AI-assisted molecular design of quinolone alkaloid analogues as pancreatic lipase inhibitor candidates

Hayato Suzuki[1], *Yoh* Noguchi[1,2*], *Kaito* Watanabe[1], *Chiawen* Ying[1], *Motokuni* Nakajima[1], *Ryota* Morikawa[1], *Yukiko* Matsuo[3] and *Masako* Takasu[4]

[1]Tokyo University of Pharmacy and Life Sciences, School of Life Sciences, Japan.
[2]The Institute of Statistical Mathematics, Japan.
[3]Tokyo University of Pharmacy and Life Sciences, School of Pharmacy, Japan.
[4]Tokyo Woman's Christian University, Japan.

Abstract. Pancreatic lipase (PL) is a key enzyme in dietary fat absorption and a validated target for preventing obesity. However, the existing PL inhibitors, such as orlistat, have side effects. We developed a molecular design workflow that integrates machine-learning-based generation with iterative docking simulations, focusing on quinolone-alkaloid-like scaffolds. The docking score of the best candidate decreased from -8.0 to -13.5 kcal/mol over 60 optimization cycles. The top-scoring structure contained a polycyclic aromatic core linked via conjugated chains, with docking poses indicating π–π stacking with Tyr114 and Phe215, consistent with a pocket-blocking mechanism that impeded substrate entry. Additionally, we observed a polar approach near Ser152. The generator produced chemotypes that resembled known inhibitors despite starting from quinolone alkaloid scaffolds. Although we focused on docking-based optimization, these findings demonstrate the potential of AI-assisted molecular design for identifying novel PL inhibitors. Further validation is required, including systematic structure–activity analyses and experimental testing.

Introduction

Artificial intelligence (AI) is increasingly being used for drug discovery. AI-driven drug discovery is expected to increase the development speed of new drugs, reduce costs, and increase success rates. For instance, generative AI was used to identify candidate compounds for treating of idiopathic pulmonary fibrosis, resulting in a preclinical candidate in one year, which advanced to the clinical stage [1].

Obesity contributes to numerous diseases that are among the leading causes of mortality, and the global prevalence of obesity is increasing [2,3]. Standard obesity treatment typically includes behavioral therapy such as exercise and dietary restrictions, although pharmacotherapy is also used in severe cases. GLP-1 receptor agonists, such as semaglutide, have attracted considerable attention as obesity treatments [4]. GLP-1 receptor agonists induce weight loss through mechanisms such as appetite suppression, delayed gastric emptying, and enhanced insulin secretion. However, the development costs of peptide-based

*Yoh Noguchi: ynoguchi@toyaku.ac.jp

therapeutics tend to be high. In contrast, low-molecular-weight compounds act as pancreatic lipase inhibitors.

Pancreatic lipase is mainly secreted by the pancreas, and pancreatic lipase inhibitors promote weight loss via inhibiting fat absorption. Lipases break down triglycerides into monoglycerides and fatty acids [5]. Inhibiting pancreatic lipase activity is particularly effective in treating obesity. Orlistat is currently approved as a pharmaceutical for obesity. Orlistat forms a covalent bond with the Ser residue in the active sites of gastric and pancreatic lipases, effectively and irreversibly inactivating the enzyme [6]. Orlistat inhibits the absorption of approximately one-third of dietary fat; however, orlistat also has various side effects such as fatty stools. Therefore, further studies are ongoing to identify novel pancreatic lipase inhibitors; for example, virtual screening studies have focused on flavonoids, such as chrysin, genistein, and naringenin [7]. In addition, bromhexine is an expectorant and an example of a drug repositioned to inhibit pancreatic lipase [8].

Matsuo et al. [9] isolated a series of novel quinolone alkaloids, such as evocarpine, as well as the indole alkaloid evodiamine, from methanolic goshuyu extracts. Evodiamine has been studied using molecular docking and molecular dynamics simulations in phosphodiesterase-5 (PDE5) [10, 11]. Quinolone alkaloids were characterized as pancreatic lipase inhibitors that suppress lipase activity in vitro, and the administration of evocarpine to mice led to reduced serum triglyceride levels [9].

Therefore, we applied AI-driven drug discovery for molecular design, starting with isolated quinolone alkaloids. The accuracy of these initial inhibitory activity predictions was expected to be low because experimental data on quinolone alkaloids are scarce. Thus, we adopted the framework of Nanjo et al. [12], which combines machine learning predictions and simulations for highly extrapolative exploration, and applied it to the search for molecular structures surrounding quinolone alkaloids as potential pancreatic lipase inhibitors. However, the binding affinities estimated using docking simulations do not necessarily strongly correlate with pancreatic lipase inhibitory activity. Some compounds selected based on docking results have exhibited strong inhibitory activity [7, 13–15]. We thus prioritized the exploration of molecular structures with small docking scores as the first step in primary screening.

