International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Identification of Novel Small Molecule Inhibitors Against nsP2 Protease of CHIKV through a Molecular Modeling Approach

Sutanu Mukhopadhyay

Department of Chemistry, Ramakrishna Mission Vivekananda Centenary College, Rahara, West Bengal, India
* Corresponding Author Email ID: *sutanubermensch[at]gmail.com*
ORCID ID: 0000-0001-8243-2138

Abstract: ***Chikungunya is a tropical viral disease spread by the female Aedes mosquitoes infected with the Chikungunya virus (CHIKV). Non-structural protein 2 (nsP2) plays a crucial role in the viral life cycle by its proteolytic activity and hence it is one of the most important drug targets. There is currently no permanent treatment available to tackle the infection. In this molecular modeling-based study, a combination of de novo ligand design, molecular docking, and ADMET-based screening is employed to identify novel inhibitor molecules targeting the active site of nsP2 protease of the CHIKV. A set of molecules have been shortlisted as potential inhibitors based on their binding affinity and drug-likeness score. Further experimental validation is required to verify the potency of the proposed leads against CHIKV nsP2 protease activity to combat the infection.***

Keywords: ADMET, artificial neural network, CHIKV, de novo drug design, docking, nsP2 protease

1. Introduction

Chikungunya is a viral infectious disease spread by the bites of infected female Aedes mosquitoes. It produces fever, joint aches and swelling, muscle pain, muscle pain, nausea, headache, and rash. It was initially reported in Southern Tanzania, and it has now been detected in over forty countries worldwide[1]–[3]. It is an RNA virus, belongs to the Togaviridae family's alphavirus genus. The disease is mostly transmitted by two mosquito species known as *Aedes aegypti*and *Aedes albopictus*[4]. The word "chikungunya" is originated from a phrase in the Makonde language that means "that which bends up," and refers to the hunched posture of people suffering from joint discomfort (arthralgia)[1].

The virus's genome is translated into nine proteins, five of which are structural (one capsid protein (C), two envelope glycoproteins (E1, E2), and two peptides, E3 and 6K) and four of which are non-structural (nsP1, nsP2, nsP3, and nsP4)[5], [6]. Among the structural proteins, capsid protein initiates the viral life cycle within the host cell via autoproteolysis, whilst envelope glycoproteins aid in the viral entrance[7]. nsP1 and nsP3 synthesize the negative sense of the RNA strand, while nsP4 controls virion polymerization in the host cell. The multifunctional protein nsP2 is a member of the papainsuperfamily of cysteine proteases, which is well recognized for its proteolytic activity and involvement in viral replication. Therefore, nsP2 is one of the most promising drug targets for developing drugs against the infection[8]. Several computational and experimental studies have been conducted in search of inhibitors of nsP2 through different routes such as molecular modeling, structure-based drug design, molecular docking, and MD simulation, pharmacophore mapping, biological evaluation, and many more[9]–[17]

Deep learning contributes to the field of drug discovery in multiple ways from searching potential binding pockets on the surface of the protein [18]–[20] to generating novel molecules based on the protein pocket information [21]–[24]. LiGANN uses deep learning-based generative adversarial networks (GANs) to generate ligands taking the protein structure information [21].

In this in silico study, the crystal structure of CHIKV nsP2 protease was obtained from protein databank [25] and, based on the active site information, novel lead compounds were created using LiGANN, a de novo drug design tool based on generative neural network (GNN) [21]. All of the suggested molecules are then manually docked to the nsP2 active site using PyRx, an autodock-embedded software [26], [27]. The highest scoring ligands' ADMET characteristics were evaluated using the SwissADME webserver [28]. Further studies are needed to investigate the efficiency of the suggested compounds in treating chikungunya virus infection.

