Predictive Modelling to Design Potential Binders of MurD, an Antibacterial Target

Sowmya Ramaswamy Krishnan, Navneet Bung and Arijit Roy

TCS Research (Life Sciences division), Tata Consultancy Services Ltd, India

Methodology overview

To design molecules specific to the allosteric site of MurD ligase protein, the conditional variational autoencoder (CVAE) model developed in our recent study (Krishnan et al., 2021) on structure-based drug design using deep learning was utilized. The model takes as input a residue-level binding site graph of the target protein of interest and can generate novel small molecules which can bind to the target binding site, based on a reinforcement learning framework. The crystal structures of MurD ligase complexed with some fragments were provided by the organizers. Based on the crystal structure, the binding site graph was constructed for the allosteric site and used as input to the pre-trained CVAE model.

The organizers specified the desired physicochemical property range (log*P*<3 and 250<=MW<=350 Da), which were considered during the molecule generation process, by posing the reinforcement learning as a multi-parameter optimization problem (Bung et al., 2022). The drug-target affinity (DTA) was also used for on-the-fly property optimization along with logP and MW. Our pre-trained drug-target affinity prediction model was used for affinity prediction, while log*P* and MW values were calculated from RDKit python package. An additive RL reward function was formulated (eqn. 1) and used to optimize the CVAE model. Specifically the drug-target affinity (log-scale bioactivity, pXC_{50}) was optimized to obtain more molecules with single digit micromolar affinity. The final RL model was used to sample 10,000 novel small molecules, which can be potential MurD ligase binders.

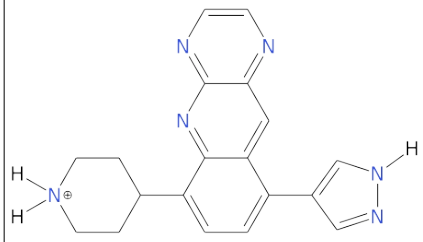
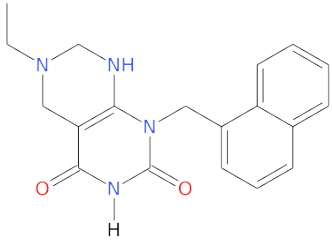
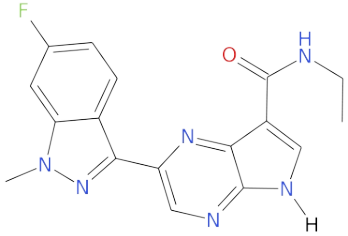
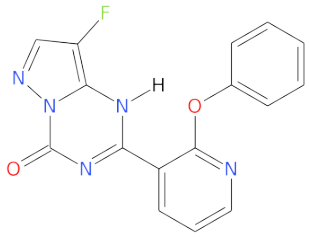
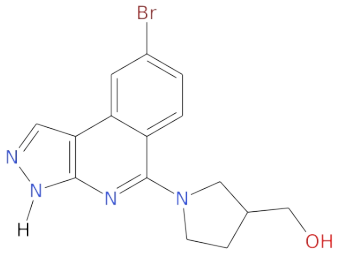
$$RL\,reward = \exp\left\{\frac{pXC_{50}}{3.0}\right\} + \left\{\begin{matrix} 11, if\ logP < 3 \\ 1, otherwise \end{matrix}\right\} + \left\{\begin{matrix} 11, if\ 250 < MW < 350 \\ 1, otherwise \end{matrix}\right\} \quad (1)$$

