Journal of Molecular Graphics and Modelling 127 (2024) 108669

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

Fragment databases from screened ligands for drug discovery (FDSL-DD)

Jerica Wilson [a], Bahrad A. Sokhansanj [b], Wei Chuen Chong [b], Rohan Chandraghatgi [b], Gail L. Rosen [b,**], Hai-Feng Ji [a,*]

[a] Department of Chemistry, Drexel University, Philadelphia, PA, 19104, USA
[b] Ecological and Evolutionary Signal-processing and Informatics Lab, Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, 19104, USA

ARTICLE INFO

Keywords:
Drug design
Drug discovery
Fragment-based drug design (FBDD)
Fragment based design
Structure based design

ABSTRACT

Fragment-based drug design (FBDD) is one major drug discovery method employed in computer-aided drug discovery. Due to its inherent limitations, this process experiences long processing times and limited success rates. Here we present a new Fragment Databases from Screened Ligands Drug Design method (FDSL-DD) that intelligently incorporates information about fragment characteristics into a fragment-based design approach to the drug development process. The initial step of the FDSL-DD is the creation of a fragment database from a library of docked, drug-like ligands for a specific target, which deviates from the traditional *in silico* FBDD strategy, incorporating structure-based design screening techniques to combine the advantages of both approaches. Three different protein targets have been tested in this study to demonstrate the potential of the created fragment library and FDSL-DD. Utilizing the FDSL-DD led to an increase in binding affinity for each protein target. The most substantial increase was exhibited by the ligand designed for TIPE2, with a 3.6 kcalmol^{-1} difference between the top ligand from the FDSL-DD and top ligand from the high throughput virtual screening (HTVS). Using drug-like ligands in the initial HTVS allows for a greater search of chemical space, with higher efficiency in fragments selection, less grid boxes, and potentially identifying more interactions.

1. Introduction

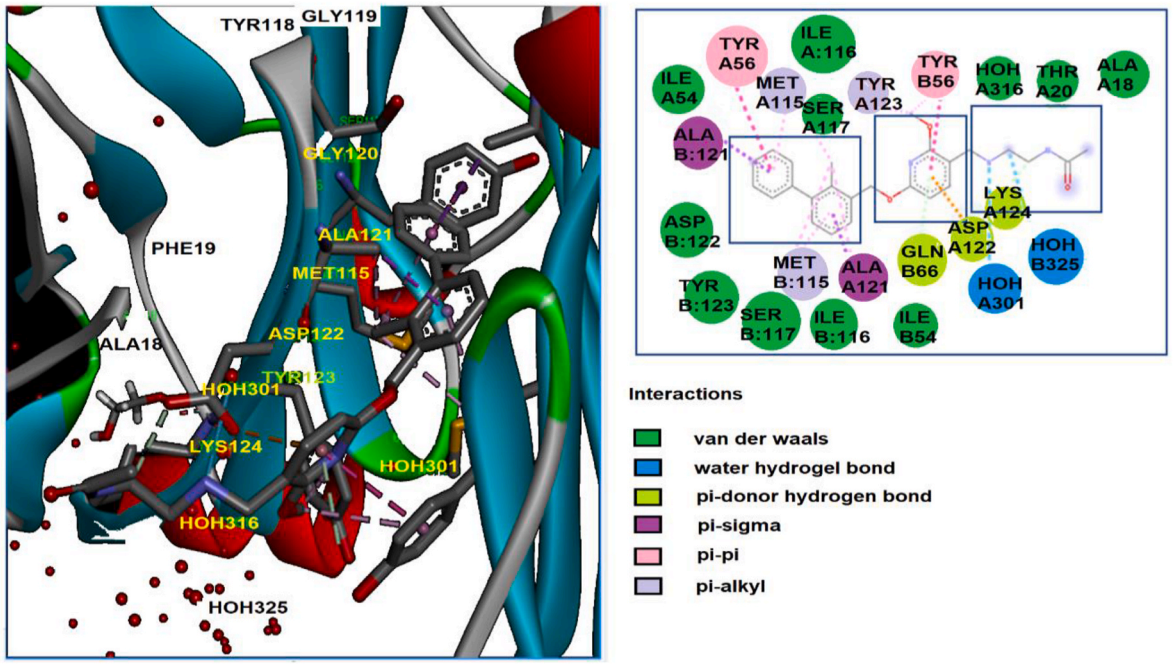
Computer-aided methods have been widely used in drug discovery processes [1–6]. Fragment-based drug design (FBDD) utilizes small molecules or fragments (molecular weight <300 gmol^{-1}), to design a lead compound. Identified fragments can be grown, linked, or merged into a more potent lead molecule [7–11]. However, fragment libraries are very large, and the number of fragment combinations and their orientations in the generation of novel ligands is combinatorically explosive [12]. As a result, FBDD remains a challenge, since the space for identifying effective drug candidates is still very large, and finding candidates that are both feasible (drug-like) and have high binding affinity to the target is a difficult task.

We report a new fragment-based method: creating a fragment database from a large, already docked, ligand screening library for a specific target, in which fragments are associated with information from the parent ligand. We term the method Fragments from Ligands Drug Design (FDSL-DD) as shown in Scheme 1. At a high level, a large number of "drug like" ligands are screened with computational docking software to 1) obtain the predicted binding affinity between each of the ligands and the protein targets and 2) where and how (i.e., what chemical bonds at what atoms) the ligand binds with the protein. After these screening and profiling steps, the ligands are computationally "fragmentized" (virtually broken up into fragments) (Scheme 1). A database is then created which includes, for each fragment, summary statistics for the binding affinity of parent ligands and protein-ligand bond profiling data from the screening step. We may then utilize the resulting fragment database to design drug candidates *in silico* (Fig. 1).

