



Accurate Prediction of Inhibitor Binding to HIV-1 Protease Using CANDOCK

Zackary Falls^{1*}, Jonathan Fine², Gaurav Chopra^{2,3,4,5,6*} and Ram Samudrala^{1*}

¹Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, United States, ²Department of Chemistry, Purdue University, West Lafayette, IN, United States, ³Purdue Institute for Drug Discovery, West Lafayette, IN, United States, ⁴Purdue Center for Cancer Research, West Lafayette, IN, United States, ⁵Purdue Institute for Inflammation, Immunology and Infectious Disease, West Lafayette, IN, United States, ⁶Purdue Institute for Integrative Neuroscience, West Lafayette, IN, United States

OPEN ACCESS

Edited by:

Renjith Thomas,
St Berchmans College, Mahatma
Gandhi University, India

Reviewed by:

Kenny Lischer,
University of Indonesia, Indonesia
Sixue Zhang,
Southern Research Institute,
United States

*Correspondence:

Zackary Falls
zmfalls@buffalo.edu
Gaurav Chopra
gchopra@purdue.edu
Ram Samudrala
ram@compbio.org

Specialty section:

This article was submitted to
Theoretical and Computational
Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 14 September 2021

Accepted: 25 November 2021

Published: 17 January 2022

Citation:

Falls Z, Fine J, Chopra G and
Samudrala R (2022) Accurate
Prediction of Inhibitor Binding to HIV-1
Protease Using CANDOCK.
Front. Chem. 9:775513.
doi: 10.3389/fchem.2021.775513

The human immunodeficiency virus 1 (HIV-1) protease is an important target for treating HIV infection. Our goal was to benchmark a novel molecular docking protocol and determine its effectiveness as a therapeutic repurposing tool by predicting inhibitor potency to this target. To accomplish this, we predicted the relative binding scores of various inhibitors of the protease using CANDOCK, a hierarchical fragment-based docking protocol with a knowledge-based scoring function. We first used a set of 30 HIV-1 protease complexes as an initial benchmark to optimize the parameters for CANDOCK. We then compared the results from CANDOCK to two other popular molecular docking protocols Autodock Vina and Smina. Our results showed that CANDOCK is superior to both of these protocols in terms of correlating predicted binding scores to experimental binding affinities with a Pearson coefficient of 0.62 compared to 0.48 and 0.49 for Vina and Smina, respectively. We further leveraged the Database of Useful Decoys: Enhanced (DUD-E) HIV protease set to ascertain the effectiveness of each protocol in discriminating active versus decoy ligands for proteases. CANDOCK again displayed better efficacy over the other commonly used molecular docking protocols with area under the receiver operating characteristic curve (AUROC) of 0.94 compared to 0.71 and 0.74 for Vina and Smina. These findings support the utility of CANDOCK to help discover novel therapeutics that effectively inhibit HIV-1 and possibly other retroviral proteases.

Keywords: molecular docking, inhibitor prediction, protein–ligand interaction, HIV-1 protease, knowledge-based force field, CANDOCK

1 INTRODUCTION

Inhibition of the Human immunodeficiency virus (HIV) protease blocks viral maturation and replication, making inhibitors of this vital enzyme an important class of drugs for the treatment of HIV infection (Wlodawer et al., 1989; Wlodawer and Vondrasek, 1998). The introduction of HIV protease inhibitors reduced the mortality rate of infected patients in the US significantly, from about 50 thousand deaths per year in 1995, down to 20 thousand by 2000 (Centers for Disease Control and Prevention, 2001a; Centers for Disease Control and Prevention, 2001b; Quinn, 2008). However, escape mutations within the viral protease have resulted in HIV strains that are resistant to these inhibitors, presenting a challenge to identify which protease inhibitors are effective against specific mutants or discover and design broad spectrum inhibitors (Ohtaka and Freire, 2005). For these

reasons, accurate prediction of inhibitor efficacy against protease mutants that arise during infection has been a focus of HIV drug discovery for decades.

Previous efforts to predict inhibitor activity against human immunodeficiency virus 1 (HIV-1) proteases include rule-based methods (Shafer et al., 1999; Kantor et al., 2001), support vector machine (SVM) models (Beerenwinkel et al., 2002; Cai et al., 2003), chemical shape and features (Yadav et al., 2012; Pandit et al., 2006; DesJarlais and Dixon, 1994; Wei et al., 2015), various docking protocols (Chang et al., 2007, 2010), and molecular dynamics (MD) simulations (Rick et al., 1998; Wang and Kollman, 2001; Wang K. et al., 2004; Jenwitheesuk and Samudrala, 2003; Jenwitheesuk et al., 2004; Jenwitheesuk and Samudrala, 2005a; Jenwitheesuk and Samudrala, 2005b). These different approaches have displayed varying results. Chang et al. compared the efficacy of Autodock 4 and Autodock Vina in predicting active versus inactive compounds using the National Cancer Institute Diversity II compound sets and showed that both protocols perform better than random [area under the curve (AUC) of 0.69 and 0.68, respectively] on this diverse compound set (Huang et al., 2006; Chang et al., 2010). Pandit et al. generated a pharmacophore model using Molecular Operating Environment (MOE) software to predict activity from a set of known protease inhibitors and non-inhibitors, correctly identifying 65 of the 75 protease inhibitors and incorrectly classifying 11 out of 75 non-inhibitors. When volume exclusion was incorporated into the model, the authors were able to decrease the number of false positives to 5 out of 75 while reducing the true positives to 60 out of 75. (Pandit et al., 2006). These results display the tradeoff between sensitivity and specificity and the limitations of this model.

The problem is more difficult in a *de novo* molecular docking scenario, when the x-ray diffraction structure of a protease mutant (or homology model) is docked to an inhibitor and the necessary rotations and transformations are calculated without *a priori* knowledge (Razzaghi-Asl et al., 2015). In contrast, approaches that use x-ray diffraction structure poses for the protease and inhibitor, combined with the features of the compound and/or MD, to predict binding affinities have met with some success (Jenwitheesuk et al., 2004; Jenwitheesuk and Samudrala, 2005a; Jenwitheesuk and Samudrala, 2005b). Jenwitheesuk and Samudrala obtained a peak correlation of 0.87 between predicted binding energies and experimental binding affinities through the use of MD simulations (Jenwitheesuk and Samudrala, 2003). These MD approaches have the advantages of allowing for the exploration of the inhibitor within the binding site and using scoring functions that are sensitive to amino acid mutations in the protease structure. However, MD requires an accurate 3D complex structure of the bound pose between ligand and protein as a starting point, thus limiting this method by requiring solved or modeled structure(s) and/or a docking protocol. Jenwitheesuk and Samudrala showed that calculated versus experimental binding correlation was 0.38 with the docking protocol Autodock alone, illustrating the benefit MD provides (Jenwitheesuk and Samudrala, 2003). Leveraging docking protocols with more sophisticated and robust scoring

TABLE 1 | List of HIV-1 protease structures extracted from Protein DataBase (PDB) with experimentally determined inhibition constants (**K_i**) extracted from BindingMOAD.

PDB ID	Resolution (Å)	Ligand ID	Experimental K _i (M)
1a8g	2.50	ZZ4	7.40e-09
1aaq	2.50	PSI	3.00e-09
1aid	2.20	THK	1.50e-05
1ajv	2.00	NMB	1.91e-08
1ajx	2.00	AH1	1.22e-08
1g2k	1.95	NM1	1.10e-08
1g35	1.80	AHF	7.30e-09
1gno	2.30	U0E	2.00e-08
1hbv	2.30	GAN	4.30e-07
1heg	2.20	PSI	1.80e-08
1hih	2.20	C20	9.00e-09
1hiv	2.00	1ZK	1.00e-09
1hos	2.30	PHP	2.80e-09
1hps	2.30	RUN	6.00e-10
1hpv	1.90	478	6.00e-10
1hpx	2.00	KNI	5.50e-12
1hvh	1.80	Q82	1.10e-08
1hvi	1.80	A77	8.40e-11
1hvj	2.00	A78	4.00e-12
1hvk	1.80	A79	7.70e-11
1hvl	1.80	A76	1.00e-09
1hvr	1.80	XK2	3.10e-10
1hvs	2.25	A77	5.00e-11
1pro	1.80	A88	5.00e-12
1qbr	1.80	XV6	2.70e-11
1qbs	1.80	DMP	3.40e-10
1qbt	2.10	146	2.40e-11
1qbu	1.80	846	5.80e-11
1sbg	2.30	IM1	1.80e-08
1sdt	1.30	MK1	5.40e-10

The PDB ID and resolution, in angstroms, are provided as well as the corresponding co-crystallized inhibitor ligand ID and inhibition constant (M) for the protease-inhibitor complex.

functions can be used to identify the pose of the inhibitors with respect to the protease as well as accurately predict the corresponding binding scores.

