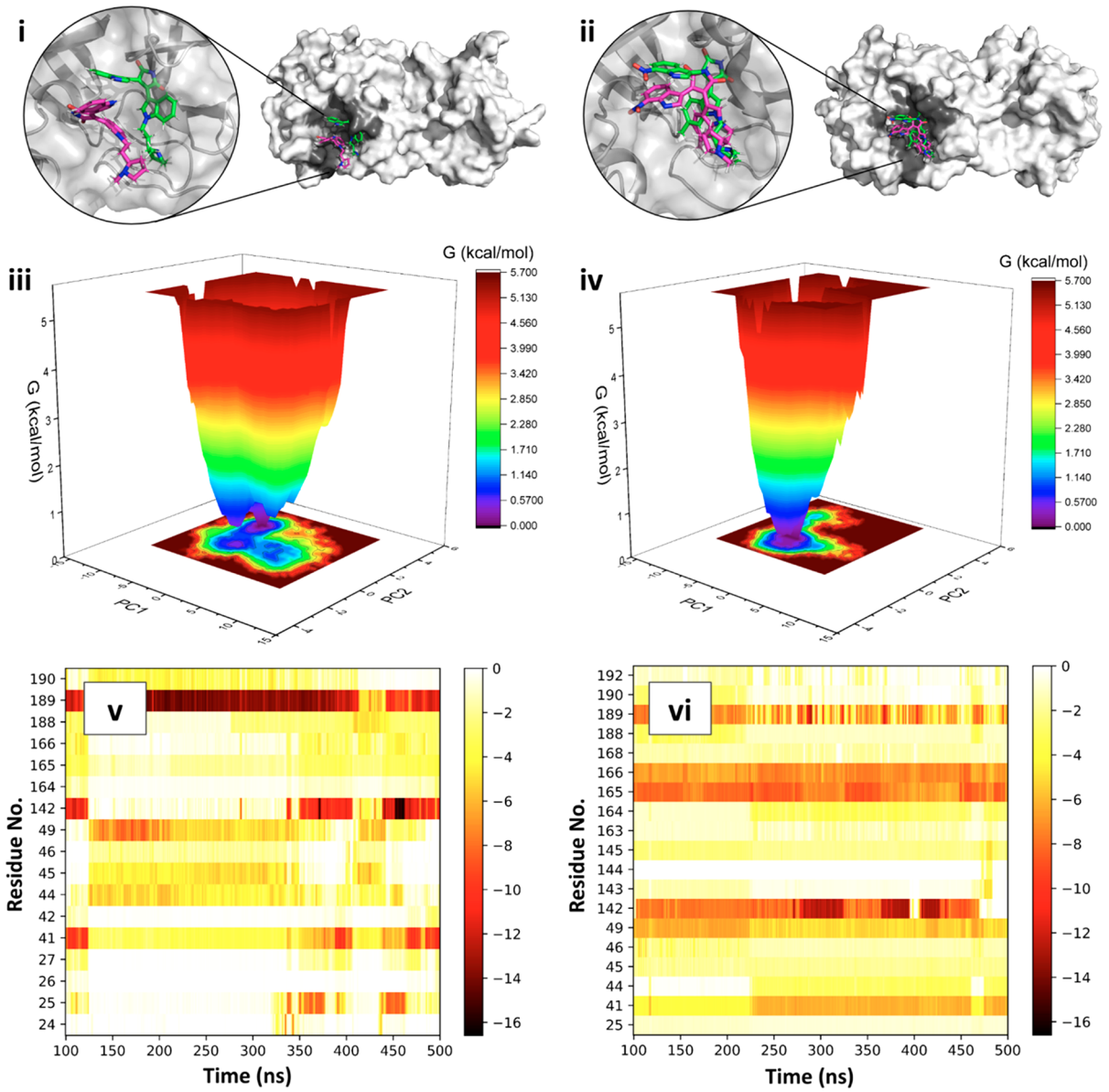


Figure 5. (i)−(ii) The protein ligand complex clustered structure-dominant morphology. (iii)−(iv) Free energy landscapes. (v)−(vi) Major energetic contributions for interacting residues over the period of the production run. All data is presented for first pose bare drug (left column) and first pose functionalized drug (right column) simulations as representative examples.