To demonstrate the potential of the created fragment library and FDSL-DD, three different protein targets have been chosen, each with substantially different chemical and structural characteristics. The first, tumor necrosis factor alpha induced protein 8-like 2 (TIPE2), is a transport protein that can induce leukocyte polarization, sustaining chronic inflammation and ultimately supporting tumorigenesis [13,14]. Inhibition of TIPE2 would provide a therapeutic option for solid tumor cancers. The second, RelA, a protein that detects amino acid starvation

* Corresponding author.
** Corresponding author.
E-mail addresses: glr26@drexel.edu (G.L. Rosen), hj56@drexel.edu (H.-F. Ji).

https://doi.org/10.1016/j.jmgm.2023.108669
Received 14 September 2023; Received in revised form 4 November 2023; Accepted 7 November 2023
Available online 22 November 2023



Scheme 1. Schematic presentation of the Fragments from Ligands Drug Design method (FDSL-DD).

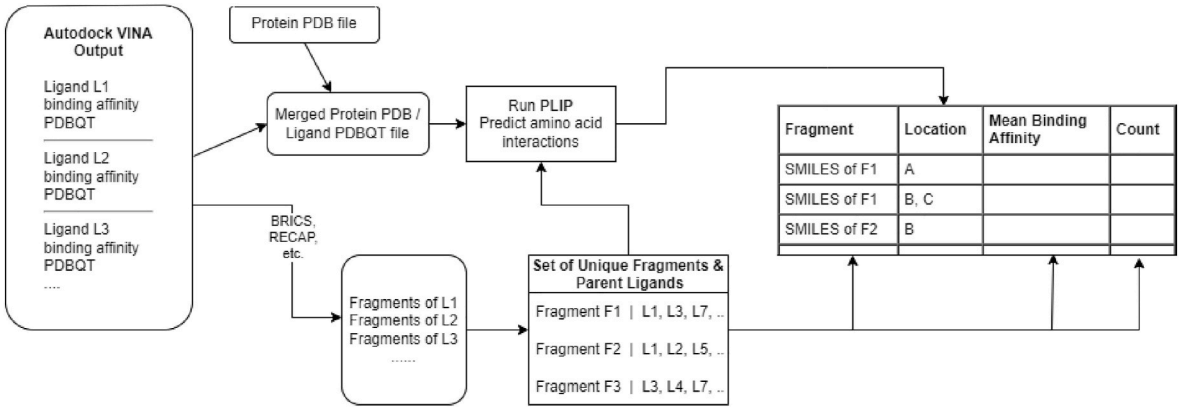


Fig. 1. Database Creation FDSL-DD Flow Diagram. The input to the FDSL-DD is the output files of ligand screening from a modeling software, such as AutoDock Vina or Schrodinger. The output files include the PDBQT representation of a ligand structure in its predicted docking conformations with the protein, along with predicted binding affinities. The minimum binding affinity solution is selected. The ligand PDBQT (Protein Data Bank, Partial Charge (Q), & Atom Type (T)) and protein PDB files are merged and then fed into PLIP (Protein-Ligand Interaction Profiler) which predicts ligand atom-amino acid bonds. The ligand PDBQT files are also converted into a SMILES representation, which is then fed into a computational fragmentation tool, such as BRICS or RECAP. The fragments from each ligand are then collected. The FDSL-DD then outputs a database or table of fragments (with SMILES representation) at different locations in the binding pocket determined by the amino acids to which the fragments bind in a parent ligand. There may be more than one entry for a fragment if it is found to bind in different locations in different parent ligands.

activating the stringent response in bacteria which leads to persister cell formation. Persister cells can withstand 1000 times the antibiotic concentrations of their normal cell counterparts [15], so inhibit RelA and antibiotics can be used to eradicate the bacteria, and mostly importantly bacterial biofilms. The final protein utilized in this study, is the receptor binding domain (RBD) of the S1 subunit of the spike protein (S-protein) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The S-protein RBD binds to human angiotensin-converting enzyme (ACE-2), facilitating viral entry [16]. It is noteworthy that our selection of the three proteins was not arbitrary; rather, we deliberately chose three ongoing projects in our lab. Over the years, both our research group and several others have failed in identifying potent inhibitors for these proteins, except for the S-protein, using conventional drug discovery methods. However, with the introduction of the innovative FDSL-DD approach, we have achieved promising and unprecedented results that were previously unattainable.

Table 1
Most frequently occurring fragments in the top 10 % of highest binding ligands for TIPE2, RelA and the S-protein.

Fragment SMILES	Protein	Fragment	Domain	Mean	Mode	Median	Count
[5*]N1CCC2(CC1)C(=O)N([10*])C(=O)N2[10*]	TIPE2	T2F1	A44	−7.72	−7.90	−7.70	81
[10*]N1C(=O)[N]C([13*])([13*])C1=O	RelA	RAF11	N187, K195, Y319	−8.33	−8.00	−8.30	584
[5*]N1CCc2nnc([14*])n2CC1	Sprotein	SPF21	F497, Y505	−7.22	−7.00	−7.10	328

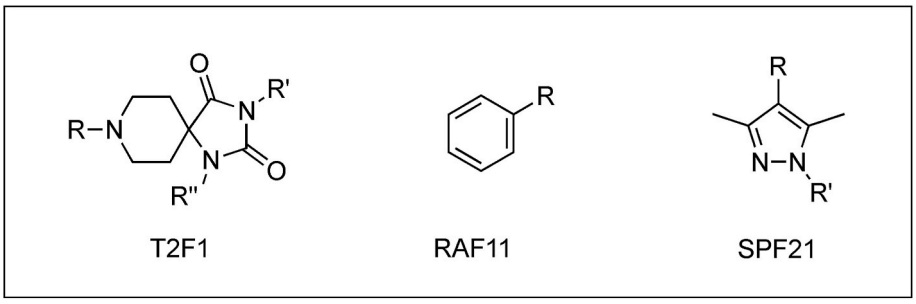


Fig. 2. Most frequently occurring fragments for each protein.