Here we used a hierarchical fragment-based docking based dynamics protocol implemented in the CANDOCK protocol (Fine and Chopra, 2018; Fine et al., 2020), in conjunction with its all-atom knowledge-based scoring function (Bernard and Samudrala, 2009), to predict the binding scores of inhibitors to HIV-1 protease. The knowledge-based scoring function calculates and optimizes atomic interactions in the binding pocket to sample biologically relevant ligand conformations giving us an ability to identify specific interactions (beyond hydrogen bonding or pi-stacking, etc.) in protein binding sites. We first optimized the parameters for the knowledge-based scoring function used in CANDOCK using a set of HIV-1 protease-inhibitor complex structures with known binding affinities from BindingMOAD (Smith et al., 2019). We then used CANDOCK to predict actives versus decoys from the Directory of Useful Decoys Enhanced (DUD-E) HIV protease subset (Mysinger et al., 2012) to affirm the discriminatory abilities of the improved protocol and compared it to two popular molecular docking methods, AutoDock Vina and Smina. Our research showed a strong correlation between experimental

TABLE 2 | Compounds extracted from PubChem with experimentally determined inhibition activity against HIV-1 protease.

PubChem compound ID	HIV-1 protease activity
CID480440	Active
CID480447	Active
CID480550	Active
CID514961	Active
CID480441	Active
CID480469	Active
CID514958	Active
CID10509626	Inactive
CID478338	Inactive
CID49796249	Inactive
CID66162	Inactive
CID10747313	Inactive
CID478339	Inactive
CID49796254	Inactive

The PubChem compound identifier and the corresponding activity is provided for each of the 14 compounds.

binding affinities and predicted binding scores for CANDOCK which resulted in a 0.62 Pearson coefficient compared to the 0.48 and 0.49 for Vina and Smina, respectively. In addition, the performance of CANDOCK on the DUD-E HIV protease set exceeded Vina and Smina with an area under the receiver operating characteristic curve (AUROC) of 0.94 for CANDOCK compared to 0.71 for Vina and 0.74 for Smina. These results demonstrate the predictive power of CANDOCK for the specific case of assessing inhibitor potency against HIV-1 protease.

2 MATERIALS AND METHODS

2.1 Curation of Human Immunodeficiency Virus 1 Protease–Inhibitor Sets for Benchmarking

We compiled a set of 30 HIV-1 protease–inhibitor complex structures (Table 1) extracted from the Protein Data Bank (PDB), all of which were solved using x-ray diffraction with a resolution of 2.5 Å or lower and have experimentally determined binding affinities between the protease and given inhibitor (Berman et al., 2000). To use these structures in our docking simulations, we had to separate and process the protease and inhibitor in each complex. First, the proteases were processed using biopython and OpenMM to remove any co-crystallized ligands so the HIV-1 protease dimer was all that remained to be used as the receptor in each molecular docking simulation (Cock et al., 2009; Eastman et al., 2017). The inhibitor was also separated from the corresponding complex in each case and converted to a Mol2 file format using OpenBabel for compatibility with the CANDOCK protocol (O’Boyle et al., 2011, 2008). Both the protease and inhibitor files for each complex were also converted to the PDBQT file format using AutoDocktools for compatibility with the AutoDock Vina protocol (Morris et al., 2009). The binding sites for each protease were defined using the

coordinates of the native ligand in the PDB structure. Experimentally observed inhibition constants, K_i , for the bound ligands in all 30 complexes were obtained from BindingMOAD (Smith et al., 2019).

In addition to the 30 HIV-1 protease–inhibitor complexes, we also extracted a set of active and decoy compounds for the HIV protease from the DUD-E which consisted of 1,395 actives and 36,278 decoys against the macromolecule dimer (Mysinger et al., 2012).

Lastly, a set of 14 compounds were extracted from PubChem, seven of which were experimentally determined to be active HIV-1 protease inhibitors and the other seven were determined to be inactive compounds. The list of compounds and their corresponding activity against HIV-1 protease is given in Table 2.

2.2 Molecular Docking Protocols

Three different molecular docking protocols: CANDOCK (Fine and Chopra, 2018; Fine et al., 2020), Autodock Vina (Trott and Olson, 2010), and Smina (Koes et al., 2013) were used herein to predict the binding affinity between the HIV-1 protease inhibitors and the protease macromolecule binding sites.

2.2.1 CANDOCK

CANDOCK is a hierarchical fragment-based docking with dynamics protocol to “grow” the ligand in the binding pocket by leveraging a generalized knowledge-based statistical scoring function to identify docked poses (Bernard and Samudrala, 2009; Fine and Chopra, 2018). The fragment-based approach identifies and breaks rotatable bonds in the ligand to reduce the ligand down to rigid subunits, which are then individually docked in the binding pocket and relinked to build the complete original ligand in the best pose. This method enables comprehensive search of both the binding pocket and the conformational space of ligand. The scoring parameters for the knowledge-based scoring function were varied during this study and are discussed in Section 2.3. The protease–inhibitor binding site was defined as a sum of centroids, with 4.5 Å radii, at each atom of the known bound ligand for each complex. We also used the Generalized Amber ForceField (GAFF) implemented in CANDOCK as a control for the knowledge-based scoring function.

2.2.2 AutoDock Vina

AutoDock Vina, referred to here as Vina, is a well-known molecular docking protocol that uses a physics-based forcefield, similar to X-Score, that is tuned on the experimental data in PDBBind (Wang et al., 2002; Trott and Olson, 2010; Su et al., 2018). In this study, Vina was used with default parameters with the following exceptions: the exhaustiveness and num_modes parameters set to 8 and 9, respectively. The binding box center was placed at the geometric center of the known bound ligand for each corresponding crystal structure. The length of all sides of the binding box were defined as two times the radius of gyration of the compound plus 9.0 Å, to ensure a large enough search space while simultaneously mimicking the binding site centroids used in CANDOCK.

2.2.3 Smina

Smina is a forked version of Vina with expanded functionality and enabled user-defined scoring functions (Koes et al., 2013). In addition, Koes et al. modified the existing Vina potential by optimizing additional energetic terms found in the original code that were not used; i.e., the coefficients were set to 0. The potential terms included an electrostatic term, a desolvation term, and a non-hydrophobic contact term, among others that were all parameterized by training on the CSAR (Community Structure-Activity Resource) 2010 dataset (Dunbar Jr et al., 2011). The parameters used for Vina, described above, were also used here. The binding box was also defined identically.

2.3 Selector and Ranker

The CANDOCK protocol generates hundreds to thousands of binding modes for a protein–ligand pair that may then be scored. To obtain a single score, the protocol orders all of these modes by score using a knowledge-based forcefield (KBF) and chooses the top result as the best binding pose and corresponding binding score. We refer to this scoring and sorting as the *selector*. Once the top binding pose is chosen, we rescore this pose using the same or different variation of the KBF, by modifying one or more of the terms. The new score is then used in the calculation of the correlation between known and predicted binding affinities. We call the parameter set used for rescoring the *ranker* (Fine et al., 2020).

The KBF is an atomic level forcefield that is generalized to all intermolecular binding interactions, e.g., protein–small molecule, protein–DNA, etc. (Bernard and Samudrala, 2009). The equation to calculate the interaction score between two molecules, Eq. 1, is analogous to the net potential of mean force:

$$S(\{r_{ab}^{ij}\}) = - \sum \ln \frac{P(r_{ab}^{ij}|C)}{P(r^{ij})} \quad (1)$$

This equation calculates the probability of two intermolecular atoms being within a given distance ($P(r_{ab}^{ij}|C)$) with respect to the probability of any two atoms being within the same distance ($P(r^{ij})$), where r_{ab}^{ij} is the distance between atom i of type a and j of type b . For this function the distances, r , are discretized into distinct spherical shells. The KBF has four terms that affect the function and are optimized herein. The four terms are *functional*, *reference*, *composition*, and *cutoff*.

The probability distributions $P(r_{ab}^{ij}|C)$ and $P(r^{ij})$ are generated by one of two *functionals*: a normalized frequency distribution function (f), or a radial distribution function (r). The former assesses how many atoms are within a given shell, whereas the latter divides the number of atoms in the shell by the volume of the spherical shell.