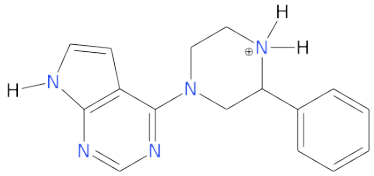
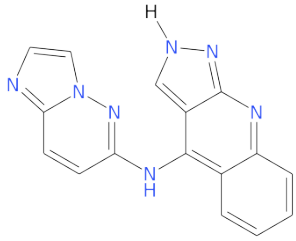
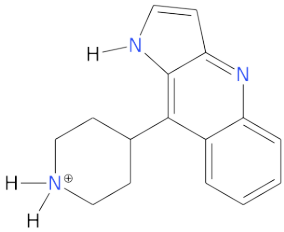
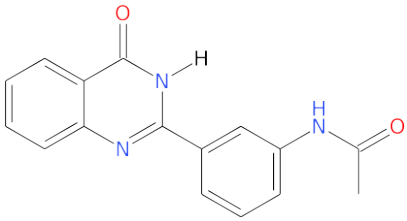
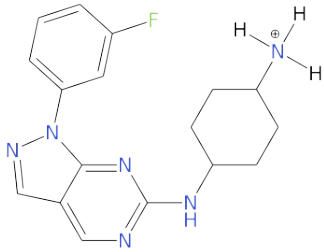
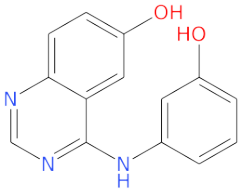
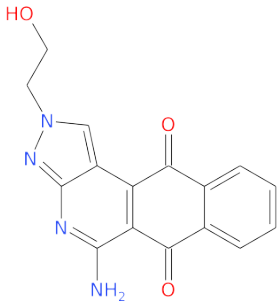
From the 10,000 molecules, the subset of chemically valid and non-redundant small molecules were selected and several property filters were applied (-1<logP<3, 250<=MW<=350 Da, predicted log-scale bioactivity>=6). To further filter the molecules based on their synthesizability, synthetic accessibility score (SAS) (Ertl and Schuffenhauer, 2009) and retrosynthetic accessibility score (RAscore) (Thakkar et al., 2021) were used for which, cut-offs of less than 4 and greater than 0.5 were chosen, respectively. Four rule-based filters to remove potentially problematic molecules namely, PAINS (Baell and Holloway, 2010), BRENK (Brenk et al., 2008), NIH (Doveston et al., 2015) and ZINC were also applied, resulting in a set of 443 small molecules. These molecules were docked at the MurD ligase allosteric site using GNINA (McNutt et al., 2021). 5 docking poses were sampled per generated molecule and the pose with the best docking score was chosen for further analyses. A docking score cut-off of less than -7.0 kcal/mol was applied to obtain a final set of 222 small molecules.

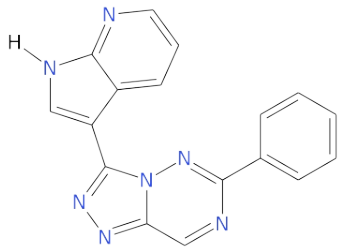
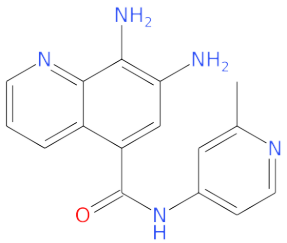
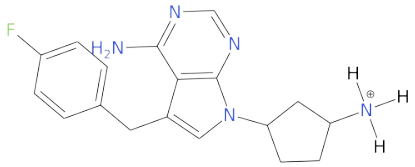
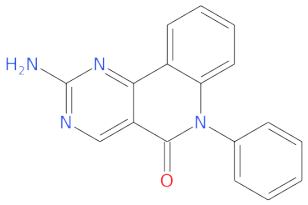
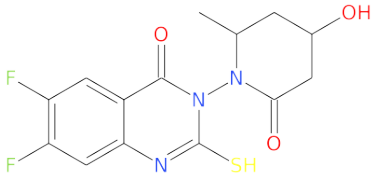
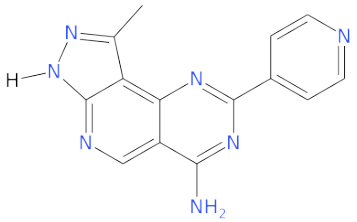
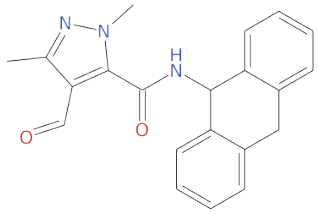
Top 30 molecules from the 222 molecules are tabulated below (Table 1). The overlap of terminal ring systems of the top 5 generated small molecules and the fragments provided by organizers are shown below (Fig. 1). These molecules were chosen based on their