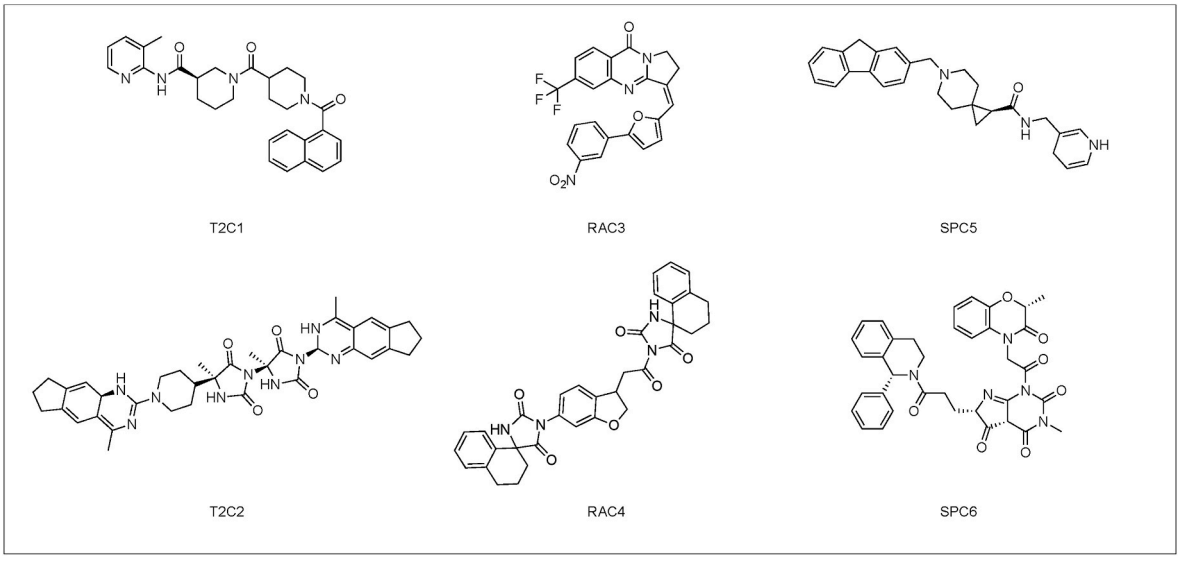


Fig. 3. Structures of top 3 ligands from the HTVS method (top) and top 3 ligands from FDSL-DD (bottom).

Table 2
Comparison of top binding ligands from high throughput virtual screening (HTVS) and top binding ligand from FDSL-DD for each protein target. Included are binding affinities from high throughput method and from the FDSL-DD and computed ADME properties.

Method	TIPE2		RelA		S-Protein	
	HTVS	FDSL-DD	HTVS	FDSL-DD	HTVS	FDSL-DD
Ligand Name	T2C1	T2C2	RAC3	RAC4	SPC5	SPC6
Binding Affinity ($kcalmol^{-1}$)	−9.8	−13.4	−10.2	−12.4	−9.1	−12.0
Molecular Weight ($gmol^{-1}$)	484.59	673.76	453.37	590.63	426.57	647.69
Solubility (ESOL)	−4.94	−4.63	−5.68	−5.84	−4.72	−5.63
Partition Coefficient (MlogP)	2.91	2.8	2.78	2.97	3.59	2.08
TPSA	82.61	150.84	93.85	125.12	44.37	136.97
Log K_p (cm/s)	−6.74	−9.55	−5.88	−7.31	−6.25	−7.96
GI Absorption	High	Low	Low	High	High	High
BBB permeant	No	No	No	No	Yes	No

Table 3
Comparison of druglikeness. The Lipinski filter; MW ≤ 500, MLogP ≤4.15. The Ghose filter; 160≤ MW ≤ 480, −0.4 ≤ WLogP ≤5.6, 40 ≤ MR ≤ 130, 20 ≤ atoms ≤70. The Veber filter; TPSA ≤140. And the Egan filter; WLogP ≤5.88 and TPSA ≤131.6. The Muegge filter; 200 ≤ MW ≤ 600, −2 ≤ XLogP ≤5, TPSA ≤150.

Method	TIPE2		RelA		S-Protein	
	HTvS	FDSL-DD	HTvS	FDSL-DD	HTvS	FDSL-DD
Ligand Name	T2C1	T2C2	RAC3	RAC4	SPC5	SPC6
Lipinski	Yes	No	Yes	Yes	Yes	No
Ghose	No	No	No	No	No	No
Veber	Yes	No	Yes	Yes	Yes	Yes
Egan	Yes	No	No	Yes	Yes	No
Muegge	Yes	No	Yes	No	Yes	No

2. Results

Fragment databases were created for each protein. Table 1 shows the most frequently occurring fragments in the top 10 % of highest binding ligands for each protein. For TIPE2, 1,3,8-triazaspiro [4.5] decane-2,4-dione, or fragment T2F1 (Fig. 2), appears 81 times from ligands with a mean binding affinity of −7.72 $kcalmol^{-1}$, and possible connections at the 1, 3, and 8 position. For RelA, benzene is the most frequently occurring fragment, with a count of 584 in ligands with a mean binding affinity of −8.33 $kcalmol^{-1}$. For the S-protein, 3,5-dimethyl-1*H*-pyrazole, or SPF21, with possible connection at positions 1 and 4, was the most frequent fragment, occurring 328 times.