(EC_{50}), for the derived compound at 2.94 μM compared to 15.38 μM for the bare drug.

To further evaluate the therapeutic potential of DB07456_23_CH_3_7_NO_2, cell proliferation MTS assays were conducted to assess the drug toxicity. Percent cell viability, normalized to the vehicle control, displayed similar patterns between the parental and the functionalized compounds (Figure 4(ii)). We found that the half maximal cytotoxic concentration (CC_{50}), shifted lower for DB07456_23_CH_3_7_NO_2 (22.63 μM) compared to un-functionalized DB07456 (35.81 μM); however, this is not a statistically significant difference. Additionally, at a concentration of 50 μM, the functionalized compound was significantly less toxic compared to DB07456 (p = 0.0032). The cumulative data show an increased selectivity index, or the ratio between cytotoxicity and antiviral activity (CC_{50}/EC_{50}), for DB07456_23_CH_3_7_NO_2 at 7.70, showing greater therapeutic value over the parental compound with a selectivity index of 2.33. The observed antiviral effect for DB07456_23_CH_3_7_NO_2 showed better EC50 value (2.94 μM) than some of the other proposed antivirals for SARS-CoV-2, such as Remdesivir (EC_{50} = 23.15 μM), Lopinavir (EC_{50} = 26.63 μM), Ribavirin (EC_{50} > 500 μM), Galidesivir (EC_{50} > 100 μM), Ritonavir (EC_{50} > 100 μM), Oseltamivir carboxylate (EC_{50} > 100 μM), Baloxivir acid (EC_{50} > 100 μM), and Favipiravir (EC_{50} > 100 μM).[44] However, it had a higher EC50 value than Homoharringtonine (EC_{50} = 2.55 μM) and Emetine (EC_{50} = 0.46 μM).[44,45] Nonetheless, the improvement in the EC50 value of the DB07456_23_CH3_7_NO2 as compared to the parent, DB07456 (EC50 = 15.38 μM), indicates the merit of our framework in improving the performance of weaker ligands through functionalization, which may be imperative in time-sensitive situations.

To further explore the protein−ligand binding mechanism and to probe the role of functional groups, we performed atomistic MD simulations in the presence of explicit water for

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499

the top 3 poses of bare and functionalized drugs obtained from docking studies. Additional details regarding the MD methodology are provided in Section S7 of SI. During 500 ns MD simulations, both bare (DB07456) and functionalized (DB07456_23_CH_3_7_NO_2) drugs were found to stay in the binding cavity of the initial docked pose, with minor repositioning to explore additional sampling space not available in static molecular docking experiments (Figure 5(i−ii) and Figure S33). The simulation trajectories are presented as Movies S3, S4, and S5. The backbone root-mean-square deviation (RMSD) analysis showed a mostly unchanging structural conformation of M^{pro} throughout the simulation for most ligand-bound systems. From the root-mean-square fluctuation (RMSF) analysis, lower fluctuation of the protein backbone atoms, including those for key catalytic residues HIS41 and CYS145, was observed in systems with ligands, as compared to the apo protein simulation (Figure S37). Lack of change in structural orientation of HIS41 and CYS145 in the presence of both bare and functionalized drugs highlights ligand fitness in the binding pocket.