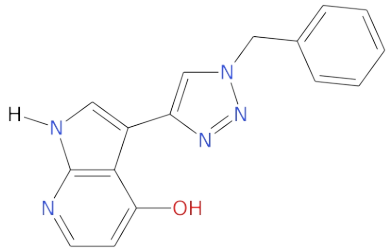
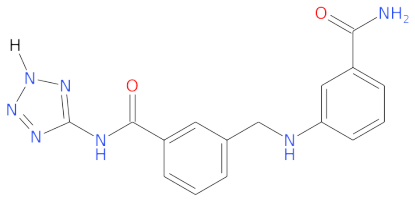
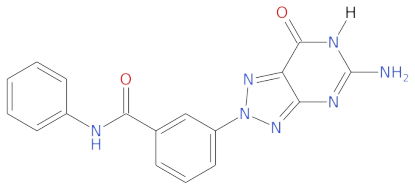
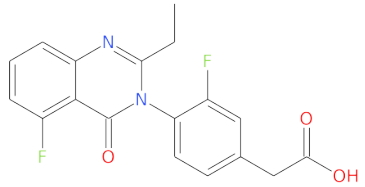
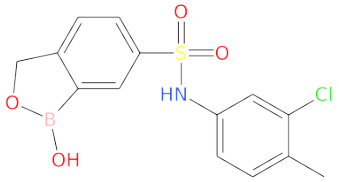
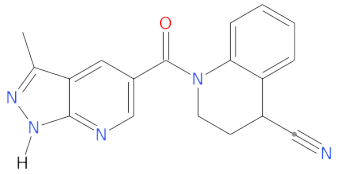
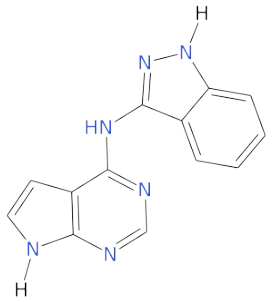
docking score and their interactions with the allosteric site residues in MurD ligase (Fig. 2-6).

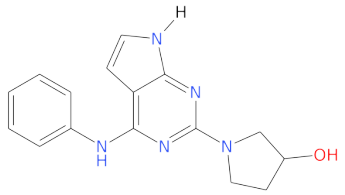
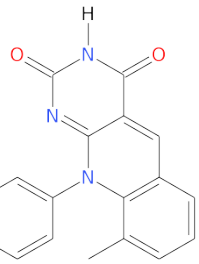
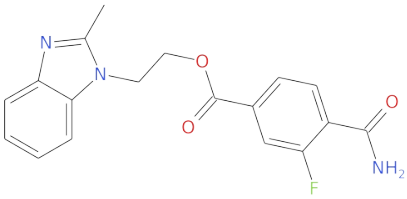
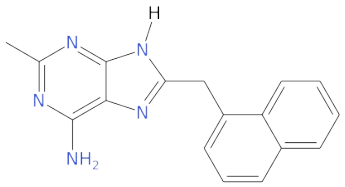
Table 1: Molecular structures and key properties of the top 30 generated small molecules from the conditional VAE model. The corresponding SMILES are provided in Appendix 1 below.

Molecule 2D structure	Molecule ID	log*P*	MW (Da)	SAS	RAscore	Docking score (kcal/mol)	Predicted log-scale bioactivity
	Mol_4104	2.0	331.16	3.77	0.98	-8.98	6.23
	Mol_6571	1.9	336.15	2.55	0.94	-8.72	6.44
	Mol_861	2.4	338.12	2.67	0.92	-8.65	6.60
	Mol_2613	2.4	323.08	2.73	0.87	-8.49	6.58
	Mol_4865	2.6	346.04	3.06	0.98	-8.46	6.60

Molecule 2D structure	Molecule ID	log*P*	MW (Da)	SAS	RAscore	Docking score (kcal/mol)	Predicted log-scale bioactivity
	Mol_3952	1.0	280.15	3.76	0.98	-8.40	6.25
	Mol_2740	2.8	301.10	2.96	0.91	-8.37	6.49
	Mol_5948	2.1	252.14	3.59	0.97	-8.36	6.18
	Mol_4092	2.5	279.10	1.95	0.98	-8.36	6.21
	Mol_7459	1.9	327.17	3.03	0.94	-8.35	6.25
	Mol_3508	2.7	253.08	2.05	0.99	-8.34	6.05
	Mol_128	0.7	308.09	2.68	0.64	-8.33	6.48

Molecule 2D structure	Molecule ID	log*P*	MW (Da)	SAS	RAscore	Docking score (kcal/mol)	Predicted log-scale bioactivity
	Mol_5494	2.7	313.10	2.50	0.94	-8.32	6.29
	Mol_3043	2.3	293.12	2.39	0.99	-8.31	6.64
	Mol_3884	2.0	326.17	3.73	0.90	-8.30	6.40
	Mol_1625	2.5	288.10	2.08	0.94	-8.25	6.32
	Mol_4056	0.9	341.06	3.92	0.65	-8.25	6.09
	Mol_5593	1.8	277.10	2.78	0.82	-8.25	6.24
	Mol_4091	2.9	345.14	2.66	0.89	-8.23	6.16