2. Materials and Methods

Protein structure retrieval: The crystal structure of the chikungunya virus nsP2 protease has been downloaded in .pdb format from the RCSB PDB website (PDB ID: 3TRK). The crystallographic water molecules and the ions are removed and the protein surface is visualized using PyMol [29]. The catalytic dyad, CYS 1013 and HID 1083 is highlighted.

de novo ligand design: The prepared protein pdb file has been uploaded as input in the LiGANN webserver (https://playmolecule.com/LiGANN/), number of ligand shape generations and decoding per shape has been set as 10 and 10 respectively. A grid box has been generated centering the catalytic dyad (CYS 1013 and HIS 1083) of the protein.

Molecular docking: Then the protein has been processed and converted into. pdbqt using PyRx "Make macromolecule" utility. A total of 88 molecules have been generated by LiGANN, using Open Babel, the SMILES ID of the compounds are converted into .sdf file format[30]. All the downloaded compounds have been energy minimized and converted into. pdbqt format before docking using PyRx. A grid box has been created around the catalyticdyad (CYS 1013 and HIS 1083) and docking has been performed using the vina wizard utility of PyRx.The 2D depiction of interactions of the protein-ligand complexes has been created using Discovery Studio visualizer[31].

ADMET prediction: ADMET stands for Absorption, Digestion, Metabolism, Excretion, and Toxicity, this is a measure of drug metabolism and pharmacokinetics (DMPK) which is crucial for drug discovery. ADMET and physicochemical properties, drug-likeness, synthetic accessibility of the top-scoring leads have been estimated using the SwissADME web server.

3. Results and Discussions

Protein structure retrieval: The crystal structure of the chikungunya virus nsP2 protease (3TRK) is downloaded in .pdb file format consists of 324 amino acids and has a resolution of 2.40 Å. The three-dimensional structure is visualized in cartoon representation; the catalytic dyads are shown in licorice representation (**Figure 1**).

de novo ligand design: The modified pdb file is given as input on the LiGANN server, which results in 88molecules based on the protein active site information. The molecules belong to different chemical classes so that there's a diverse group of ligands.

Molecular docking: Among the 88 ligands, 17 ligands have a binding affinity score of ≤ -6.0 Kcal/mol; but all of them are not interacting with at least one of the catalytic dyad residues. In the case of the best scoring lead (**Figure 2.A**), *molecule A* (-7.6 kcal/mol) has a total of seven non-bonding interactions with the neighboring amino acids at the active site of the protein including two pi-sulfur interactions with CYS 1013(distance: 4.91 Å) and MSE 1238 (Selenomethionine), a hydrogen bond with TYR 1047, two pi-sigma interactions with ALA 1046 and MSE 1242, two pi-alkyl interactions with ALA 1046 and TYR 1079.*Molecule D, H, L, M, N* interacts with CYS 1013 with pi-sulfur non-bonded interactions while *molecule E*interacts through pi-alkyl interaction. A list of all the interactions is listed in **Table 1** and all the 2D protein-ligand interaction images are shown in **Figure 2**.

ADMET prediction: The SwissADME web-server was used to determine various physicochemical parameters, leadlikeness, and synthetic accessibility of the top 17 lead compounds, as shown in **Table 2**. All the molecules have a molecular weight of ≤ 500 g/mol,the number of hydrogen acceptor atoms (HA) and hydrogen donor atoms (HD) are ≤ 10 and ≤ 5 respectively which are the criteria in Lipinski's rule of five for drug-likeness. According to Veber's rule of drug-likeness, a drug-like molecule should have ≤ 10 rotatable bonds and Topological Polar Surface Area (TPSA) of ≤ 140 Å², most of the molecules passed the first criterion (*Molecule A,B, D, E, F, G, H, I, L,N, O*) and all the molecules satisfied the second criterion. Log Po/w is a measure of lipophilicity of the molecules; here a consensus value is given as an average of five values predicted in different manners. A drug-like molecule must have optimal aqueous solubility; most compounds are moderately soluble in water. The bioavailability score indicates the percentage of the drug that may reach the target location; all molecules have a bioavailability score of 0.55. After summing up all the information related to leadlikeness, 6 molecules are found as leadlike (*Molecule A, B, D, F, G, N*) according to SwissADME predictions. Synthetic accessibility score, predicted by the server signifies the ease of synthesizing the molecules in the laboratory, where score 1 means the molecule can be synthesized very easily and 10 means it is very difficult to synthesize the molecule.