To identify the differences in protein secondary structures, in the presence and absence of ligands, results from the apo- and ligand-bound protein simulations were investigated and compared. Interestingly, it was observed that β-sheets within the region B3 lost some structural order and were observed as loops in the simulation with the apo protein but not in simulations with ligand-bound proteins (Figures S38−S40). This loss of secondary structure in the apo protein simulation highlights the structural stabilization of the protein due to the presence of ligand at the onset of simulation. Analysis of the free energy landscape (FEL) values showed that in the absence of the drug, the protein explored multiple local minima separated by low energy barriers and the trough of FEL was broader. In the presence of bare or functionalized drug simulations, as observed in Figure S42, the troughs were narrower and the proteins explored less conformational space compared to apo protein, indicating that the presence of drugs limited conformational sampling. As shown in Figure 5(iii−iv), FEL troughs were even narrower for the functionalized drug simulations as compared to bare drug simulations indicating their conformational stability in the presence of ligand. Analyzing the residue-wise interaction energies between the protein and docked ligands, it was observed that functionalized drugs interacted with more residues than the bare drugs, resulting in added energetic stability of these ligands in the binding pocket. The time-evolution of residue-wise nonbonded energy values showed that all poses of bare and functionalized ligands showed favorable energetic stabilization due to specific protein residues. For the first pose bare drug simulations, as shown in Figure 5(v), dominant interactions were observed with GLN189, while for first pose functionalized drug simulations, prolonged and favorable interactions were observed with ASN142, MET165 and GLU166. Figure S44 shows the time-evolution of residue-wise nonbonded energy values for the other two drug poses. It was also observed that functionalized drugs displayed prolonged interactions with residues HIS41 and MET49 which were much weaker for the bare drug simulations as shown in Figures S45 and S46, indicating the strong affinities of the functional groups to these residues. The analysis of the distance between COMs of ligands and protein tiered residues (Figures S47) revealed that functionalized drugs exhibited lower mobility compared to unfunctionalized drugs, possibly due to robust and enduring protein−ligand interactions. Additionally, initial fluctuations in the solvent-accessible surface area (SASA) of protein binding pocket residues were observed, ultimately stabilizing toward the end of the simulation indicating both types of drugs attained stable binding configurations in the protein cavity (Figures S48). In the case of all functionalized drug simulations, the overall number and duration of major hydrophilic and hydrophobic protein−ligand interactions appeared to be higher than those for the bare drug simulations indicating stronger ligand binding (Figures S49 to S52). This analysis also suggests that the functionalized drug fluctuated between two or three dominant conformations in these complexes but always strongly interacted with the key residues. These enhanced and prolonged interactions between functionalized drugs and protein residues may result in its improved EC_{50} values compared to bare drugs.

We presented an integrated computational and experimental approach to accelerate drug design by designing novel drugs for M^{pro} of SARS-CoV-2. Our computational framework is used to shortlist drug candidates from a drug database based on their binding affinity toward M^{pro}. These shortlisted drugs are functionalized, to produce hits with better binding affinity, using a data-driven approach composed of hybrid-evolutionary algorithms, which significantly reduced the design space. Experimental synthesis and validation of one of the functionalized drug candidates further demonstrates the reliability of our computational approach. A mechanistic understanding developed by atomistic MD simulations reveals that functionalized drugs have enhanced and prolonged interactions with the key residues, resulting in their better performance compared to parent drug. Overall, the proposed PSO-integrated GA approach is very versatile and is currently being applied to functionalize and design other materials such as metal organic frameworks (MOFs),[46] glycomaterials, polymers, etc. Furthermore, readers can easily adapt our framework from the pseudocode presented in the SI to accelerate drug discovery for any given target protein. Our ongoing efforts are focused on designing new drugs by functionalizing known drugs for SARS-CoV-2 such as Nirmatrelvir using our data-driven framework.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c01749.