Molecule 2D structure	Molecule ID	log*P*	MW (Da)	SAS	RAscore	Docking score (kcal/mol)	Predicted log-scale bioactivity
	Mol_1702	2.5	291.11	2.53	0.97	-8.19	6.01
	Mol_2838	1.1	337.12	2.16	0.99	-8.18	6.28
	Mol_7478	1.3	347.11	2.43	0.91	-8.16	6.60
	Mol_3128	2.8	344.09	2.29	0.96	-8.15	6.32
	Mol_1735	1.6	337.03	2.59	0.98	-8.13	6.30
	Mol_5782	2.9	317.12	3.27	0.97	-8.11	6.79
	Mol_593	2.5	250.09	2.47	0.95	-8.11	6.38

Molecule 2D structure	Molecule ID	log*P*	MW (Da)	SAS	RAscore	Docking score (kcal/mol)	Predicted log-scale bioactivity
	Mol_665	2.2	295.14	2.81	0.98	-8.08	6.52
	Mol_1072	2.4	303.10	2.37	0.63	-8.06	6.33
	Mol_4672	2.4	341.11	2.11	0.99	-8.06	6.62
	Mol_6616	2.9	289.13	2.42	0.91	-8.01	6.32

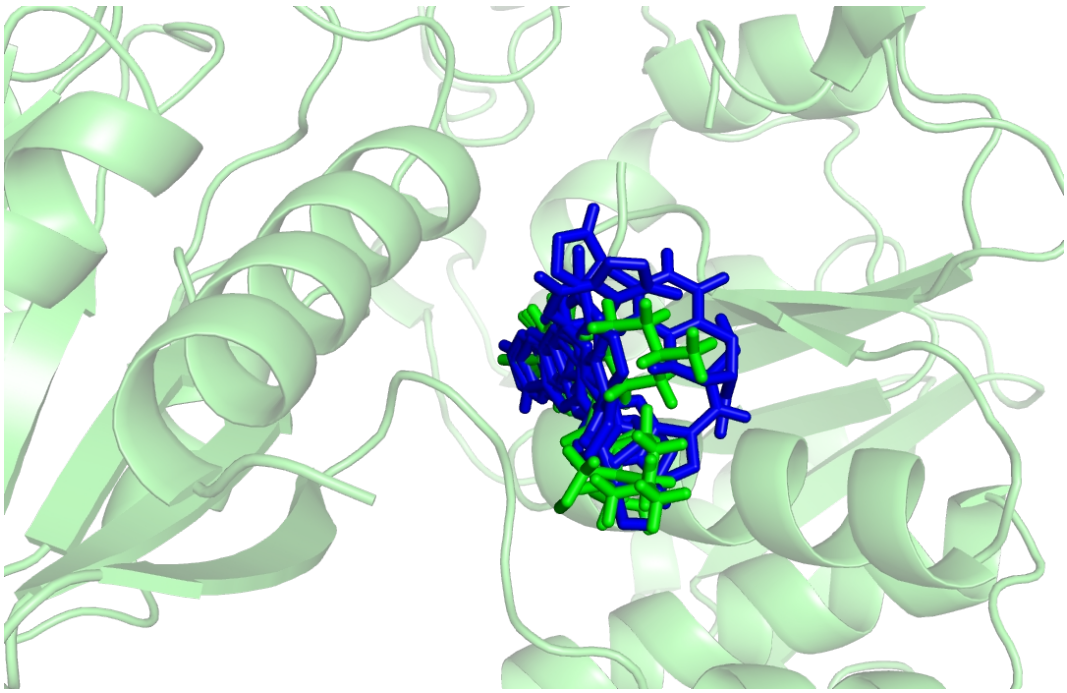


Figure 1: Overlap of the terminal ring systems of top 5 generated small molecules (blue sticks) and the fragments provided by organizers (green sticks) in the allosteric site of MurD ligase. The ring systems are constrained by the presence of Lys311 and Ile147 on the either sides of their planes.