4. Conclusions

Chikungunya is a widespread threat to public health. There is currently no viable medicine or vaccination that can totally cure the disease. Using a deep learning methodology, this investigation discovers several new leads. The lead compounds demonstrate significant efficacy against one of the virus's most important therapeutic targets, nsP2, in terms of PyRx binding affinity score, protein-ligand non-bonding interactions, and physicochemical features. Further experimental research is necessary to confirm their effectiveness and find viable treatments for the viral infection.

Table 1: Details of the binding affinities and interacting amino acid residues for top 17 ligands

Molecule	*SMILES*	*Binding affinity score (kcal/mol)*	*Neighboring interacting residues*
A	CN1CCC(CC1)C1ONC(N1)[C@@H]1CNc2c(CC1)cccc2	-7.6	CYS 1013. ALA 1046(2), TYR 1047, TYR 1079, MSE 1238, MSE 1242
B	N1CC(C1)CNC1C(Cc2ccccc2)Cc2n1ncn2	-7.0	TYR 1047, TYR 1079, TRP 1084, ASP 1246
C	CCC(NC(=O)CSc1nnc(s1)NCCCN(Cc1ccccc1)C)C	-6.7	TRP 1014, ALA 1046, TYR 1047 (3), TYR 1079 (2), ASN 1082, TRP 1084, MSE 1242
D	Sc1nnc(n1CCC1CNCC1)CCc1ccccc1	-6.7	CYS 1013, ALA 1046, LEU 1205, MSE 1242, ASP 1246

E	CNCCCC(=O)N1CCN(c2nnc(CCCN3CCCC3)n2C)CC1	-6.7	CYS 1013, ALA 1046, TYR 1047, GLU 1050, TYR 1079, ASN 1082, LYS 1091 (unfavorable), MSE 1242
F	CN(C)CC(CNc1cccnn1)C1CCCc2ccccc21	-6.6	ALA 1046, TYR 1047 (2), SER 1048, TYR 1079
G	CC1CC(CN)CN1c1nnc(Cc2ccccc2)n1CC1CC1	-6.5	ASN 1082, TRP 1084, LEU 1205
H	COCC(C)Nc1cc2ccccc2nn1	-6.4	ASN 1011 (2), CYS 1013, ALA 1046 (2), TYR 1047, TYR 1079, ASN 1082, MSE 1242
I	CN(C)CCN(CCc1ccccc1)Cc1ccccn1	-6.3	TYR 1047, TYR 1079 (2), TRP 1084, LEU 1205, ASP 1246
J	CNCCCNC(=O)CCC(CN(C)C)NC(=O)OCc1ccccc1	-6.3	TYR 1047, TYR 1079, ASN 1082, TRP 1084, LYS 1091
K	CC(C)NC(=O)NCCCC(=O)NCCN(C)Cc1ccco1	-6.2	ASN 1082 (2), TYR 1079
L	CNC(C)(C)Cc1nc(CCCCc2ccccc2)no1	-6.2	CYS 1013, ALA 1046 (2), TYR 1079, ASN 1082
M	C=CCN(CCN1CCC(C)(C(=O)OC)CC1)C(C)c1ccc(S(=O)(=O)N(CC)CC)cc1	-6.1	CYS 1013, ALA 1046 (2), TYR 1079, ASN 1082, TRP 1084 (2), LEU 1205, MSE 1242
N	CC(C)N(Cc1ccccc1)Cc1nncn1C(C)C	-6.0	CYS 1013, ALA 1046 (2), TYR 1079 (2)
O	CC(C)N=C(NCCc1cccs1)NCc1ccco1	-6.0	ALA 1046, TYR 1079, TRP 1084 (unfavorable)
P	CCN(C)CCNC(CCSCc1ccccc1)CN(C)CCc1nn[nH]n1	-6.0	ALA 1046, TYR 1047, TYR 1079, ASN 1082, TRP 1084, MSE 1242, ASP 1246
Q	COCCN1C2CCC1CC(N(C)CCCCN(Cc1ccccc1)Cc1ccccn1)C2	-6.0	TYR 1047 (2), TYR 1079, TRP 1084