The *reference* probability is calculated one of two ways: cumulative (c) or mean (m), which are the sum over all atom pairs or averaged over the number of atom pairs, respectively.

The last two terms, *composition* and *cutoff*, refer to the atom type pairs and the maximum distance (r) considered for calculations. The two options for *composition* are complete (c), which enables the use of all atom type pairs, and reduced (r), which limits the summation to only atom types a and b found

within the given intermolecular complex. *Cutoff* can range from 4 to 15 Å.

The probabilities $P(r_{ab}^{ij}|C)$ and $P(r^{ij})$ were generated using the experimental data found in the Cambridge Structural Database (CSD) for both the protein–ligand and protein–DNA interactions (Allen, 2002).

The “default” parameter set for CANDOCK was defined as *rmr6–rmr6*, which equates to radial *functional*, mean *reference*, reduced *composition*, and 6 Å *cutoff* for both the selector and ranker.

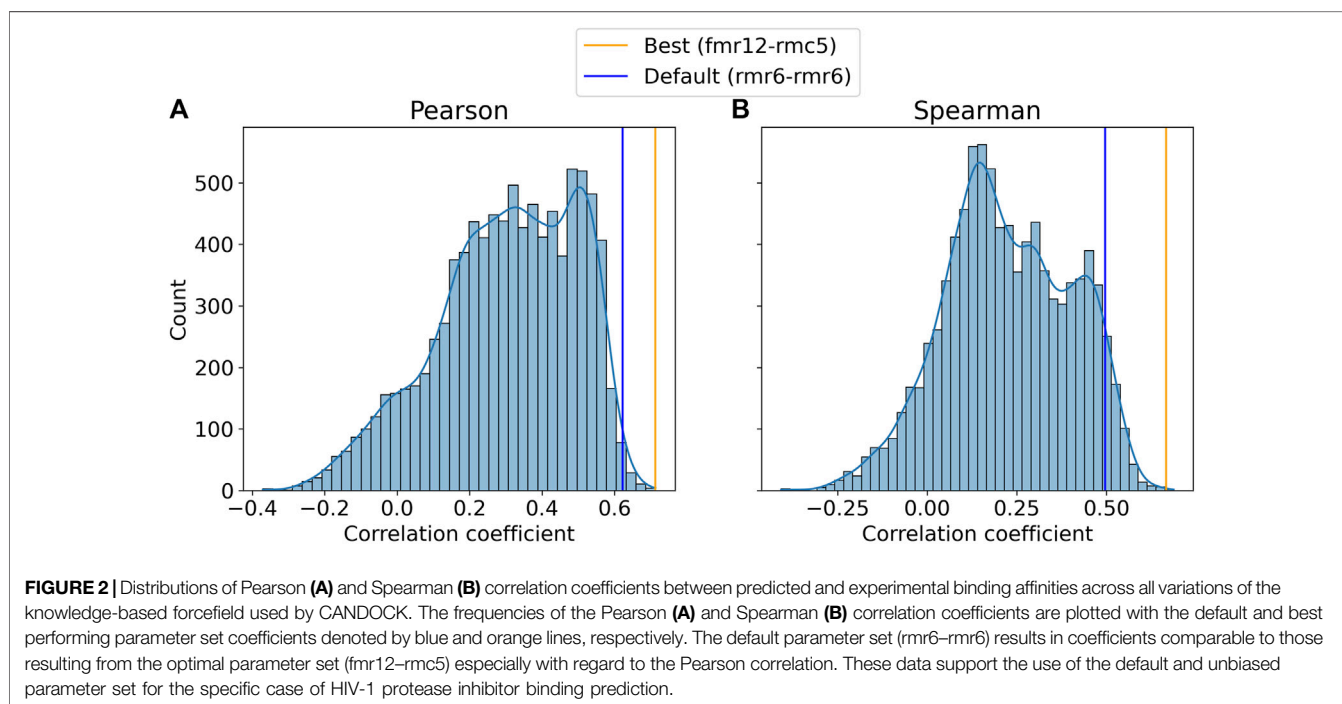
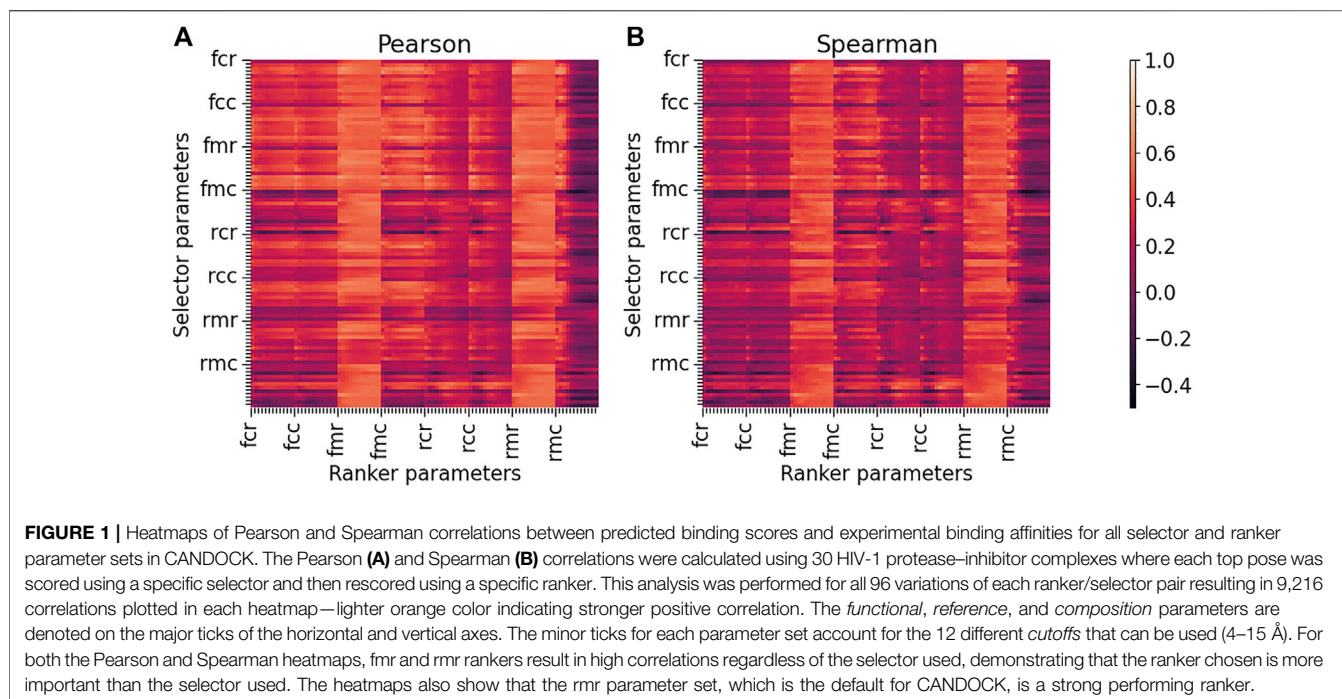
3 RESULTS AND DISCUSSION

3.1 Comprehensive Analysis of Parameters for the CANDOCK Knowledge-Based Scoring Function

The CANDOCK protocol uses a generalized statistical scoring function for scoring molecular interactions. This KBF is implemented with four different parameters: *functional*, *reference*, *composition*, and *cutoff* (Section 2.3). Each of the first three parameters have two options, and cutoff can range from 4 to 15 Å, resulting in 96 variations of the KBF. We generated the top poses for each of the 30 HIV-1 protease–inhibitor complexes using the 96 KBF variations for both the selector and ranker to ascertain the optimal combination.

We calculated top poses and corresponding binding scores for the 30 HIV-1 protease–inhibitor complexes using all 96 variations of the KBF (selector) and subsequently rescored the top pose for all complexes with each of the 96 KBF variations (ranker), resulting in 9,216 different selector–ranker combinations. For each selector–ranker pair, we calculated the Pearson and Spearman correlation between the known K_i from BindingMOAD and the predicted binding scores for all 30 protease–inhibitor pairs. The correlations calculated showed us how accurately each selector–ranker parameter set predicted binding scores when compared to the experimental binding affinities. We populated a heatmap with all of the calculated correlations presented in Figure 1. This visualization showed us that *rmr* and *fmr* with all cutoffs were very high-performing rankers, where variation of the selector shows negligible effect on the correlations. These results were interesting from a high-level perspective by showing the efficacy of the *fmr* and *rmr* rankers, but we wanted to assess specific parameter sets to determine which would be the most accurate for the HIV-1 protease–inhibitor complexes.