From these generated fragment libraries superior binding ligands were constructed for each protein, with the highest binding pictured in Fig. 3. For comparison, Table 2 shows the highest binding ligands from the HTVS of the Enamine library with each protein against the highest binding ligand produced from the FDSL-DD. Select predicted ADME properties that are utilized in the training environment have also been noted. An increase in binding affinity is seen for each constructed ligand. The most substantial increase was exhibited by the ligand designed for

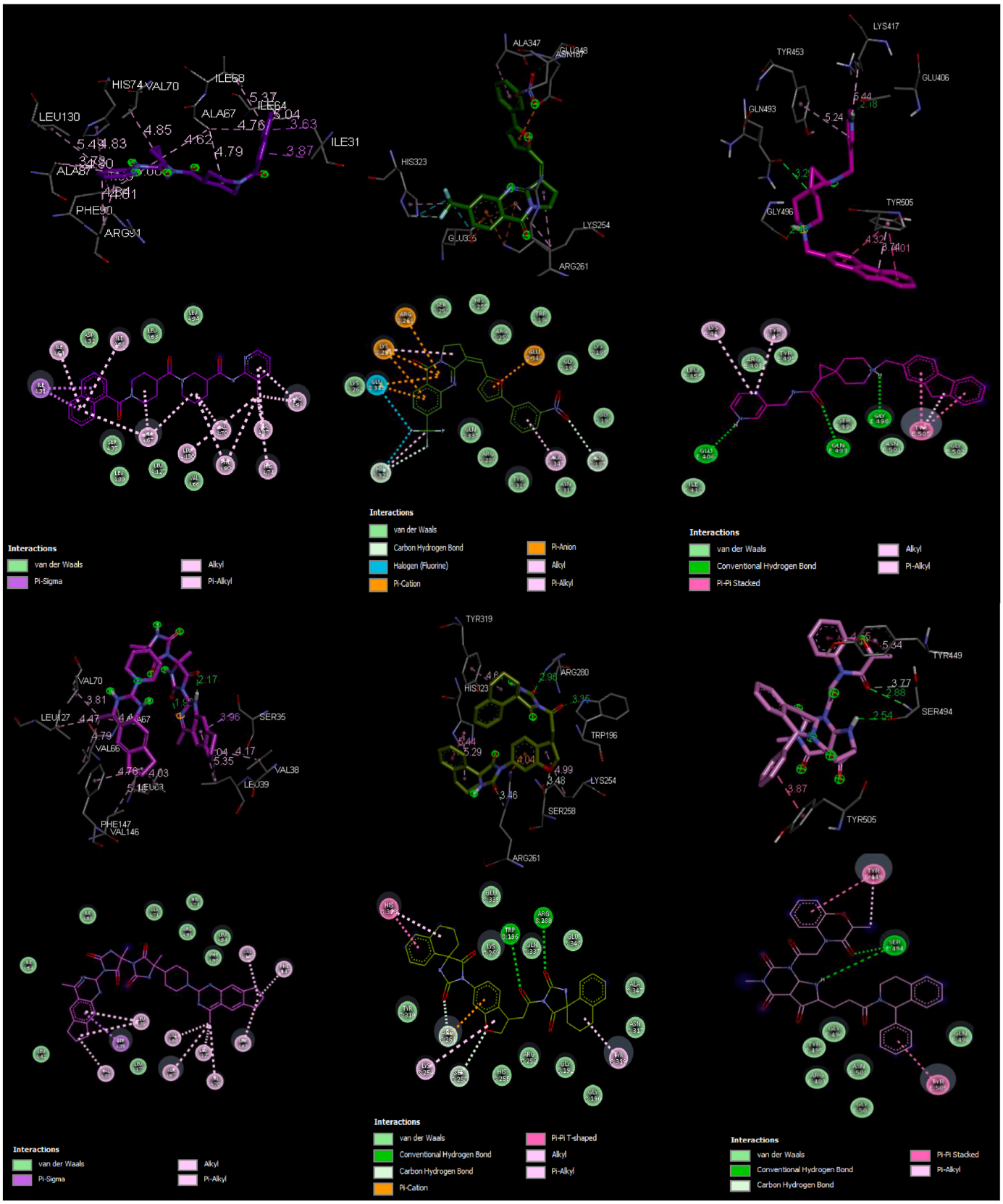


Fig. 4. 3D and 2D representation of interactions of screened ligands vs FDSL-DD ligands. a. TIPE2-T2C1. B. RelA-RAC3. C. S-protein-SPC5. D. TIPE2-T2C2. E. RelA-RAC4. F. S-protein-SPC6.