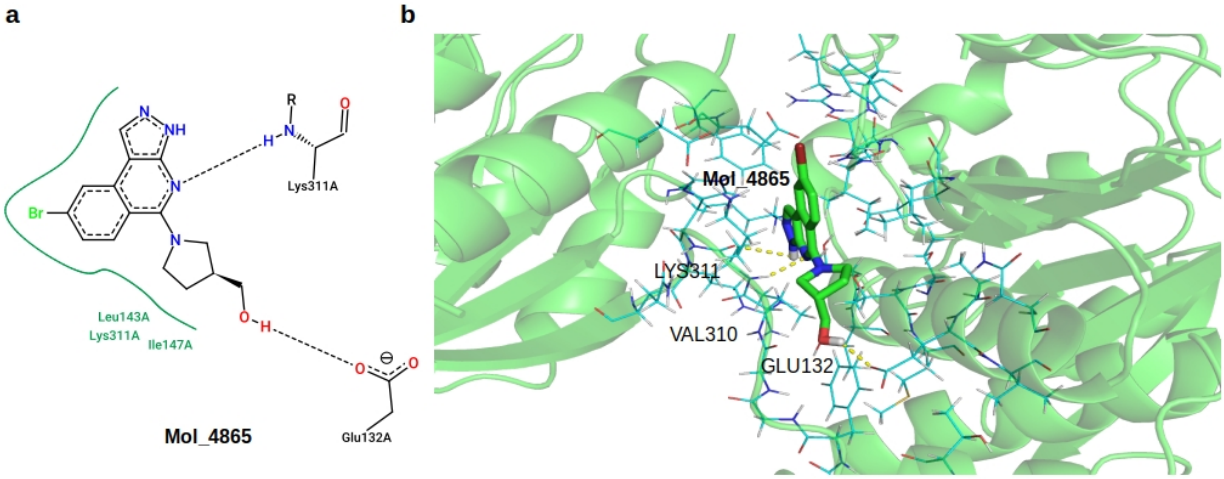


Figure 2: (**a**) 2D and (**b**) 3D interaction diagrams of the generated small molecule - Mol_4865 with the allosteric site residues of MurD ligase. The 2D interaction diagram was generated using the PoseView tool online.

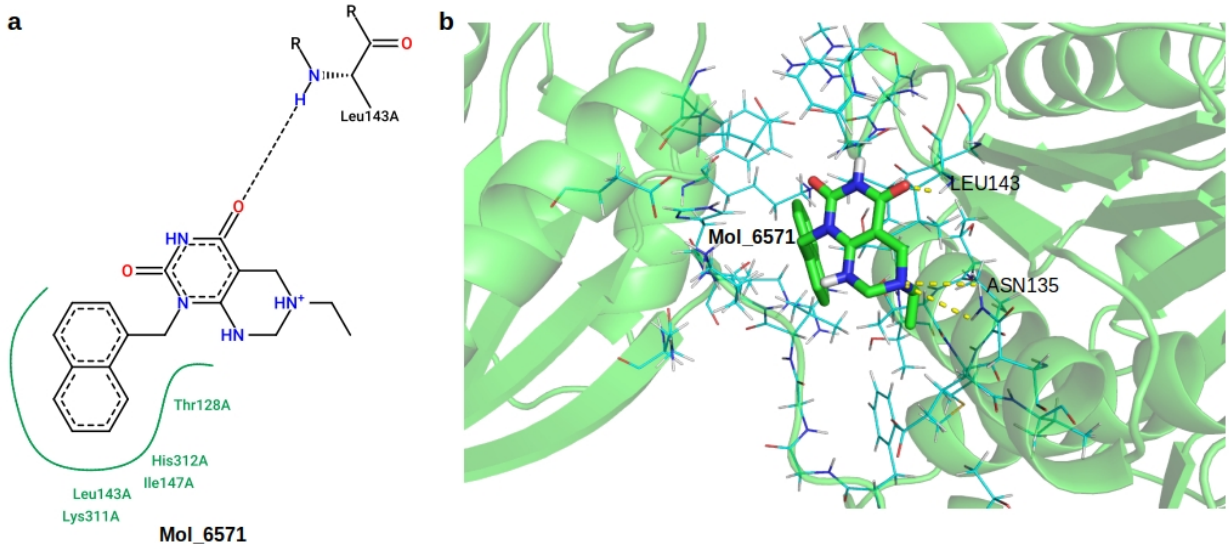


Figure 2: (**a**) 2D and (**b**) 3D interaction diagrams of the generated small molecule - Mol_6571 with the allosteric site residues of MurD ligase. The 2D interaction diagram was generated using the PoseView tool online.