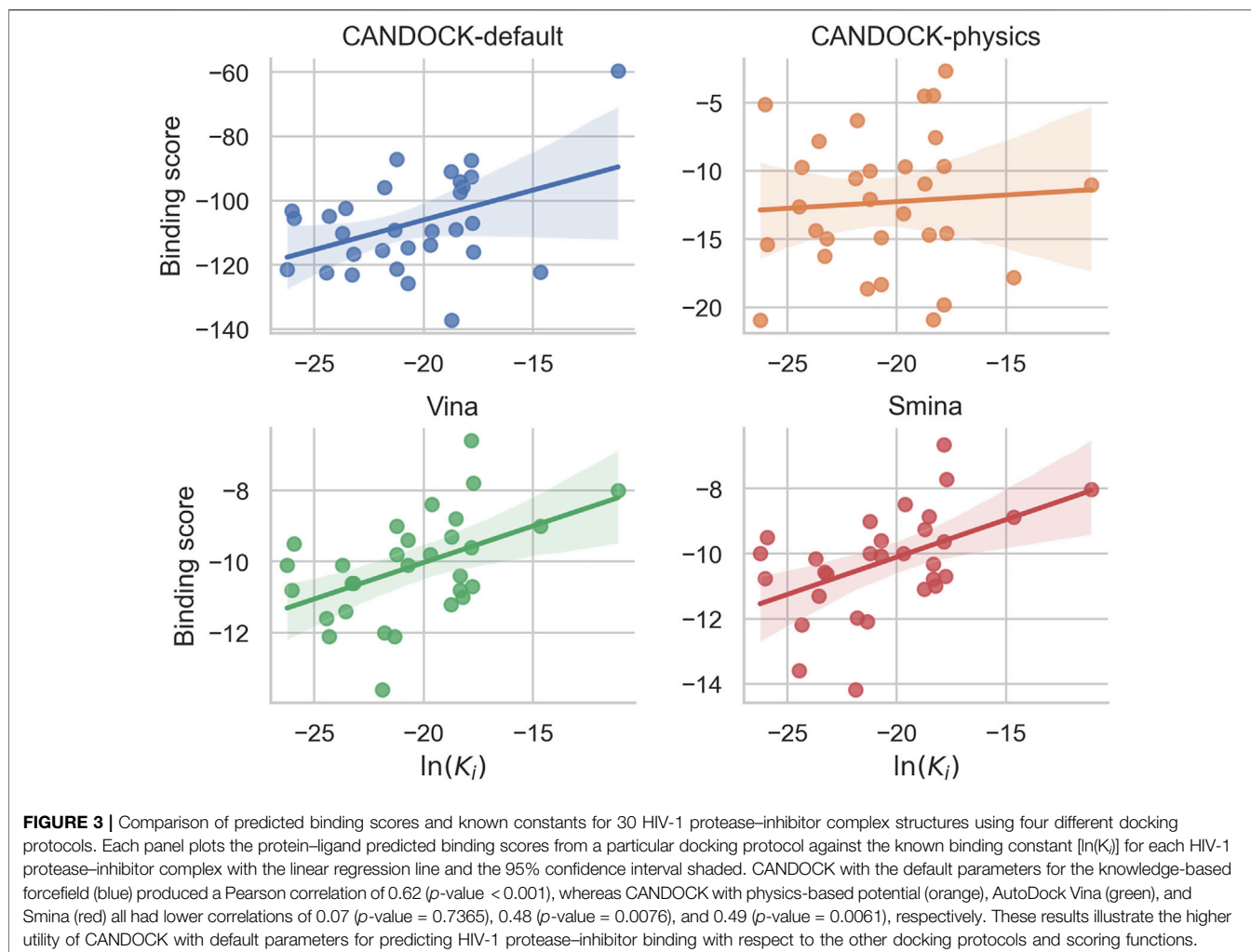
The CANDOCK protocol was previously parameterized on the complete CASF-2016, containing 285 protein–ligand complexes across 57 proteins (Fine and Chopra, 2018; Su et al., 2018; Fine et al., 2020). Results from this analysis determined *rmr6* as selector and *rmc15* as ranker were the best performing parameter sets, and varying the selector did not have a major impact on the resulting correlations between binding affinity and binding score. The authors also commented on the results for the subset of HIV protease–inhibitor complexes



in CASF-2016 stating that rmc15–rmr6, the reverse of the previously stated selector–ranker, was the best performing set. We used the knowledge obtained in this earlier study along with our own analysis of the heatmap of all correlations (Figure 1) to justify rmr6–rmr6 as our “default” selector–ranker because it is the default values set in the CANDOCK program, the rmr6 ranker was previously shown to be the best for the HIV

protease set from CASF-2016, and the ranker parameters are more important than the selector parameters for the accuracy of the method.

We analyzed our HIV-1 protease results by comparing the results of the default parameter set, rmr6–rmr6, to the best performing parameter set, fmr12–rmc5. Figure 2 shows that the best performing parameter set, fmr12–rmc5, yielded



Pearson and Spearman correlations of 0.71 (p -value < 0.0001) and 0.67 (p -value < 0.0001), respectively, for the 30 HIV-1 protease-inhibitor complexes. The best performing parameter set marginally outperforms the default parameter set for CANDOCK, rmr6-rmr6, with Pearson and Spearman correlations of 0.62 (p -value < 0.001) and 0.50 (p -value < 0.01), respectively.

Both CANDOCK parameter sets were in the top 1% of Pearson scores, suggesting that the previously established default parameter set for assessing the ability of CANDOCK to accurately calculate the relative binding strength of HIV-1 protease-inhibitor complexes performs near optimally (Figure 2). This led us to continue further analysis with this unbiased parameter set rather than using the best one, thereby eliminating the risk of any overtraining.

3.2 Comparison to Other Docking Methods and Forcefields

We compared CANDOCK to two other well-established molecular docking protocols to determine their relative utility (Figure 3). For this comparison, we used protocols based on the

Vina and Smina software (Morris et al., 2009; Trott and Olson, 2010). In addition, we also ran the predictions made by CANDOCK using a physics-based forcefield (CANDOCK-physics), GAFF, as a control (Wang et al., 2004a; Wang et al., 2006).

Figure 3 shows that Vina, which uses a scoring function involving knowledge-based and empirical components, generated Pearson and Spearman correlations of 0.48 and 0.50, respectively. Smina, a modified Vina with a custom scoring potential, performed similarly with correlations of 0.49 and 0.49 for Pearson and Spearman, respectively. Recall that the respective correlations were 0.62 and 0.50 for CANDOCK using the default KBF parameter set and 0.71 and 0.67 using the best performing one. Lastly, CANDOCK-physics was a control that resulted in 0.06 Pearson and 0.09 Spearman correlations. The physics-based potentials used to predict protein-ligand interactions showed little to no correlation to known binding affinities for the HIV-1 protease-inhibitor set.

These results combined indicate that CANDOCK-default is able to accurately predict the relative binding affinities with greater confidence than other docking protocols.

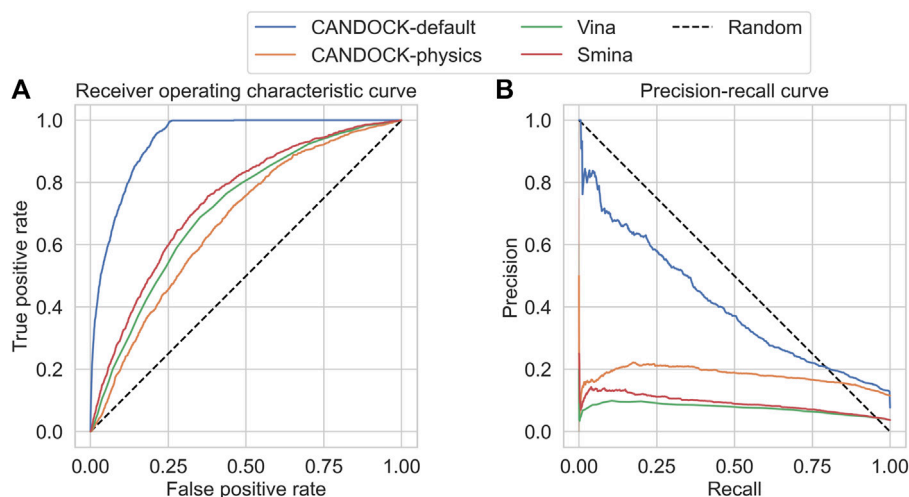


FIGURE 4 | Receiver operating characteristic (ROC) and precision–recall (PR) curves for the HIV protease DUD-E set using four different docking protocols. The resulting area under the ROC (AUROC) curve values on the **(A)** are 0.94 for CANDOCK with the default KBF parameters (blue), 0.67 for CANDOCK with physics-based potential (orange), 0.71 for Vina (green), and 0.74 for Smina (red). The resulting area under the precision–recall curves (AUPRC) values on the **(B)** are 0.41 for CANDOCK-default, 0.18 for CANDOCK-physics, 0.08 for Vina, and 0.09 for Smina. These results provide further evidence that CANDOCK with its default KBF parameters outperforms the other docking protocols and scoring functions for prediction of binding affinity and activity of HIV-1 protease inhibitors.