Performance of traditional GA search (MP4)
Performance of PSO-integrated GA search (MP4)
Trajectory for the apo protein simulation (MP4)
Trajectory for the ligand-bound protein simulation with bare drug pose 1 (MP4)
Trajectory for the ligand-bound protein simulation with functionalized drug pose (MP4)
Additional framework details, experimental and simulation protocols, and longer descriptions of analysis results and figures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Sanket A. Deshmukh − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-7573-0057; Phone: +1 540-231-8785; Email: sanketad@vt.edu

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499

Anne M. Brown − *Research and Informatics, Interdisciplinary Program in Genetics, Bioinformatics, and Computational Biology, Center for Emerging, Zoonotic, and Arthropod-borne Pathogens, and Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-6951-8228; Phone: +1 540-231-9231; Email: ambrown7@vt.edu

Authors

Samrendra K. Singh − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Kelsie King − *Research and Informatics and Interdisciplinary Program in Genetics, Bioinformatics, and Computational Biology, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-5393-2077

Cole Gannett − *Department of Chemistry and Center for Emerging, Zoonotic, and Arthropod-borne Pathogens, Virginia Tech, Blacksburg, Virginia 24061, United States*

Christina Chuong − *Department of Biomedical Sciences and Pathobiology, VA-MD Regional College of Veterinary Medicine and Center for Emerging, Zoonotic, and Arthropod-borne Pathogens, Virginia Tech, Blacksburg, Virginia 24061, United States*

Soumil Y. Joshi − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0003-1531-0098

Charles Plate − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Parisa Farzeen − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Emily M. Webb − *Department of Entomology, Virginia Tech, Blacksburg, Virginia 24061, United States*

Lakshmi Kumar Kunche − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-6914-0365

James Weger-Lucarelli − *Department of Biomedical Sciences and Pathobiology, VA-MD Regional College of Veterinary Medicine and Center for Emerging, Zoonotic, and Arthropod-borne Pathogens, Virginia Tech, Blacksburg, Virginia 24061, United States*

Andrew N. Lowell − *Department of Chemistry, Center for Emerging, Zoonotic, and Arthropod-borne Pathogens, and Faculty of Health Sciences, Translational Biology, Medicine, and Health, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-5357-5279

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.3c01749

Author Contributions

[▽](S.K.S., K.K., C.G., C.C., and S.Y.J.) These authors contributed equally. S.A.D. conceived the idea. S.A.D. and A.M.B. codirected the project. S.K.S. and C. P. set up the data driven computational framework and performed analysis. S.Y.J. and S.A.D. assisted in analysis of this data. P.F. and L.K.L. performed the MD simulations. P.F., L.K.L., S.Y.J., A.M.B., and S.A.D. analyzed the related data. K.K. performed molecular docking, pharmacophore modeling, and related analysis. C.G. and A.N.L. synthesized and characterized the functionalized drugs. A.N.L. guided experimental synthesis. C.C., E.M.W., and J.W.-L. performed testing, validation, and related analysis of experiments with M^{pro}. J.W.-L. guided these experimental efforts. All authors contributed to discussing results and composing the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to acknowledge the Advanced Research Computing (ARC), Virginia Tech, for providing the computational resources. S.A.D. acknowledges the ICTAS Junior Faculty Award from Virginia Tech. This work was supported by GlycoMIP, a National Science Foundation Materials Innovation Platform funded through Cooperative Agreement DMR-1933525. S.A.D. and J.W.-L. acknowledge the support from PIPP Phase 1 funding # 2200045.

■ REFERENCES

(1) Collins, S. P.; Daff, T. D.; Piotrkowski, S. S.; Woo, T. K. Materials Design by Evolutionary Optimization of Functional Groups in Metal-Organic Frameworks. *Sci. Adv.* **2016**, *2* (11), No. e1600954.

(2) Huang, H.; Qiao, Y.; Yuan, Y.; Zhang, J. Surface Functionalization for Heterogeneous Catalysis. In *Reference Module in Materials Science and Materials Engineering*; Elsevier: 2022. DOI: 10.1016/B978-0-12-822425-0.00073-7.

(3) Dirocco, D. A.; Dykstra, K.; Krska, S.; Vachal, P.; Conway, D. V.; Tudge, M. Late-Stage Functionalization of Biologically Active Heterocycles through Photoredox Catalysis. *Angew. Chem., Int. Ed. Engl.* **2014**, *53* (19), 4802−4806.