TIPE2, with a 3.6 kcalmol^{-1} difference between the top ligand from the FDSL-DD and top ligand from the HTVS. Solubility and partition coefficient remained within the same range of the HTVS ligands. The molecular weight saw a significant increase for T2C2 and SPC6. And while approved drugs have been increasing in molecular weight and surpassing the 500 gmol^{-1} maximum guideline in recent years [8] bringing this value down with future adjustments could also contribute to improving solubility. The FDSL-DD synthesized ligand, RAC4, is the best overall in terms of druglikeness (Table 3), meeting the constraints of three widely accepted guidelines; Linpinski, Veber and Egan. Imposing stricter penalties into the selection method will promote better candidates, such as RAC4, that not only exhibit satisfactory binding affinities but desirable ADME and pharmacokinetic properties.

Examination of the interactions (Fig. 4) of T2C1 versus T2C2 reveals both compounds interact with the amino acid residues of the binding pocket in the same way; alkyl-alkyl, π-alkyl, π-sigma and van der Waals interactions. The difference between these two compounds is size. T2C2 is nearly 200 gmol^{-1} larger than T2C1 with an additional 8 Å in length. As such the significant increase in binding affinity is seemingly an increased surface area.

Between RAC3 and RAC4 there is a 2.4 kcalmol^{-1} increase in affinity for the FDSL-DD constructed ligand. This stronger binding potential is exhibited in the addition of hydrogen bonding and π-π stacking for RAC4. The remainder of the interactions between RelA and RAC4 are similar to the interactions between RelA and RAC3; we see van der Waals, carbon-hydrogen, π-cation, alkyl, and π-alkyl interactions contributing to the binding affinity for both RAC4 and RAC3.

The potential inhibitors for the Spike protein, SPC5 and SPC6, both exhibit van der Waals, hydrogen bonding, π-π stacking and π-alkyl interactions. The built ligand, SPC6, also includes carbon-hydrogen bond interactions and is significantly larger in size, contributing to the increase in binding affinity.

3. Materials and methods

3.1. Preparation of receptor and ligands

The crystal structures of the protein files; TIPE2 (PDB ID: 3F4M), RelA (PDB ID: 5IQR), and S-protein (PDB ID: 6M0J); were retrieved from the RCSB Protein Data Bank. The structures were cleaned; removing all waters, co-crystalized proteins, and co-crystalized atoms; and prepared in AutoDockTools-1.5.6 [17] with the addition of polar hydrogens and calculation of Gasteiger charges. Ligands from an Enamine Ltd. "drug-like" library consisting of around 250,000 molecules were retrieved and optimized using OpenBabel [18] through the generation of 3D structures, addition of charges and minimization using the MMFF94 force field.

3.2. Grid box and molecular docking

Grid boxes were generated for each protein in accordance with their binding site. The grid box for RelA was centered at x = 297.894, y = 163.593, and z = 219.301 with dimensions of 25.000 Å. For the S-protein the box of dimensions 22.000 Å × 42.000 Å × 22.000 Å were centered at x = −27.878, y = 25.205, and z = 5.514. Due to large binding cavity of TIPE2, 4 grid boxes were generated. These grid boxes span the pocket entrance, occluding the cavity). All grid box quadrants are of 12.000 Å dimensions with quadrant 1 centered at x = 60.677, y = 5.646, and z = 17.000. Quadrant 2 is centered at x = 62.636, y = 11.365, and z = 19.959. Quadrant 3 is centered at x = 68.067, y = 10.738, and z = 18.594. And quadrant 4 is centered at x = 67.024, y = 5.362, and z = 17.113. All ligands were docked with each protein using their respective grid boxes. High throughput molecular docking calculations of mass libraries were performed using AutoDock Vina 1.1.2 [19,20]. on the University of Drexel's high-performance computer cluster, Picotte. Docking with TIPE2 generated 4 times the output, as every ligand was docked in each quadrant grid box.

3.3. FDSL-DD

The Autodock VINA output file includes nine binding solutions, each with a predicted protein-ligand binding affinity. The FDSL-DD selects the lowest binding affinity solution, and it extracts the PDBQT file and binding affinity value. The ligand PDBQT-format file is converted to a PDB format and merged with the protein PDB file using VINA and UNIX command line tools to transform, concatenate, and reorder the text in the files. This provides the input for the Protein Ligand Interaction Profiler (PLIP) [21]. PLIP performs a rule-based prediction of interactions between ligand atoms and protein amino acids, including the bond type and atom–residue pairs.

The PDBQT file of the ligand is also converted to a MOL format file, or Molfile, which is a text file containing 'information about the position of a molecule [22]. The MolFile is then broken into fragments using BRICS [23], an algorithm designed to break bonds in a chemically realistic manner. The BRICS fragmentation methodology is an automatic decomposition tool that employs a set of rules which avoids redundant fragments, unwanted chemical motifs, and small terminal fragments. This generates a diverse set of fragments based off a given ligand. Other fragmentation methodologies can be used, including RECAP [24], in-place of BRICS if desired. The fragmentation algorithm is