3.3 Discrimination of Active Versus Decoy Human Immunodeficiency Virus 1 Protease Inhibitors

We next assessed the discriminatory ability of CANDOCK to effectively identify active inhibitors of HIV-1 protease over decoys. We again compared the results of CANDOCK with default KBF parameters to Vina, Smina, and CANDOCK-physics. To accomplish this, we ran docking simulations using all four protocols on a set of active and decoy HIV-1 protease inhibitors from DUD-E (Mysinger et al., 2012).

For each protocol, the receiver operating characteristic (ROC) curve was generated based upon the sorted binding scores the protocol calculated for each active/decoy–protein pair; additionally, the AUROC was calculated for each protocol (Figure 4). The resulting AUROC values for all protocols were above 0.5, meaning they all performed better than random. CANDOCK with default KBF parameters performed the best with an AUROC of 0.94. The other three methods had much lower AUROC values with CANDOCK-physics at 0.67, Vina at 0.71, and Smina at 0.74. While all protocols performed well, CANDOCK-default displayed its superiority for binding score prediction.

We plotted the precision–recall curves for each protocol and calculated the area under these curves (AUPRC) to more thoroughly assess the protocols in discriminating active from decoy inhibitors (Figure 4). All four protocols scored under 0.5, which is common when there are imbalanced class sizes like in the DUD-E set. CANDOCK-default still outperformed the other three protocols with a AUPRC of 0.41. The remaining three protocols, CANDOCK-physics, Vina, and Smina, had AUPRCs of 0.18, 0.08, and 0.09, respectively.

These results show that in addition to CANDOCK predicting binding scores that correlate well with known binding affinities,

the protocol is also capable of effectively choosing active inhibitors over decoys.

3.4 Discrimination of Known Active Versus Inactive Human Immunodeficiency Virus 1 Protease Inhibitors

To further investigate the efficacy of the CANDOCK protocol, we assessed its discrimination capability between known active HIV-1 protease inhibitors and known inactive compounds. This benchmark is similar to the DUD-E set; however, the inactive compounds are experimentally confirmed, as opposed to decoys that are generated based on chemical similarity to the active compounds. This provides a more robust test for CANDOCK to be compared to the other methods in their ability to identify HIV-1 protease inhibitors.

For each protocol, we compared the resulting binding scores for each active and inactive compound (Figure 5) and visualized their separation based upon the strength of binding (active compounds should be in the lower left of the plot and the inactives should be in the upper right). CANDOCK-default and Vina distinguish the actives versus inactives clearly, while they group more closely for CANDOCK-physics and Smina.

To more explicitly determine the discriminatory ability, we subsequently generated the ROC curve based on the sorted binding scores the protocol calculated for each active/inactive–protein pair; additionally, the AUROC was calculated for each protocol (Figure 6). CANDOCK with default KBF parameters and Vina both performed the best with an AUROC of 1.00. The other two methods had lower AUROC values with CANDOCK-physics at 0.98 and Smina at 0.92. While all protocols performed well, CANDOCK-default and Vina

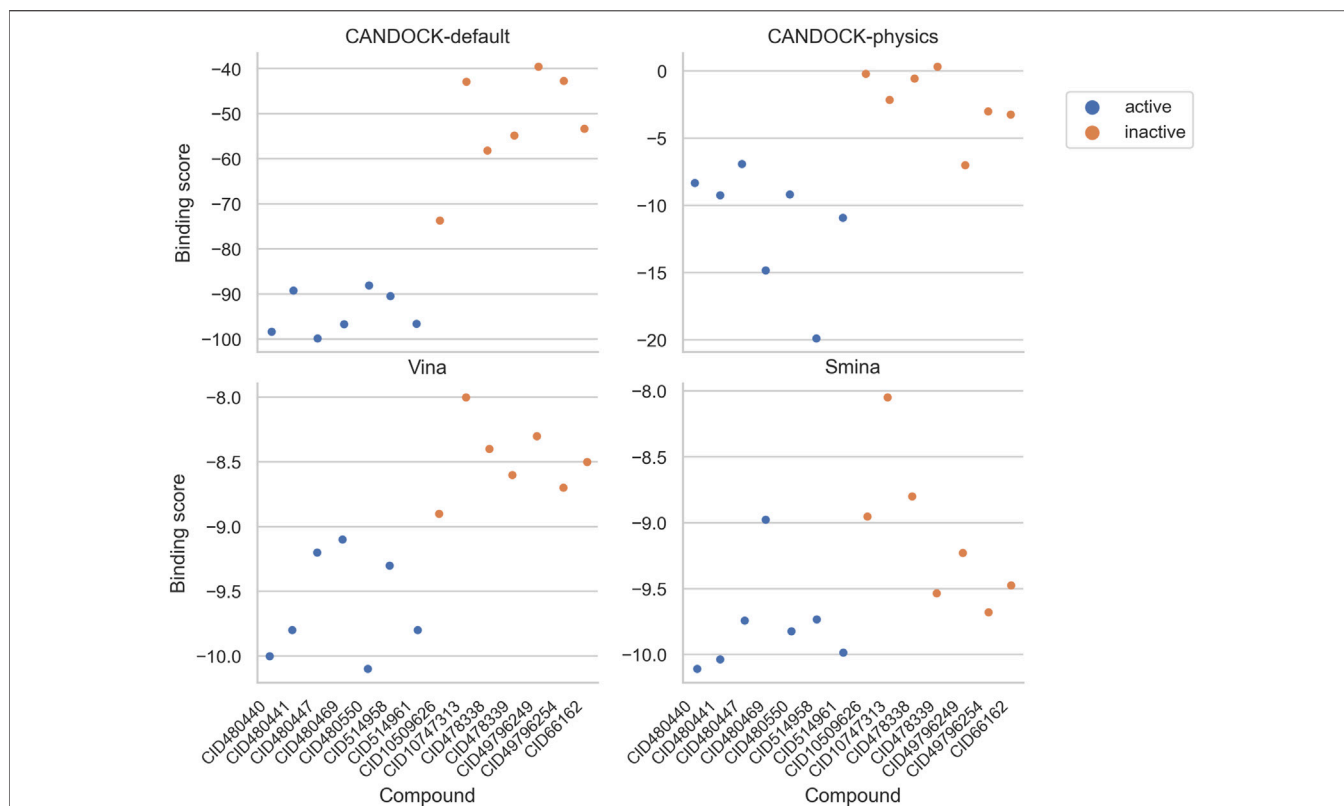


FIGURE 5 | Predicted binding scores for HIV-1 protease inhibitors and inactive compounds using four different docking protocols. Each panel plots the protein–ligand predicted binding scores from a particular docking protocol for each of the seven active (blue) and seven inactive (orange) compounds. These results depict a distinct separation in binding scores for actives versus inactives using CANDOCK with default parameters and Vina.