(4) Canossa, S.; Wuttke, S. Functionalization Chemistry of Porous Materials. *Adv. Funct. Mater.* **2020**, *30* (41), No. 2003875.

(5) Nandanapalli, K. R.; Mudusu, D.; Lee, S. Functionalization of Graphene Layers and Advancements in Device Applications. *Carbon N. Y.* **2019**, *152*, 954−985.

(6) Sanità, G.; Carrese, B.; Lamberti, A. Nanoparticle Surface Functionalization: How to Improve Biocompatibility and Cellular Internalization. *Front Mol. Biosci* **2020**, *7*, No. 587012.

(7) Tas, S.; Kopec′, M.; van der Pol, R.; Cirelli, M.; de Vries, I.; Bolukbas, D. A.; Tempelman, K.; Benes, N. E.; Hempenius, M. A.; Vancso, G. J.; de Beer, S. Chain End-Functionalized Polymer Brushes with Switchable Fluorescence Response. *Macromol. Chem. Phys.* **2019**, *220*, No. 1800537.

(8) Cernak, T.; Dykstra, K. D.; Tyagarajan, S.; Vachal, P.; Krska, S. W. The Medicinal Chemist's Toolbox for Late Stage Functionalization of Drug-like Molecules. *Chem. Soc. Rev.* **2016**, *45*, 546−576.

(9) Börgel, J.; Ritter, T. Late-Stage Functionalization. *Chem.* **2020**, *6* (8), 1877−1887.

(10) Holland, J. H. Genetic Algorithms. *Sci. Am.* **1992**, *267* (1), 66−73.

(11) Kramer, O. Genetic Algorithms. In *Genetic Algorithm Essentials*; Kramer, O., Ed.; Springer International Publishing: Cham, 2017; pp 11−19. DOI: 10.1007/978-3-319-52156-5_2.

(12) Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*; IEEE: 1995; Vol. 4, pp 1942−1948. .

(13) Paszkowicz, W. Genetic Algorithms, a Nature-Inspired Tool: Survey of Applications in Materials Science and Related Fields. *Materials and Manufacturing Processes* **2009**, *24*, 174−197.

(14) Rezaee Jordehi, A.; Jasni, J. Particle Swarm Optimisation for Discrete Optimisation Problems: A Review. *Artificial Intelligence Review* **2015**, *43* (2), 243−258.

(15) Paszkowicz, W. Genetic Algorithms, a Nature-Inspired Tool: A Survey of Applications in Materials Science and Related Fields: Part II. *Materials and Manufacturing Processes* **2013**, *28*, 708−725.

(16) Jennings, P. C.; Lysgaard, S.; Hummelshøj, J. S.; Vegge, T.; Bligaard, T. Genetic Algorithms for Computational Materials Discovery Accelerated by Machine Learning. *npj Computational Materials* **2019**, *5* (1), 1−6.

(17) Patra, T. K.; Meenakshisundaram, V.; Hung, J.-H.; Simmons, D. S. Neural-Network-Biased Genetic Algorithms for Materials