Table 2: Physicochemical properties, Lead likeness and synthetic accessibility of the top 17 leads

Molecule	*MW (g/mol)*	*Num. rotatable bonds*	*HA*	*HD*	*TPSA (Å²)*	*Consensus Log Po/w*	*Water solubility class*	*Bioavailability score*	*Leadlikeness*	*Synthetic accessibility*
A	316.44	2	4	3	48.56	1.94	Moderately soluble	0.55	YES	4.40
B	283.37	5	4	2	54.77	1.46	Moderately soluble	0.55	YES	3.65
C	407.60	13	4	2	123.69	3.39	Poorly soluble	0.55	NO	4.23
D	302.44	6	3	1	81.54	2.59	Moderately soluble	0.55	YES	3.13
E	377.53	10	5	1	69.53	1.05	Soluble	0.55	NO	3.81
F	310.44	6	3	1	41.05	3.09	Poorly soluble	0.55	YES	3.44
G	325.45	6	3	1	59.97	2.50	Moderately soluble	0.55	YES	3.96
H	217.27	4	3	1	47.04	1.97	Moderately soluble	0.55	NO	2.60
I	283.41	8	3	0	19.37	2.80	Poorly soluble	0.55	NO	2.21
J	364.48	15	5	3	82.70	1.80	Moderately soluble	0.55	NO	3.58
K	324.42	13	4	3	86.61	1.17	Moderately soluble	0.55	NO	2.89
L	287.40	8	4	1	50.95	3.43	Poorly soluble	0.55	NO	3.36
M	479.68	13	7	0	78.54	3.50	Moderately soluble	0.55	NO	4.31
N	272.39	6	3	0	33.95	2.58	Moderately soluble	0.55	YES	2.41
O	291.41	8	2	2	77.80	2.87	Moderately soluble	0.55	NO	3.26
P	405.60	15	6	2	98.27	2.59	Poorly soluble	0.55	NO	4.12
Q	450.66	13	5	0	31.84	3.91	Poorly soluble	0.55	NO	5.06

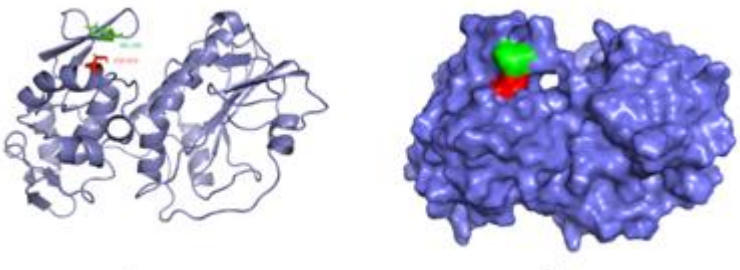

Figure 1: (A) Cartoon representation, (B) Surface representation of the CHIKV nsP2 protease (Red: CYS 1013, Green: HIS 1083)