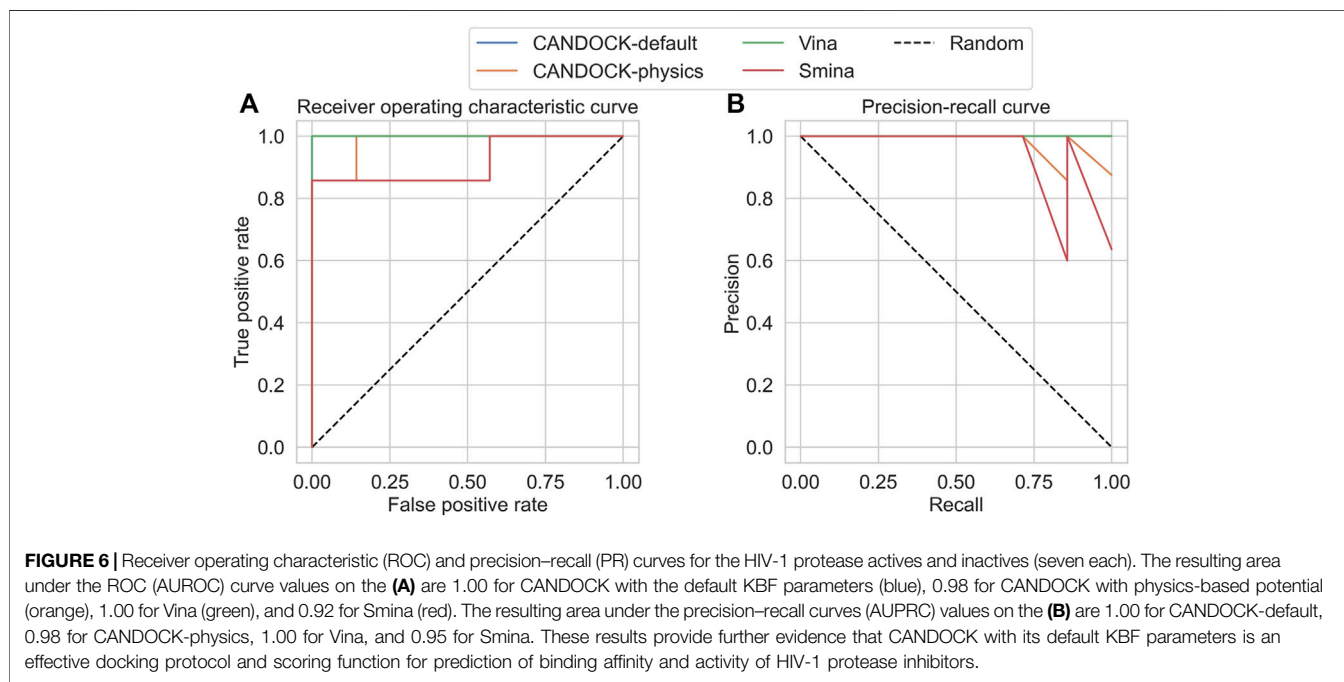
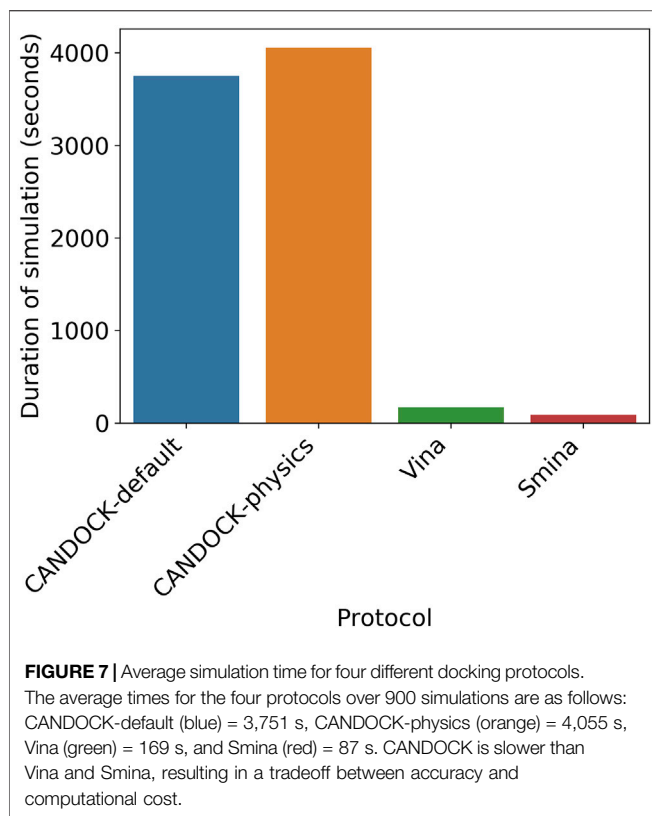


FIGURE 6 | Receiver operating characteristic (ROC) and precision–recall (PR) curves for the HIV-1 protease actives and inactives (seven each). The resulting area under the ROC (AUROC) curve values on the (A) are 1.00 for CANDOCK with the default KBF parameters (blue), 0.98 for CANDOCK with physics-based potential (orange), 1.00 for Vina (green), and 0.92 for Smina (red). The resulting area under the precision–recall curves (AUPRC) values on the (B) are 1.00 for CANDOCK-default, 0.98 for CANDOCK-physics, 1.00 for Vina, and 0.95 for Smina. These results provide further evidence that CANDOCK with its default KBF parameters is an effective docking protocol and scoring function for prediction of binding affinity and activity of HIV-1 protease inhibitors.



displayed comparable efficacy for active versus inactive discrimination.

We plotted the precision–recall curves for each protocol and calculated the AUPRC to more thoroughly assess the protocols in discriminating actives from inactives (Figure 6). CANDOCK-default and Vina still outperformed the other two protocols with a AUPRC of 1.00, compared to AUPRC of 0.98 for CANDOCK-physics and 0.95 for Smina.

Overall, CANDOCK-default is effective at predicting active inhibitors for HIV-1 protease from a set of actives and inactives. Vina showed comparable results for this test, and both slightly outperformed Smina and CANDOCK-physics.

3.5 Comparison of Docking Simulation Times

As another point of comparison for all four protocols, we assessed the average time it takes to generate the top pose for the HIV-1 protease–inhibitor complex. We ran the 30 protease–inhibitor complexes 30 times resulting in 900 simulations for each protocol. We then averaged the time of simulation, in seconds, for all 900 simulations and calculated the averages for each protocol (Figure 7). The results show that CANDOCK takes much longer to complete a single simulation, despite the scoring function with CANDOCK-default averaging 3,751 s and CANDOCK-physics averaging 4,055 s. Vina and Smina are comparable to each other and much faster than CANDOCK, with an average time for simulation of 169 and 87 s, respectively.

Balancing the computational cost with the accuracy of the protocol is important to consider, especially if the compound sets being tested are very large (>1,000) or the available computing power is limited.

3.6 Limitations and Future Work

The size of the set of HIV-1 protease–inhibitor complexes used for the parameterization and correlation calculations is limited to only 30 members. A larger set would enable better parameterization and assessment of the protocols by providing more variability in the chemical space of the ligands. Moreover, a larger set of known active and inactive compounds would allow for a much more rigorous comparison of these methods in their ability to discriminate active inhibitors from inactive compounds.

The datasets used herein all focused on the wildtype HIV-1 protease, but the prevalence of mutations that confer drug resistance is of great concern in HIV treatment. Future work will assess the sensitivity of CANDOCK and the other protocols to inhibitors of HIV-1 protease mutations. Since CANDOCK has already been shown to be effective in predicting and designing specific inhibitors, demonstrating the efficacy of the CANDOCK protocol in the prediction of binding affinities to any protease mutant will greatly aid in future drug design efforts for effective, broad-spectrum protease inhibitors to replace the cocktails that are currently used (Larocque et al., 2017; Ma et al., 2017). Moreover, if our results were generalizable to predict accurate binding between HIV-1 protease mutants and their inhibitors, it would allow for a precision medicine approach to protease inhibitor efficacy prediction (Jenwitheesuk and Samudrala, 2003; Jenwitheesuk et al., 2004; Wang et al., 2004b; Jenwitheesuk and Samudrala, 2005a;2005b; Jenwitheesuk et al., 2005; Jenwitheesuk and Samudrala, 2007; Jenwitheesuk et al., 2008).

Leveraging some of our recently developed programs, based on machine learning and graph neural networks, we can iteratively select synthetically feasible bioactive protease inhibitors based on bioactivity data and CANDOCK-generated pose of molecules (Majumder et al., 2018; Wijewardhane et al., 2020). Exploration of CANDOCK efficacy on other HIV-1 targets, such as reverse transcriptase, would enable proteomic-based drug discovery, which we have shown to be useful for drug repurposing, and could lead to more potent HIV-1 therapeutics (Chopra et al., 2016; Chopra and Samudrala, 2016; Hernandez-Perez et al., 2017; Majumder et al., 2017; Fine et al., 2019; Mangione et al., 2020; Robertson et al., 2020).