Design: Evolutionary Algorithms That Learn. *ACS Comb. Sci.* **2017**, *19*, 96−107.
(18) Rodemerck, U.; Baerns, M.; Holena, M.; Wolf, D. Application of a Genetic Algorithm and a Neural Network for the Discovery and Optimization of New Solid Catalytic Materials. *Appl. Surf. Sci.* **2004**, *223* (1), 168−174.
(19) Rahmat-Samii, Y. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) in Engineering Electromagnetics. In *17th International Conference on Applied Electromagnetics and Communications, 2003. ICECom 2003*; IEEE: 2003; pp 1−5. DOI: 10.1109/ICECOM.2003.1290941.
(20) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Jr. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66* (1), 334−395.
(21) Davis, A. M.; Plowright, A. T.; Valeur, E. Directing Evolution: The next Revolution in Drug Discovery? *Nat. Rev. Drug Discovery* **2017**, *16* (10), 681−698.
(22) Devi, R. V.; Sathya, S. S.; Coumar, M. S. Evolutionary Algorithms for de Novo Drug Design − A Survey. *Applied Soft Computing* **2015**, *27*, 543−552.
(23) Murakami, N.; Hayden, R.; Hills, T.; Al-Samkari, H.; Casey, J.; Del Sorbo, L.; Lawler, P. R.; Sise, M. E.; Leaf, D. E. Therapeutic Advances in COVID-19. *Nat. Rev. Nephrol.* **2023**, *19* (1), 38−52.
(24) Andrzejczyk, J.; Jovic, K.; Brown, L. M.; Pascetta, V. G.; Varga, K.; Vashisth, H. Molecular Interactions and Inhibition of the SARS-CoV-2 Main Protease by a Thiadiazolidinone Derivative. *Proteins* **2022**, *90* (11), 1896−1907.
(25) Thiel, V.; Herold, J.; Schelle, B.; Siddell, S. G. Viral Replicase Gene Products Suffice for Coronavirus Discontinuous Transcription. *J. Virol.* **2001**, *75* (14), 6676−6681.
(26) Dai, W.; Zhang, B.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; 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.; Peng, J.; Jiang, H. Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020**, *368*, 1331.
(27) Ibrahim, M. A. A.; Abdelrahman, A. H. M.; Hegazy, M.-E. F. In-Silico Drug Repurposing and Molecular Dynamics Puzzled out Potential SARS-CoV-2 Main Protease Inhibitors. *J. Biomol. Struct. Dyn.* **2021**, *39* (15), 5756−5767.
(28) Daoud, S.; Alabed, S. J.; Dahabiyeh, L. A. Identification of Potential COVID-19 Main Protease Inhibitors Using Structure-Based Pharmacophore Approach, Molecular Docking and Repurposing Studies. *Acta Pharm.* **2021**, *71* (2), 163−174.
(29) Choudhary, M. I.; Shaikh, M.; Tul-Wahab, A.; Ur-Rahman, A. In Silico Identification of Potential Inhibitors of Key SARS-CoV-2 3CL Hydrolase (Mpro) via Molecular Docking, MMGBSA Predictive Binding Energy Calculations, and Molecular Dynamics Simulation. *PLoS One* **2020**, *15* (7), No. e0235030.
(30) Gupta, Y.; Maciorowski, D.; Zak, S. E.; Jones, K. A.; Kathayat, R. S.; Azizi, S.-A.; Mathur, R.; Pearce, C. M.; Ilc, D. J.; Husein, H.; Herbert, A. S.; Bharti, A.; Rathi, B.; Durvasula, R.; Becker, D. P.; Dickinson, B. C.; Dye, J. M.; Kempaiah, P. Bisindolylmaleimide IX: A Novel Anti-SARS-CoV2 Agent Targeting Viral Main Protease 3CLpro Demonstrated by Virtual Screening Pipeline and in-Vitro Validation Assays. *Methods* **2021**, *195*, 57−71.
(31) Gentile, F.; Oprea, T. I.; Tropsha, A.; Cherkasov, A. Surely You Are Joking, Mr Docking! *Chem. Soc. Rev.* **2023**, *52*, 872.