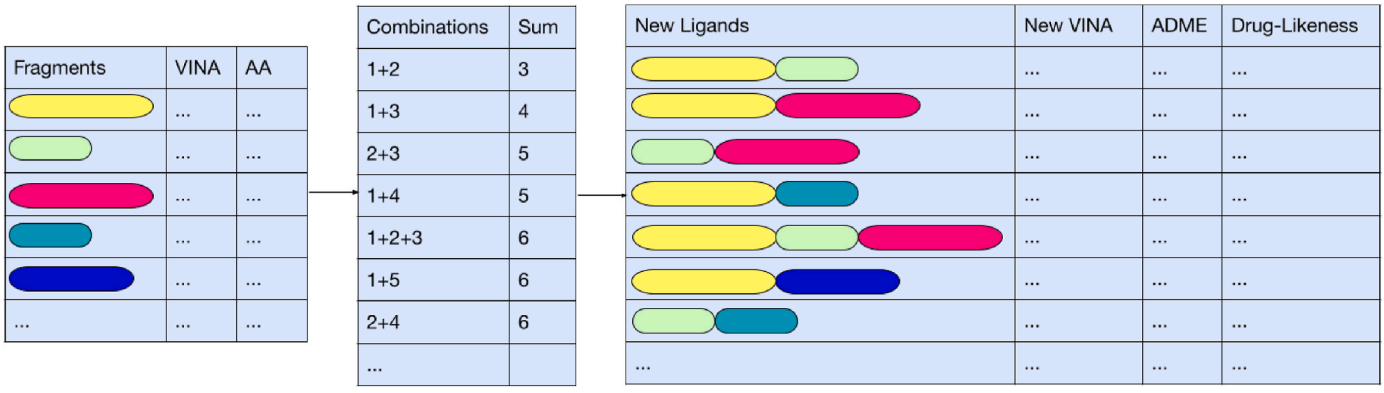


Fig. 5. Scheme used for generating fragments. The ranked fragment list is first created based on the Autodock VINA binding scores of their parent ligands and fragment counts as determined using the FDSL-DD shown in Fig. 1. The search space is generated by using triangular numbers to generate all possible combinations of ranked fragments up to a given threshold. The example shown above is to generate all possible ranks up to rank 4, which is based on the triangular number 6, which includes all possible ways of summing to 3, 4, 5, and 6. These are then the ranks of fragments that are used in combination. Possible candidate ligands are then generated from the prioritized rank list and evaluated using Autodock VINA, as well as for ADME and drug-likeness properties.

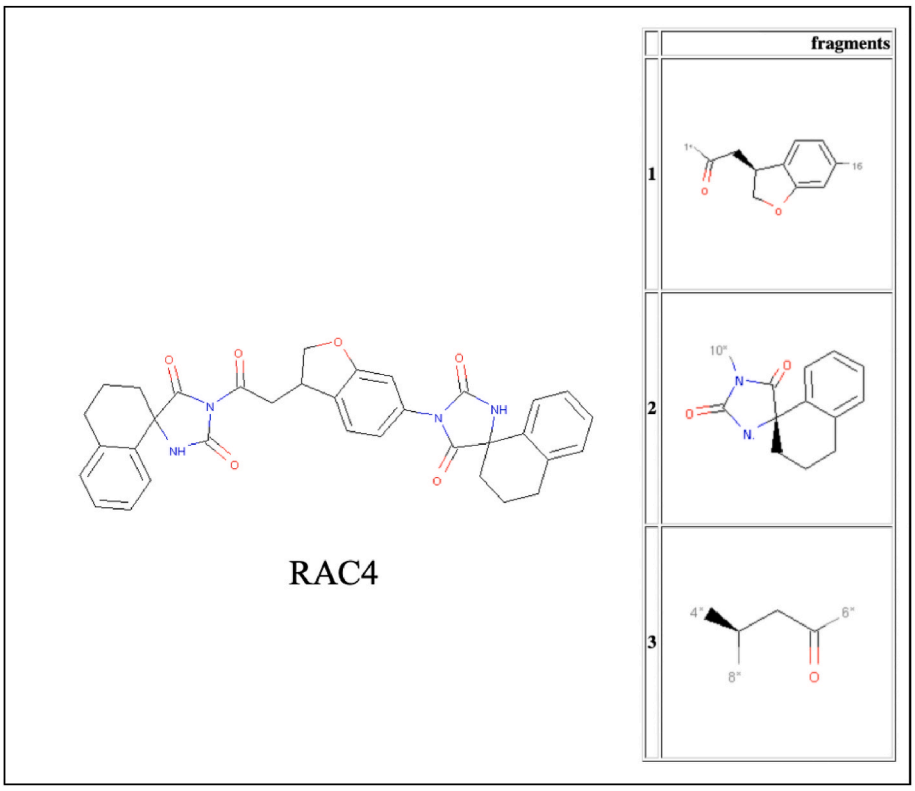


Fig. 6. An example of a final generated ligand and the fragments fed into the BRICS algorithm to generate it.

implemented in Python 3.8 using the RDKit open source chemoinformatics software package, http://www.rdkit.org. The fragments are then associated with their "parent" ligands (the ligands that were fragmentized). Many fragments will have multiple parent ligands, i.e., they will have appeared as the fragments of multiple ligands.

The location of the fragment in the binding pocket is then identified for each parent ligand following the procedure implemented by Tang et al., 2014 [25]. The ligand atoms constituting the fragment are identified by finding the maximum common substructure (MCS) between the fragment and ligand using RDKit. The XML-formatted output of PLIP is parsed to obtain the residues identified as binding to the ligand atoms corresponding to the fragment. The fragment is then associated with those residues. In many applications, groups of protein residues will be defined as binding pocket subregions, e.g., subregions A, B, or C. If a fragment's constituent atoms bind to the residues in a subregion, then the fragment is associated with that subregion. In some ligand contexts, a fragment will bind to residues in multiple subregions.