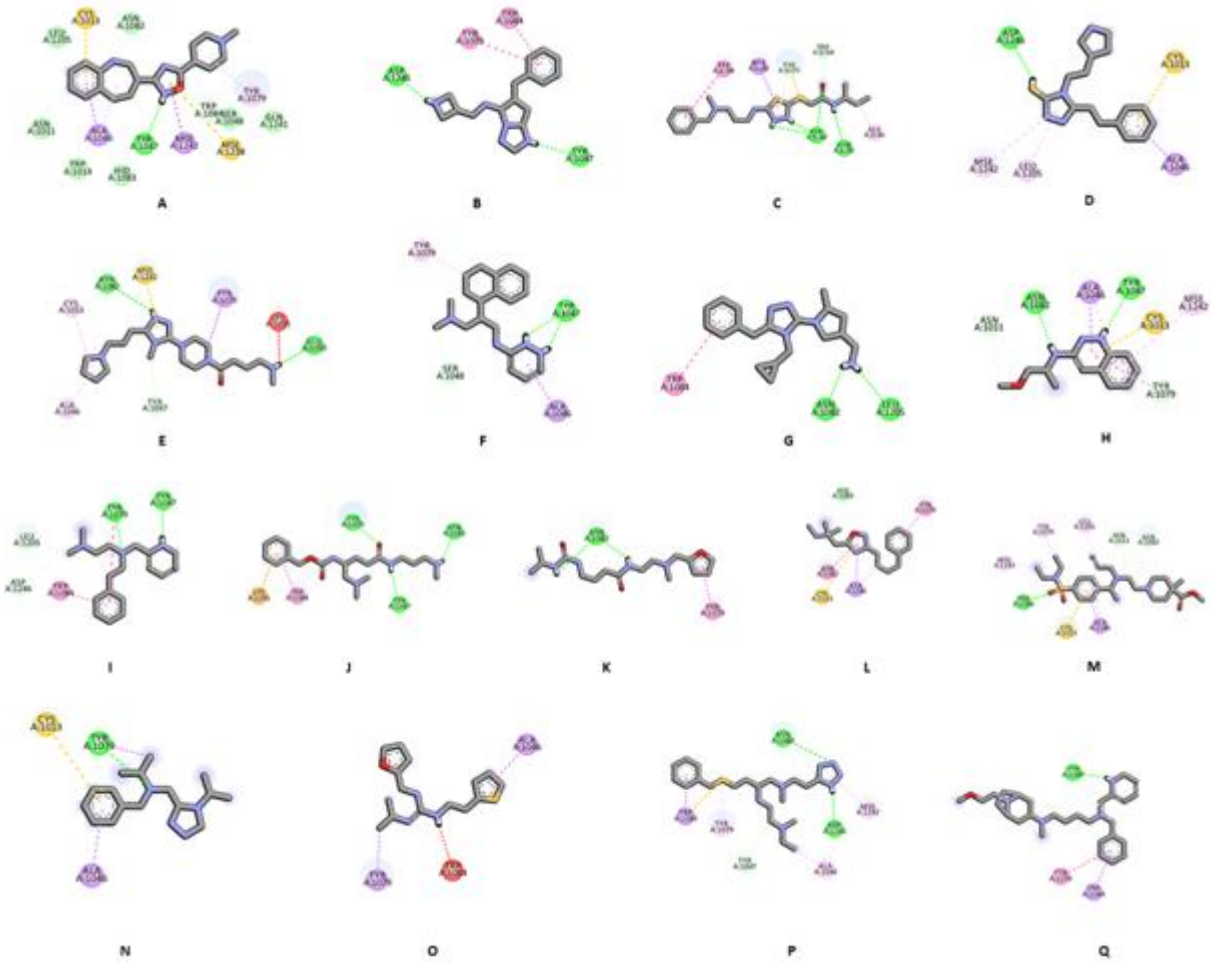

Figure 2: 2D depiction of protein-ligand interactions

References

[1] "Chikungunya fact sheet." https://www.who.int/news-room/fact-sheets/detail/chikungunya.

[2] M. C. Robinson, "An epidemic of virus disease in southern province, tanganyika territory, in 1952–1953. i. clinical features," *Trans. R. Soc. Trop. Med. Hyg.*, vol. 49, no. 1, pp. 28–32, 1955, doi: 10.1016/0035-9203(55)90080-8.

[3] P. Bodenmann and B. Genton, "Chikungunya: an epidemic in real time," *Lancet*, vol. 368, no. 9531, p. 258, Jul. 2006, doi: 10.1016/S0140-6736(06)69046-6.

[4] M. Sourisseau *et al.*, "Characterization of Reemerging Chikungunya Virus," *PLOS Pathog.*, vol. 3, no. 6, p. e89, Jun. 2007, doi: 10.1371/JOURNAL.PPAT.0030089.