(32) Huang, C.; Shuai, H.; Qiao, J.; Hou, Y.; Zeng, R.; Xia, A.; Xie, L.; Fang, Z.; Li, Y.; Yoon, C.; Huang, Q.; Hu, B.; You, J.; Quan, B.; Zhao, X.; Guo, N.; Zhang, S.; Ma, R.; Zhang, J.; Wang, Y.; Yang, R.; Zhang, S.; Nan, J.; Xu, H.; Wang, F.; Lei, J.; Chu, H.; Yang, S. A New Generation Mpro Inhibitor with Potent Activity against SARS-CoV-2 Omicron Variants. *Signal Transduct Target Ther* **2023**, *8* (1), 128.
(33) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *31* (2), 455−461.
(34) 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, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582* (7811), 289−293.
(35) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10* (5), 449−461.
(36) Yang, S.-Y. Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discovery Today* **2010**, *15* (11−12), 444−450.
(37) Macchiagodena, M.; Pagliai, M.; Procacci, P. Identification of Potential Binders of the Main Protease 3CLpro of the COVID-19 via Structure-Based Ligand Design and Molecular Modeling. *Chem. Phys. Lett.* **2020**, *750*, No. 137489.
(38) Fier, P. S.; Hartwig, J. F. Synthesis and Late-Stage Functionalization of Complex Molecules through C-H Fluorination and Nucleophilic Aromatic Substitution. *J. Am. Chem. Soc.* **2014**, *136* (28), 10139−10147.
(39) White, M. C.; Zhao, J. Aliphatic C-H Oxidations for Late-Stage Functionalization. *J. Am. Chem. Soc.* **2018**, *140* (43), 13988−14009.
(40) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53* (7), 2719−2740.
(41) Faul, M. M.; Engler, T. A.; Sullivan, K. A.; Grutsch, J. L.; Clayton, M. T.; Martinelli, M. J.; Pawlak, J. M.; LeTourneau, M.; Coffey, D. S.; Pedersen, S. W.; Kolis, S. P.; Furness, K.; Malhotra, S.; Al-awar, R. S.; Ray, J. E. Synthetic Approaches to indolo[6,7-a]pyrrolo[3,4-C]carbazoles: Potent Cyclin D1/CDK4 Inhibitors. *J. Org. Chem.* **2004**, *69* (9), 2967−2975.
(42) Scalacci, N.; Brown, A. K.; Pavan, F. R.; Ribeiro, C. M.; Manetti, F.; Bhakta, S.; Maitra, A.; Smith, D. L.; Petricci, E.; Castagnolo, D. Synthesis and SAR Evaluation of Novel Thioridazine Derivatives Active against Drug-Resistant Tuberculosis. *Eur. J. Med. Chem.* **2017**, *127*, 147−158.
(43) Zajdel, P.; Kos, T.; Marciniec, K.; Satała, G.; Canale, V.; Kamiński, K.; Hołuj, M.; Lenda, T.; Koralewski, R.; Bednarski, M.; Nowiński, L.; Wójcikowski, J.; Daniel, W. A.; Nikiforuk, A.; Nalepa, I.; Chmielarz, P.; Kuśmierczyk, J.; Bojarski, A. J.; Popik, P. Novel Multi-Target Azinesulfonamides of Cyclic Amine Derivatives as Potential Antipsychotics with pro-Social and pro-Cognitive Effects. *Eur. J. Med. Chem.* **2018**, *145*, 790−804.
(44) Choy, K.-T.; Wong, A. Y.-L.; Kaewpreedee, P.; Sia, S. F.; Chen, D.; Hui, K. P. Y.; Chu, D. K. W.; Chan, M. C. W.; Cheung, P. P.-H.; Huang, X.; Peiris, M.; Yen, H.-L. Remdesivir, Lopinavir, Emetine, and Homoharringtonine Inhibit SARS-CoV-2 Replication in Vitro. *Antiviral Res.* **2020**, *178*, No. 104786.
(45) Valipour, M. Different Aspects of Emetine's Capabilities as a Highly Potent SARS-CoV-2 Inhibitor against COVID-19. *ACS Pharmacol Transl Sci.* **2022**, *5* (6), 387−399.
(46) Singh, S. K.; Sose, A. T.; Wang, F.; Bejagam, K. K.; Deshmukh, S. A. Data Driven Discovery of MOFs for Hydrogen Gas Adsorption. *J. Chem. Theory Comput.* **2023**, *19* (19), 6686−6703.

https://doi.org/10.1021/acs.jpclett.3c01749
J. Phys. Chem. Lett. 2023, 14, 9490−9499