4 CONCLUSION

We evaluated four different docking protocols for their effectiveness in predicting HIV-1 protease–inhibitor binding affinities. We assessed these protocols by correlation to known binding affinities and by ability to discriminate between known active and decoy inhibitors. The results from both these computational experiments showed that the CANDOCK protocol with its all-atom knowledge-based forcefield was superior to these other protocols.

Overall, we show that CANDOCK accurately predicts the relative binding affinities for HIV-1 protease inhibitors when compared to and greatly outperforms popular publicly available molecular docking protocols. The efficacy demonstrated by CANDOCK in our study indicates that it will be very useful for the repurposing, discovery, and design of novel HIV-1 protease inhibitors.

DATA AVAILABILITY STATEMENT

The data generated for this study can be found at http://compbio.buffalo.edu/data/in_hiv_candock.

AUTHOR CONTRIBUTIONS

ZF and RS conceived the research design and methods for this study. ZF implemented the methods, executed the research design, and drafted the manuscript. RS, GC, and JF edited and refined the manuscript. RS supervised the overall study. JF, GC, and RS are the co-creators of the CANDOCK protocol. All authors have read and agreed to the published version of the manuscript.

REFERENCES

- Allen, F. H. (2002). The Cambridge Structural Database: a Quarter of a Million crystal Structures and Rising. *Acta Crystallogr. Sect B* 58, 380–388. doi:10.1107/s0108768102003890
- Beerenwinkel, N., Schmidt, B., Walter, H., Kaiser, R., Lengauer, T., Hoffmann, D., et al. (2002). Diversity and Complexity of Hiv-1 Drug Resistance: a Bioinformatics Approach to Predicting Phenotype from Genotype. *Proc. Natl. Acad. Sci.* 99, 8271–8276. doi:10.1073/pnas.112177799
- Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Bernard, B., and Samudrala, R. (2009). A Generalized Knowledge-Based Discriminatory Function for Biomolecular Interactions. *Proteins: Struct. Funct. Bioinf.* 76, 115–128. doi:10.1002/prot.22323
- Cai, C., Han, L., Ji, Z. L., Chen, X., and Chen, Y. Z. (2003). Svm-prot: Web-Based Support Vector Machine Software for Functional Classification of a Protein from its Primary Sequence. *Nucleic Acids Res.* 31, 3692–3697. doi:10.1093/nar/kg600
- Centers for Disease Control and Prevention (2001a). HIV and AIDS--United States, 1981–2000. *MMWR Morb Mortal Wkly Rep.* 50, 430–434.
- Centers for Disease Control and Prevention (2001b). Hiv/aids Surveillance—General Epidemiology. Available at: <https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2001-vol-13-2.pdf>.
- Chang, M. W., Lindstrom, W., Olson, A. J., and Belew, R. K. (2007). Analysis of Hiv Wild-type and Mutant Structures via In Silico Docking against Diverse Ligand Libraries. *J. Chem. Inf. Model.* 47, 1258–1262. doi:10.1021/ci700044s
- Chang, M. W., Ayeni, C., Breuer, S., and Torbett, B. E. (2010). Virtual Screening for Hiv Protease Inhibitors: a Comparison of Autodock 4 and Vina. *PLoS one* 5, e11955. doi:10.1371/journal.pone.0011955
- Chopra, G., and Samudrala, R. (2016). Exploring Polypharmacology in Drug Discovery and Repurposing Using the cando Platform. *Curr. Pharm. Des.* 22, 3109–3123. doi:10.2174/1381612822666160325121943
- Chopra, G., Kaushik, S., Elkin, P. L., and Samudrala, R. (2016). Combating Ebola with Repurposed Therapeutics Using the cando Platform. *Molecules* 21, 1537. doi:10.3390/molecules21121537
- Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: Freely Available python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* 25, 1422–1423. doi:10.1093/bioinformatics/btp163
- DesJarlais, R. L., and Dixon, J. S. (1994). A Shape-And Chemistry-Based Docking Method and its Use in the Design of Hiv-1 Protease Inhibitors. *J. computer-aided Mol. Des.* 8, 231–242. doi:10.1007/bf00126742
- Dunbar, J. B., Jr, Smith, R. D., Yang, C.-Y., Ung, P. M.-U., Lexa, K. W., Khazanov, N. A., et al. (2011). Csar Benchmark Exercise of 2010: Selection of the Protein–Ligand Complexes. *J. Chem. Inf. Model.* 51, 2036–2046. doi:10.1021/ci200082t
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). Openmm 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* 13, e1005659. doi:10.1371/journal.pcbi.1005659
- Fine, J. A., and Chopra, G. (2018). Candock: Conformational Entropy Driven Analytics for Class-specific Proteome-wide Docking. *Biophys. J.* 114, 57a. doi:10.1016/j.bpj.2017.11.363
- Fine, J., Lackner, R., Samudrala, R., and Chopra, G. (2019). Computational Chemoproteomics to Understand the Role of Selected Psychoactives in Treating Mental Health Indications. *Sci. Rep.* 9, 1–15. doi:10.1016/b978-0-12-812962-3.00001-0
- Fine, J., Konc, J., Samudrala, R., and Chopra, G. (2020). Candock: Chemical Atomic Network-Based Hierarchical Flexible Docking Algorithm Using Generalized Statistical Potentials. *J. Chem. Inf. Model.* 60, 1509–1527. doi:10.1021/acs.jcim.9b00686
- Hernandez-Perez, M., Chopra, G., Fine, J., Conteh, A. M., Anderson, R. M., Linnemann, A. K., et al. (2017). Inhibition of 12/15-lipoxygenase Protects against β -cell Oxidative Stress and Glycemic Deterioration in Mouse Models of Type 1 Diabetes. *Diabetes* 66, 2875–2887. doi:10.2337/db17-0215
- Huang, N., Shoichet, B. K., and Irwin, J. J. (2006). Benchmarking Sets for Molecular Docking. *J. Med. Chem.* 49, 6789–6801. doi:10.1021/jm0608356
- Jenwitheesuk, E., and Samudrala, R. (2003). Improved Prediction of Hiv-1 Protease-Inhibitor Binding Energies by Molecular Dynamics Simulations. *BMC Struct. Biol.* 3, 2. doi:10.1186/1472-6807-3-2
- Jenwitheesuk, E., and Samudrala, R. (2005a). Prediction of Hiv-1 Protease Inhibitor Resistance Using a Protein-Inhibitor Flexible Docking Approach. *Antivir. Ther.* 10, 157–166. doi:10.1097/01.aids.0000162343.96674.4c
- Jenwitheesuk, E., and Samudrala, R. (2005b). Virtual Screening of Hiv-1 Protease Inhibitors against Human Cytomegalovirus Protease Using Docking and Molecular Dynamics. *Aids* 19, 529–531. doi:10.1097/01.aids.0000162343.96674.4c

FUNDING

This work was supported in part by a 2010 NIH Director's Pioneer Award (1DP1OD006779), NIH Clinical and Translational Sciences Award (UL1TR001412), NLM T15 Award (T15LM012495), NCI/VA BD-STEP Fellowship in Big Data Sciences, and startup funds from the Department of Biomedical Informatics at the University at Buffalo. This work was also supported by 2019 NIH NCATS ASPIRE Challenge Awards and 2020 NIH NCATS ASPIRE Reduction-to-Practice Awards. Additional support in part by an NCATS Clinical and Translational Sciences Award from the Indiana Clinical and Translational Sciences Institute (UL1TR002529) and the Purdue University Center for Cancer Research NIH grant P30 (CA023168) is acknowledged.

ACKNOWLEDGMENTS

We are grateful for the computational support provided by the University at Buffalo's Center for Computational Research.