[5] E. G. Strauss and J. H. Strauss, "Structure and Replication of the Alphavirus Genome," *The Togaviridae and Flaviviridae*, pp. 35–90, 1986, doi: 10.1007/978-1-4757-0785-4_3.

[6] K. Takkinen, "Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus.," *Nucleic Acids Res.*, vol. 14, no. 14, p. 5667, Jul. 1986, doi: 10.1093/NAR/14.14.5667.

[7] B. Simizu, K. Yamamoto, K. Hashimoto, and T. Ogata, "Structural proteins of Chikungunya virus," *J. Virol.*, vol. 51, no. 1, pp. 254–258, Jul. 1984, doi: 10.1128/JVI.51.1.254-258.1984.

[8] R. Sreejith *et al.*, "Mapping interactions of Chikungunya virus nonstructural proteins," *Virus Res.*, vol. 169, no. 1, pp. 231–236, Oct. 2012, doi: 10.1016/J.VIRUSRES.2012.08.006.

[9] K. D. Singh *et al.*, "Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease," *J. Mol. Model. 2011 181*, vol. 18, no. 1, pp. 39–51, Mar. 2011, doi: 10.1007/S00894-011-1018-3.

[10] J. J. Fros, E. van der Maten, J. M. Vlak, and G. P. Pijlman, "The C-Terminal Domain of Chikungunya Virus nsP2 Independently Governs Viral RNA Replication, Cytopathicity, and Inhibition of Interferon Signaling," *J. Virol.*, vol. 87, no. 18, pp. 10394–10400, Sep. 2013, doi: 10.1128/JVI.00884-13/ASSET/9573B4E0-1267-4479-9C02-B99F109E435D/ASSETS/GRAPHIC/ZJV9990980770003.JPEG.

[11] M. Lucas-Hourani *et al.*, "A phenotypic assay to identify chikungunya virus inhibitors targeting the nonstructural protein nsP2," *J. Biomol. Screen.*, vol. 18, no. 2, pp. 172–179, Feb. 2013, doi: 10.1177/1087057112460091.

[12] P. T. V. Nguyen, H. Yu, and P. A. Keller, "Identification of chikungunya virus nsP2 protease inhibitors using structure-base approaches," *J. Mol. Graph. Model.*, vol. 57, pp. 1–8, Apr. 2015, doi: 10.1016/J.JMGM.2015.01.001.

[13] P. Kumar, D. Kumar, and R. Giri, "Targeting the nsp2 Cysteine Protease of Chikungunya Virus Using FDA Approved Library and Selected Cysteine Protease

Inhibitors," *Pathog. 2019, Vol. 8, Page 128*, vol. 8, no. 3, p. 128, Aug. 2019, doi: 10.3390/PATHOGENS8030128.

[14] N. Khan, R. Bhat, A. K. Patel, and P. Ray, "Discovery of small molecule inhibitors of chikungunya virus proteins (nsP2 and E1) using in silico approaches," *https://doi.org/10.1080/07391102.2020.1731602*, vol. 39, no. 4, pp. 1373–1385, 2020, doi: 10.1080/07391102.2020.1731602.

[15] L. Ivanova, K. Rausalu, E. Žusinaite, J. Tammiku-Taul, A. Merits, and M. Karelson, "1,3-Thiazolbenzamide Derivatives as Chikungunya Virus nsP2 Protease Inhibitors," *ACS Omega*, vol. 6, no. 8, pp. 5786–5794, Mar. 2021, doi: 10.1021/ACSOMEGA.0C06191/SUPPL_FILE/AO0C06191_SI_001.PDF.

[16] L. Ivanova *et al.*, "Novel Analogues of the Chikungunya Virus Protease Inhibitor: Molecular Design, Synthesis, and Biological Evaluation," *ACS Omega*, vol. 6, no. 16, pp. 10884–10896, Apr. 2021, doi: 10.1021/ACSOMEGA.1C00625/SUPPL_FILE/AO1C00625_SI_001.PDF.