As shown in Fig. 1, a table or database is then created with entries for each distinct fragment-subregion combination. The fragment is stored in a SMILES format [26], which is a string that includes the structural information required to reconstruct the ligand. The fragment will have multiple entries if it is found binding to multiple binding pocket subregions in different ligand contexts. For each fragment-subregion combination, binding affinity statistics are also computed. Generally, the database will include the mean binding affinity predicted by AutoDock Vina for all parent ligands including that fragment-subregion combination. It may also, or instead, include the median or mode binding affinity. Another statistic that may be calculated is the deviation between the mean parent ligand binding affinity for that fragment-subregion combination and the overall mean binding affinity for all ligands in the screening experiment. Finally, the database will also include the number of parent ligands ("Count" in Fig. 1) in which the fragment-subregion combination is found.

3.4. Computational ligand design

A basic prioritized search for novel ligands based on the fragment library generated by the FDSL-DD is described as follows and in Fig. 5. As an initial step, fragments are sorted based on the inferred binding affinity of their parent ligand, as shown in Fig. 1. Additional sorting can be performed – for example, fragments binding to particular regions (as shown in Fig. 1) may be placed in different pools, and fragments may be drawn from them according to the scheme shown in Fig. 5 and described here. A set of novel ligands is generated based on a triangular number sequence. As shown in Fig. 5, all possible ways of summing the numbers m through n, where m is the first ranked fragment (generally $m = 1$) and n is the last ranked considered for combination within the ranked list of fragments are generated. For example, for $m = 1$ and $n = 5$, the combinations are $1 + 2$, $1 + 3$, $1 + 4$, and $2 + 3$. That means that candidate ligands are the combination of fragments ranked 1 and 2, 1 and 3, 1 and 4, and 2 and 3 respectively. Because most fragments are between 100 and 300 g/mol, combinations are only added with up to 5 fragments to comply with Lipinski's 500 g/mol rule. This significantly decreases computational burden, as only a subset of ligands that can be generated with 5 operands are evaluated.

BRICS fragments are combined using BRICS rules using the BRICS. Build command in RDKit. This is possible because the MolFile or SMARTS representations of BRICS fragments will store information about the broken bonds in BRICS fragmentation in isotopes. The BRICS. Build package in RDKit can utilize the isotope information to attempt to recombine fragments in new combinations according to the information in the isotopes. If the resulting molecule can be parsed by RDKit, then it is successful potential molecule. If isotopes remain, indicating potential binding sites, then other fragments can be added on to the growing ligand. If a different computational fragmentation procedure is used, then different computational methods can be used to assemble fragments. For example, while not implemented for the results shown in the paper, the pipeline allows for RECAP fragments, if generated, to be

combined through text processing of SMILES strings by replacing the (*) wildcard character used to indicate broken bonds (i.e. open binding sites).

It is important to note that not every fragment fed into the BRICS recomposition algorithm is guaranteed to be included in the final molecule. Fig. 6 shows, for example, RAC4 and its parent fragments. As shown in Fig. 6, in the case of RAC4, the third fragment fed did not end up in the final ligand. Additionally, BRICS may repeat fragments to fill the valences of each incorporated fragment. In the case of RAC4, fragment 1 and 3 had more than one fragment end. Regardless of which fragment was used, at least one repeat of fragment 2 is necessary to fill in all valences since fragment 2 is the only fragment with only one exposed end. Although repeat fragments are not ideal from a diversity perspective, in some cases, they enable molecules with better binding affinities to be generated, like in the case of RAC4.

The resulting ligands are then evaluated on the basis of their binding affinity using Autodock VINA. Optionally, *in silico* absorption, distribution, metabolism, and excretion (ADME) properties are also calculated using SwissADME, as well as other drug likeness properties such as the Lipinski's rule of 5 parameters, using built-in RDKit features in the rdkit.Chem.Lipinski module [27]. The highest ranking ligands on the basis of binding affinities and ADME properties are also visualized in protein-ligand complexes with PyMOL [28] and ChimeraX-1.2.1 [29,30].

The source code for the methods described herein will be made available on request for research and educational purposes only and under a license that prohibits any commercial or third party use.

4. Conclusions

In our study, we have developed a new method, FDSL-DD, that combined AI and FDDD, to effectively reduce the amount of time spent in pre-clinical phases of the drug development process. In the new method, *in silico* docking studies were performed on a large database of molecules to create a library of molecular fragments for specific targets. These fragments are then combined to form new inhibitors with higher binding affinity. Creating a fragment database from a large, already docked, ligand screening library using artificial intelligence based on fragment characteristics will expand the prospects for discovering and designing new therapeutic options. Moreover, ligand screening supplies additional information that allows for more effective generation of candidate ligands. Although further improvements are needed to fine-tune the structures so the compounds can pass the Lipinski's rule of five [9] and other rules for druggability [31], our approach to computationally integrate knowledge from a prior ligand screening step proved to be an effective method for developing stronger inhibitors.

Author contributions

Conceptualization, H.J. and G.R.; methodology, all authors; software, all authors; validation, J. W., B. S., W. C., R. C.; formal analysis and investigation, J. W., B. S., W. C., R. C.; data curation, J. W., B. S., W. C., R. C.; writing—; J. W., B. S., W. C., R. C.; writing—; J. W., B. S., W. C., R. C.; G. R, H.J.; supervision, project administration, and funding acquisition, G. R., H.J. All authors have read and agreed to the published version of the manuscript.

Funding and acknowledgments

Work reported here was run on hardware supported by Drexel's University Research Computing Facility, and this work was partially supported by the NSF REU supplement to grant #1919691.

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

Declaration of competing interest

There is no interest of conflict.

Data availability

No data was used for the research described in the article.