- Jenwitheesuk, E., and Samudrala, R. (2007). Identification of Potential Hiv-1 Targets of Minocycline. *Bioinformatics* 23, 2797–2799. doi:10.1093/bioinformatics/btm424
- Jenwitheesuk, E., Wang, K., Mittler, J. E., and Samudrala, R. (2004). Improved Accuracy of Hiv-1 Genotypic Susceptibility Interpretation Using a Consensus Approach. *Aids* 18, 1858–1859. doi:10.1097/00002030-200409030-00020
- Jenwitheesuk, E., Wang, K., Mittler, J. E., and Samudrala, R. (2005). Pirspred: a Web Server for Reliable Hiv-1 Protein-Inhibitor Resistance/susceptibility Prediction. *Trends Microbiol.* 13, 150–151. doi:10.1016/j.tim.2005.02.003
- Jenwitheesuk, E., Horst, J. A., Rivas, K. L., Van Voorhis, W. C., and Samudrala, R. (2008). Novel Paradigms for Drug Discovery: Computational Multitarget Screening. *Trends Pharmacol. Sci.* 29, 62–71. doi:10.1016/j.tips.2007.11.007
- Kantor, R., Machekano, R., Gonzales, M. J., Dupnik, K., Schapiro, J. M., and Shafer, R. W. (2001). Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database: an Expanded Data Model Integrating Natural Language Text and Sequence Analysis Programs. *Nucleic Acids Res.* 29, 296–299. doi:10.1093/nar/29.1.296
- Koes, D. R., Baumgartner, M. P., and Camacho, C. J. (2013). Lessons Learned in Empirical Scoring with Smina from the Csar 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* 53, 1893–1904. doi:10.1021/ci300604z
- Larocque, E., Naganna, N., Ma, X., Opoku-Temeng, C., Carter-Cooper, B., Chopra, G., et al. (2017). Aminoisoquinoline Benzamides, Flt3 and Src-Family Kinase Inhibitors, Potently Inhibit Proliferation of Acute Myeloid Leukemia Cell Lines. *Future Med. Chem.* 9, 1213–1225. doi:10.4155/fmc-2017-0067
- Ma, X., Zhou, J., Wang, C., Carter-Cooper, B., Yang, F., Larocque, E., et al. (2017). Identification of New Flt3 Inhibitors that Potently Inhibit Aml Cell Lines via an Azo Click-It/staple-It Approach. *ACS Med. Chem. Lett.* 8, 492–497. doi:10.1021/acsmchemlett.6b00468
- Majumder, J., Lantz, T., Fine, J., and Chopra, G. (2017). Abstract A35: Drug Repurposing for Castration Resistant Prostate Cancer Based on Disease-Disease Relationships. *Cancer Res.* doi:10.1158/1538-7445.Transcontrol16-A35
- Majumder, J., Fine, J. A., Lantz, T. C., Conder, C. J., and Chopra, G. (2018). Abstract Lb-076: Cancer Cell Specific Lethality by Degrading Specific Protein Target Network Identified Using a Chemical Screening Based Machine Learning Method. *Cancer Res.* doi:10.1158/1538-7445.AM2018-LB-076
- Mangione, W., Falls, Z., Melendy, T., Chopra, G., and Samudrala, R. (2020). Shotgun Drug Repurposing Biotechnology to Tackle Epidemics and Pandemics. *Drug Discov. Today* 25, 1126. doi:10.1016/j.drudis.2020.05.002
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). Autodock4 and Autodocktools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* 30, 2785–2791. doi:10.1002/jcc.21256
- Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of Useful Decoys, Enhanced (Dud-e): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* 55, 6582–6594. doi:10.1021/jm300687e
- O'Boyle, N. M., Morley, C., and Hutchison, G. R. (2008). Pybel: a Python Wrapper for the Openbabel Cheminformatics Toolkit. *Chem. Cent. J.* 2, 1–7. doi:10.1186/1752-153X-2-5
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: An Open Chemical Toolbox. *J. Cheminf.* 3, 33. doi:10.1186/1758-2946-3-33
- Ohtaka, H., and Freire, E. (2005). Adaptive Inhibitors of the Hiv-1 Protease. *Prog. Biophys. Mol. Biol.* 88, 193–208. doi:10.1016/j.pbiomolbio.2004.07.005
- Pandit, D., So, S.-S., and Sun, H. (2006). Enhancing Specificity and Sensitivity of Pharmacophore-Based Virtual Screening by Incorporating Chemical and Shape Features - a Case Study of Hiv Protease Inhibitors. *J. Chem. Inf. Model.* 46, 1236–1244. doi:10.1021/ci050511a
- Quinn, T. C. (2008). Hiv Epidemiology and the Effects of Antiviral Therapy on Long-Term Consequences. *AIDS (London, England)* 22, S7. doi:10.1097/01.aids.0000327510.68503.e8
- Razzaghi-Asl, N., Sepehri, S., Ebadi, A., Miri, R., and Shahabipour, S. (2015). Effect of Biomolecular Conformation on Docking Simulation: a Case Study on a Potent Hiv-1 Protease Inhibitor. *Iran J. Pharm. Res.* 14, 785–802.
- Rick, S. W., Topol, I. A., Burt, S. K., and Erickson, J. W. (1998). Molecular Mechanisms of Resistance: Free Energy Calculations of Mutation Effects on Inhibitor Binding to Hiv-1 Protease. *Protein Sci.* 7, 1750–1756. doi:10.1002/pro.556007809
- Robertson, M. A., Padgett, L. R., Fine, J. A., Chopra, G., and Mastracci, T. L. (2020). Targeting Polyamine Biosynthesis to Stimulate Beta Cell Regeneration in Zebrafish. *Islets* 12, 99–107. doi:10.1080/19382014.2020.1791530
- Shafer, R. W., Stevenson, D., and Chan, B. (1999). Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database. *Nucleic Acids Res.* 27, 348–352. doi:10.1093/nar/27.1.348
- Smith, R. D., Clark, J. J., Ahmed, A., Orban, Z. J., Dunbar, J. B., Jr, and Carlson, H. A. (2019). Updates to Binding Moad (Mother of All Databases): Polypharmacology Tools and Their Utility in Drug Repurposing. *J. Mol. Biol.* 431, 2423–2433. doi:10.1016/j.jmb.2019.05.024
- Su, M., Yang, Q., Du, Y., Feng, G., Liu, Z., Li, Y., et al. (2018). Comparative Assessment of Scoring Functions: The Casf-2016 Update. *J. Chem. Inf. Model.* 59, 895–913. doi:10.1021/acs.jcim.8b00545
- Trott, O., and Olson, A. J. (2010). Autodock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- Wang, W., and Kollman, P. A. (2001). Computational Study of Protein Specificity: the Molecular Basis of Hiv-1 Protease Drug Resistance. *Proc. Natl. Acad. Sci.* 98, 14937–14942. doi:10.1073/pnas.251265598
- Wang, R., Lai, L., and Wang, S. (2002). Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput.-Aided Mol. Des.* 16, 11–26. doi:10.1023/a:1016357811882
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004a). Development and Testing of a General Amber Force Field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/jcc.20035
- Wang, K., Jenwitheesuk, E., Samudrala, R., and Mittler, J. E. (2004b). Simple Linear Model Provides Highly Accurate Genotypic Predictions of Hiv-1 Drug Resistance. *Antivir. Ther.* 9, 343–352.
- Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Model.* 25, 247–260. doi:10.1016/j.jmgm.2005.12.005
- Wei, Y., Li, J., Chen, Z., Wang, F., Huang, W., Hong, Z., et al. (2015). Multistage Virtual Screening and Identification of Novel Hiv-1 Protease Inhibitors by Integrating Svm, Shape, Pharmacophore and Docking Methods. *Eur. J. Med. Chem.* 101, 409–418. doi:10.1016/j.ejmech.2015.06.054
- Wijewardhane, P. R., Jethava, K. P., Fine, J. A., and Chopra, G. (2020). Combined Molecular Graph Neural Network and Structural Docking Selects Potent Programmable Cell Death Protein 1/programmable Death-Ligand 1 (Pd-1/pd-1) Small Molecule Inhibitors. *ChemRxiv.* doi:10.26434/chemrxiv.12083907.v1
- Wlodawer, A., and Vondrasek, J. (1998). Inhibitors of Hiv-1 Protease: a Major success of Structure-Assisted Drug Design. *Annu. Rev. Biophys. Biomol. Struct.* 27, 249–284. doi:10.1146/annurev.biophys.27.1.249
- Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., et al. (1989). Conserved Folding in Retroviral Proteases: crystal Structure of a Synthetic Hiv-1 Protease. *Science* 245, 616–621. doi:10.1126/science.2548279
- Yadav, D., Paliwal, S., Yadav, R., Pal, M., and Pandey, A. (2012). Identification of Novel Hiv 1-protease Inhibitors: Application of Ligand and Structure Based Pharmacophore Mapping and Virtual Screening. *PLoS one* 7, e48942. doi:10.1371/journal.pone.0048942

Conflict of Interest: GC is the Director of Merck-Purdue Center for Measurement Science funded by Merck Sharp Dohme Corp., a subsidiary of Merck Co. Inc., Kenilworth, NJ, United States.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Falls, Fine, Chopra and Samudrala. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.