[17] M. Islamuddin *et al.*, "Inhibition of Chikungunya Virus Infection by 4-Hydroxy-1-Methyl-3-(3-morpholinopropanoyl)quinoline-2(1 H)-one (QVIR) Targeting nsP2 and E2 Proteins," *ACS Omega*, vol. 6, no. 14, pp. 9791–9803, Apr. 2021, doi: 10.1021/ACSOMEGA.1C00447/SUPPL_FILE/AO1C00447_SI_001.PDF.

[18] J. Jiménez, S. Doerr, G. Martínez-Rosell, A. S. Rose, and G. De Fabritiis, "DeepSite: protein-binding site predictor using 3D-convolutional neural networks," *Bioinformatics*, vol. 33, no. 19, pp. 3036–3042, Oct. 2017, doi: 10.1093/BIOINFORMATICS/BTX350.

[19] M. Simonovsky and J. Meyers, "Deeply Tough: Learning Structural Comparison of Protein Binding Sites," *J. Chem. Inf. Model.*, vol. 60, no. 4, pp. 2356–2366, Apr. 2020, doi: 10.1021/ACS.JCIM.9B00554/SUPPL_FILE/CI9B00554_SI_001.PDF.

[20] R. Aggarwal, A. Gupta, V. Chelur, C. V. Jawahar, and U. D. Priyakumar, "DeepPocket: Ligand Binding Site Detection and Segmentation using 3D Convolutional Neural Networks," *J. Chem. Inf. Model.*, 2021, doi: 10.1021/ACS.JCIM.1C00799/SUPPL_FILE/CI1C00799_SI_001.PDF.

[21] M. Skalic, D. Sabbadin, B. Sattarov, S. Sciabola, and G. De Fabritiis, "From Target to Drug: Generative Modeling for the Multimodal Structure-Based Ligand Design," *Mol. Pharm.*, vol. 16, no. 10, pp. 4282–4291, Oct. 2019, doi: 10.1021/ACS.MOLPHARMACEUT.9B00634.

[22] M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen, "Molecular de-novo design through deep reinforcement learning," *J. Cheminform.*, vol. 9, no. 1, pp. 1–14, Sep. 2017, doi: 10.1186/S13321-017-0235-X/FIGURES/9.

[23] M. H. S. Segler, T. Kogej, C. Tyrchan, and M. P. Waller, "Generating focused molecule libraries for drug discovery with recurrent neural networks," *ACS Cent. Sci.*, vol. 4, no. 1, pp. 120–131, Jan. 2018, doi: 10.1021/ACSCENTSCI.7B00512/SUPPL_FILE/OC7B00512_SI_002.ZIP.

[24] A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, and G. Schneider, "Generative Recurrent Networks for De Novo Drug Design," *Mol. Inform.*, vol. 37, no. 1–2, p. 1700111, Jan. 2018, doi: 10.1002/MINF.201700111.

[25] "wwPDB: 3TRK." https://www.wwpdb.org/pdb?id=pdb_00003trk .

[26] S. Dallakyan and A. J. Olson, "Small-molecule library screening by docking with PyRx," *Methods Mol. Biol.*, vol. 1263, pp. 243–250, 2015, doi: 10.1007/978-1-4939-2269-7_19.

[27] S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, and A. J. Olson, "Virtual Screening with AutoDock: Theory and Practice," *Expert Opin. Drug Discov.*, vol. 5, no. 6, pp. 597–607, Jun. 2010, doi: 10.1517/17460441.2010.484460.

[28] A. Daina, O. Michielin, and V. Zoete, "SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Sci. Rep.*, vol. 7, no. 1, p. 42717, 2017, doi: 10.1038/srep42717.

[29] Schrödinger, LLC, "The {PyMOL} Molecular Graphics System, Version~1.8," Nov. 2015.

[30] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An Open chemical toolbox," *J. Cheminform.*, vol. 3, no. 10, pp. 1–14, Oct. 2011, doi: 10.1186/1758-2946-3-33/TABLES/2.

[31] "BIOVIA Discovery Studio - BIOVIA - Dassault Systèmes®." https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/.