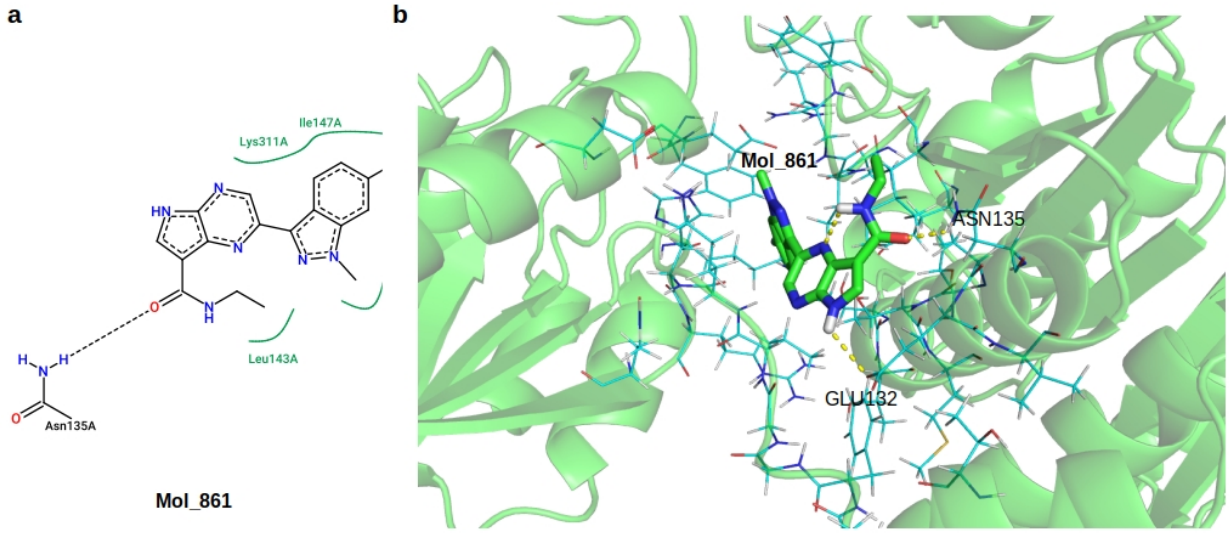


Figure 3: (**a**) 2D and (**b**) 3D interaction diagrams of the generated small molecule - Mol_861 with the allosteric site residues of MurD ligase. The 2D interaction diagram was generated using the PoseView tool online.

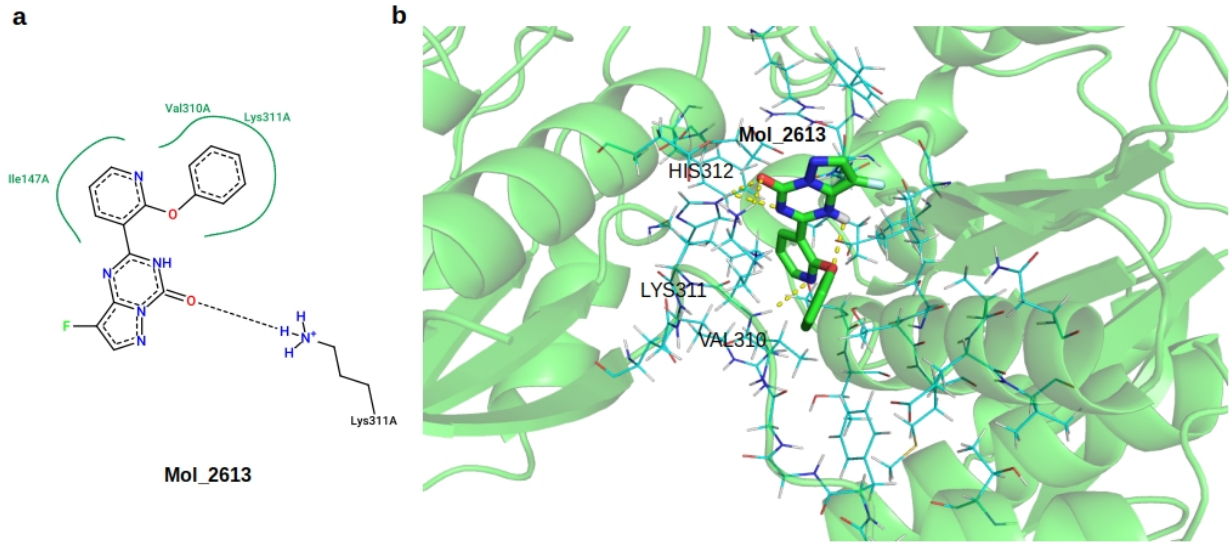


Figure 4: (**a**) 2D and (**b**) 3D interaction diagrams of the generated small molecule - Mol_2613 with the allosteric site residues of MurD ligase. The 2D interaction diagram was generated using the PoseView tool online.