References

[1] D. Sun, W. Gao, H. Hu, S. Zhou, Why 90% of clinical drug development fails and how to improve it? Acta Pharm. Sin. B (2022).
[2] A.V. Sadybekov, V. Katritch, Computational approaches streamlining drug discovery, Nature 616 (7958) (2023) 673–685.
[3] E.N. Muratov, R. Amaro, C.H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J.L. Medina-Franco, K.M. Merz, T.I. Oprea, V. Poroikov, G. Schneider, M.H. Todd, A. Varnek, D.A. Winkler, A.V. Zakharov, A. Cherkasov, A. Tropsha, A critical overview of computational approaches employed for COVID-19 drug discovery, Chem. Soc. Rev. 50 (16) (2021) 9121–9151.
[4] R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R.K. Ambasta, P. Kumar, Artificial intelligence to deep learning: machine intelligence approach for drug discovery, Mol. Divers. 25 (3) (2021) 1315–1360.
[5] H. Zhu, Big data and artificial intelligence modeling for drug discovery, Annu. Rev. Pharmacol. Toxicol. 60 (2020) 573–589.
[6] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, S. Zhao, Applications of machine learning in drug discovery and development, Nat. Rev. Drug Discov. 18 (6) (2019) 463–477.
[7] N. Fleming, How artificial intelligence is changing drug discovery, Nature 557 (2018).
[8] M.D. Shultz, Two decades under the influence of the rule of five and the changing properties of approved oral drugs, J. Med. Chem. 62 (4) (2019) 1701–1714.
[9] O. Ichihara, J. Barker, R.J. Law, M. Whittaker, Compound design by fragment-linking, Mol Inform 30 (4) (2011) 298–306.
[10] W.P. Jencks, On the attribution and additivity of binding energies, Proc. Natl. Acad. Sci. USA 78 (7) (1981) 4046–4050.
[11] A. Daniel, J.A.W. Erlanson, Andrew braisted, tethering: fragment-based drug discovery, Annu. Rev. Biophys. Biomol. Struct. 33 (2004) 199–223.
[12] L. Hoffer, J.P. Renaud, D. Horvath, Fragment-based drug design: computational & experimental state of the art, Combinatorial chemistry & high throughput screening 14 (6) (2011) 500–520.
[13] S.A. Fayngerts, Z. Wang, A. Zamani, H. Sun, A.E. Boggs, T.P. Porturas, W. Xie, M. Lin, T. Cathopoulis, J.R. Goldsmith, A. Vourekas, Y.H. Chen, Direction of leukocyte polarization and migration by the phosphoinositide-transfer protein TIPE2, Nat. Immunol. 18 (12) (2017) 1353–1360.
[14] J.W. Dehong Yan, Honghong Sun, Zamani Ali, Jiacheng Bi, Honglin Zhang, Qingguo Ruan, Xiaolu Yang, Youhai H. Chen, Xiaochun Wan, TIPE2 Protein Specifies the Functional Polarization of Myeloid-Derived Suppressor Cells during Tumorigenesis, 2019. *Unpublished Manuscript*.
[15] L. Hall-Stoodley, L. Nistico, K. Sambanthamoorthy, B. Dice, D. Nguyen, W. J. Mershon, C. Johnson, F.Z. Hu, P. Stoodley, G.D. Ehrlich, J.C. Post, Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in Streptococcus pneumoniae clinical isolates, BMC Microbiol. 8 (2008) 173.
[16] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2) (2020) 281–292 e6.
[17] M.F. Sanner, Python: a programming language for software integration and development, J. Mol. Graph. Model. 17 (1999) 57–61.
[18] M.B. Noel M O'Boyle, Craig A. James, Chris Morley, Tim Vandermeersch, Geoffrey R. Hutchison, Open Babel, An open chemical toolbox, J. Cheminf. 3 (33) (2011).
[19] O. Trott, A.J. Olson, AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010) 455–461.
[20] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings, J. Chem. Inf. Model. 61 (8) (2021) 3891–3898.
[21] M.F. Adasme, K.L. Linnemann, S.N. Bolz, F. Kaiser, S. Salentin, V.J. Haupt, M. Schroeder, Plip 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA, Nucleic Acids Res. 49 (W1) (2021) W530–W534.
[22] J.G.N. Arthur Dalby, W. Douglas Hounshell, Ann K.I. Gushurst, David L. Grier, Burton A. Leland, John Laufer, Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited, 1992.
[23] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, M. Rarey, On the art of compiling and using 'drug-like' chemical fragment spaces, ChemMedChem 3 (10) (2008) 1503–1507.
[24] D.B.J. Xiao Qing Lewell, Stephen P. Watson, Michael M. Hann, RECAPsRetrosynthetic combinatorial analysis procedure: a powerful new

technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry, J. Chem. Inf. Comput. Sci. 31 (1998) 511–522.
[25] G.W. Tang, R.B. Altman, Knowledge-based fragment binding prediction, PLoS Comput. Biol. 10 (4) (2014), e1003589.
[26] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1998) 31–36.
[27] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017), 42717.
[28] L. Schrödinger, The PyMOL Molecular Graphics System, 2015, Version 2.0.
[29] G.T. Pettersen Ef, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris, T. E. Ferrin, U.C.S.F. ChimeraX, Structure visualization for researchers, educators, and developers, Protein Sci. 30 (1) (2021) 70–82.
[30] E.F.G.T.D. Pettersen, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T. E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[31] G.R. Bickerton, G.V. Paolini, J. Besnard, S. Muresan, A.L. Hopkins, Quantifying the chemical beauty of drugs, Nat. Chem. 4 (2) (2012) 90–98.