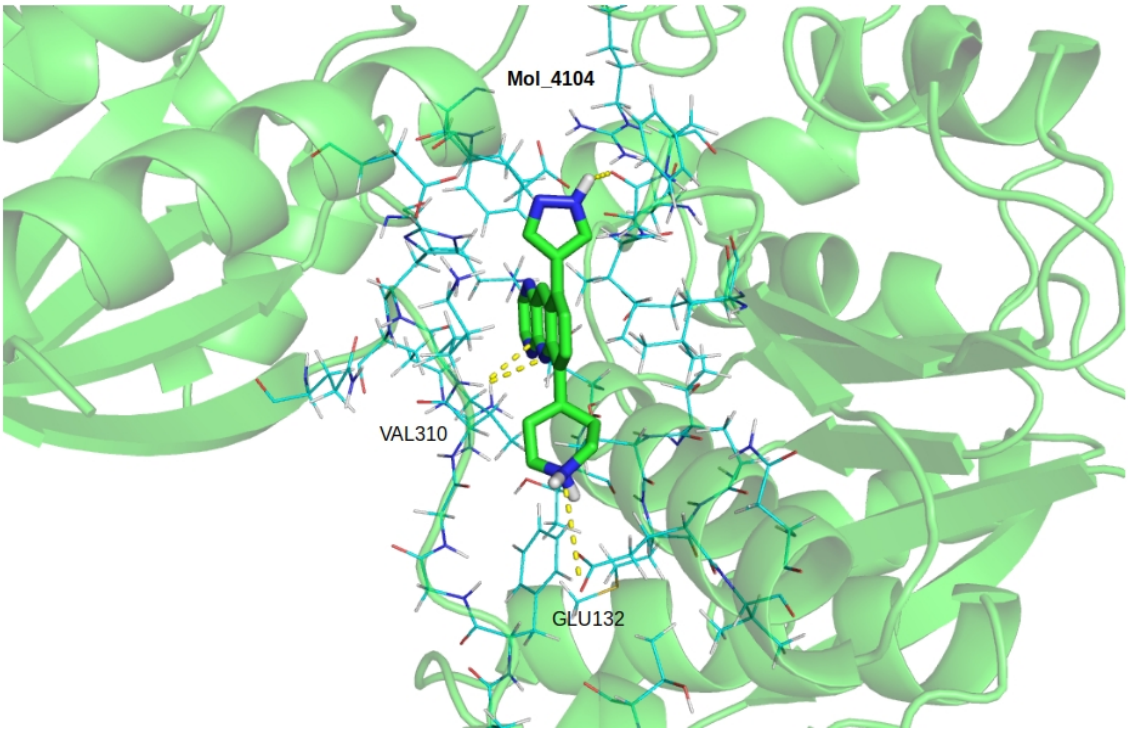


Figure 5: 3D interaction diagram of the generated small molecule - Mol_4104 with the allosteric site residues of MurD ligase.

References

Krishnan, S. R.; Bung, N.; Vangala, S. R.; Srinivasan, R.; Bulusu, G.; Roy, A. *De Novo* Structure-Based Drug Design Using Deep Learning. *J. Chem. Inf. Model.* **2021**. doi: 10.1021/acs.jcim.1c01319

Bung, N.; Krishnan, S. R.; Roy, A. An *In Silico* Explainable Multiparameter Optimization Approach for *De Novo* Drug Design against Proteins from the Central Nervous System. *J. Chem. Inf. Model.* **2022**. doi: 10.1021/acs.jcim.2c00462

Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **2009**, *1*, 8.

Thakkar, A.; Chadimová, V.; Bjerrum, E. J.; Engkvist, O.; Reymond, J. Retrosynthetic accessibility score (RAscore) - rapid machine learned synthesizability classification from AI driven retrosynthetic planning. *Chem. Sci.* **2021**, *12*, 3339-3349.

Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for their Exclusion in Bioassays. *J. Med. Chem*. **2010**, *53*, 2719−2740.

Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I. H.; Frearson, J.; Wyatt, P. G. Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. *ChemMedChem* **2008**, *3*, 435.

Doveston, R. G.; Tosatti, P.; Dow, M.; Foley, D. J.; Li, H. Y.; Campbell, A. J.; House, D.; Churcher, I.; Marsden, S. P.; Nelson, A. A Unified Lead-Oriented Synthesis of over Fifty Molecular Scaffolds. *Org. Biomol. Chem*. **2015**, *13*, 859−865.

McNutt, A. T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; Koes, D. R. GNINA 1.0: molecular docking with deep learning. *J. Cheminform*. **2021**, *13*, 43.

Appendix

Table S1: Simplified Molecular Input Line Entry System (SMILES) strings of the top 30 molecules shown in Table 1.

SMILES (Canonicalized using RDKit)	Molecule ID
c1cnc2nc3c(C4CC[NH2+]CC4)ccc(-c4cn[nH]c4)c3cc2n1	Mol_4104
CCN1CNc2c(c(=O)[nH]c(=O)n2Cc2cccc3ccccc23)C1	Mol_6571
CCNC(=O)c1c[nH]c2ncc(-c3nn(C)c4cc(F)ccc34)nc12	Mol_861
O=c1nc(-c2cccnc2Oc2ccccc2)[nH]c2c(F)cnn12	Mol_2613
OCC1CCN(c2nc3[nH]ncc3c3cc(Br)ccc23)C1	Mol_4865
c1ccc(C2CN(c3ncnc4[nH]ccc34)CC[NH2+]2)cc1	Mol_3952
c1ccc2c(Nc3ccc4nccn4n3)c3c[nH]nc3nc2c1	Mol_2740
c1ccc2c(C3CC[NH2+]CC3)c3[nH]ccc3nc2c1	Mol_5948
CC(=O)Nc1cccc(-c2nc3ccccc3c(=O)[nH]2)c1	Mol_4092
[NH3+]C1CCC(Nc2ncc3cnn(-c4cccc(F)c4)c3n2)CC1	Mol_7459
Oc1cccc(Nc2ncnc3ccc(O)cc23)c1	Mol_3508
Nc1nc2nn(CCO)cc2c2c1C(=O)c1ccccc1C2=O	Mol_128
c1ccc(-c2ncc3nnc(-c4c[nH]c5ncccc45)n3n2)cc1	Mol_5494
Cc1cc(NC(=O)c2cc(N)c(N)c3ncccc23)ccn1	Mol_3043
Nc1ncnc2c1c(Cc1ccc(F)cc1)cn2C1CCC([NH3+])C1	Mol_3884
Nc1ncc2c(=O)n(-c3ccccc3)c3ccccc3c2n1	Mol_1625
CC1CC(O)CC(=O)N1n1c(S)nc2cc(F)c(F)cc2c1=O	Mol_4056
Cc1n[nH]c2ncc3c(N)nc(-c4ccncc4)nc3c12	Mol_5593
Cc1nn(C)c(C(=O)NC2c3ccccc3Cc3ccccc32)c1C=O	Mol_4091
Oc1ccnc2[nH]cc(-c3cn(Cc4ccccc4)nn3)c12	Mol_1702
NC(=O)c1cccc(NCc2cccc(C(=O)Nc3nn[nH]n3)c2)c1	Mol_2838
Nc1nc2nn(-c3cccc(C(=O)Nc4ccccc4)c3)nc2c(=O)[nH]1	Mol_7478
CCc1nc2cccc(F)c2c(=O)n1-c1ccc(CC(=O)O)cc1F	Mol_3128
Cc1ccc(NS(=O)(=O)c2ccc3c(c2)B(O)OC3)cc1Cl	Mol_1735
Cc1n[nH]c2ncc(C(=O)N3CCC(C#N)c4ccccc43)cc12	Mol_5782
c1ccc2c(Nc3ncnc4[nH]ccc34)n[nH]c2c1	Mol_593
OC1CCN(c2nc(Nc3ccccc3)c3cc[nH]c3n2)C1	Mol_665
Cc1cccc2cc3c(=O)[nH]c(=O)nc-3n(-c3ccccc3)c12	Mol_1072
Cc1nc2ccccc2n1CCOC(=O)c1ccc(C(N)=O)c(F)c1	Mol_4672
Cc1nc(N)c2nc(Cc3cccc4ccccc34)[nH]c2n1	Mol_6616