Methods

2.1 Docking simulation

AutoDock Vina 1.2.5 [16] was used for docking simulation. The receptor file was created in UCSF Chimera [17] using the structure of the human pancreatic lipase and colipase complex (PDB ID: 1LPB [5]) with water and ligands removed. The 3D ligand structures were generated using SMILES, and energy-minimized conformers were used in Open Babel [18]. Gasteiger charges were assigned.

The center coordinates of the binding site (x: 9.82 Å, y: 23.50 Å, z: 50.87 Å) and the size of the docking box (35 × 35 × 35 Å) were set to cover the pancreatic lipase binding pocket. Thirteen residues were designated as flexible: Phe77, Ile78, Asp79, Tyr114, His151, Ser152, Leu153, Ile209, Leu213, Phe215, Arg256, His263, and Leu264. These residues interact with aurone derivatives according to virtual screening results [14]. Although Asp176 is part of the catalytic triad, direct ligand contact was not observed with this residue [14]. Asp176 was not included among the flexible residues in the docking runs to reduce computational cost. The computational parameters were as follows: the scoring function, num_modes, energy_range, and exhaustiveness were Vina, 100, 3 kcal/mol, and 64, respectively.

2.2 Molecular design

Figure 1 illustrates the molecular exploration workflow. A prediction model was constructed using training data on the molecular structures and their docking scores. Virtual molecular structures were sampled according to the predicted docking scores to obtain the structures on which the docking simulations were performed. The resulting structures and docking scores were added to the training data, and this cycle was repeated 60 times. The initial dataset in the first cycle comprised docking results for 33 molecules, including 14 quinolone alkaloids isolated by Matsuo et al. [9] and their virtual stereoisomers generated using RDKit via the aforementioned procedure.

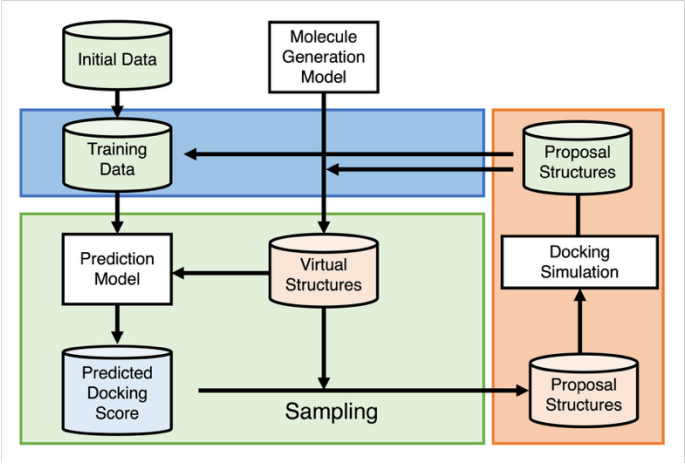

Fig. 1. Schematic diagram of molecular design workflow. Cylinders represent data, and rectangles represent models or calculations. Green contains molecular structures and physical property data, blue contains physical property data only, and orange indicates structural data only.

Virtual molecular structures were generated using an N-gram model-based molecular generator implemented in XenonPy [19, 20]. This generator learns the sequence patterns of SMILES strings and produces new molecules based on the learned patterns. The training set consisted of 85 molecular structures reported in a 2021 review of pancreatic lipase inhibitors [21] and 1,174 quinolone alkaloid structures retrieved from PubChem [22]. A total of 30 randomly rooted SMILES strings were generated for each molecule because a SMILES string representation depends on the root atom chosen, yielding 37,770 strings. The model was trained using an N-gram model with subsequences of up to 10 tokens, meaning that the model considered token sequences ranging from unigrams to 10-grams when learning the sequence patterns. Tokens 1–5 were randomly deleted during generation from the end of a SMILES string, after which the N-gram model was used to reconstruct the truncated portion. Generated structures with a molecular weight of ≥ 600, lacked a quinolone alkaloid scaffold, or contained ring systems other than five- or six-membered rings were excluded.

The docking scores were predicted using XenonPy and Bayesian ridge regression. The explanatory variables were a combination of the feature-class circular fingerprint (radius 3 and 2048 bits) [23] and MACCS keys (166 bits) [24]. Using the probabilistic predictions obtained from Bayesian ridge regression, the Probability of Improvement (PI) for each structure s_i was estimated:

$$PI(s_i) = \int_{-\infty}^{y_{\min}} \frac{1}{\sqrt{2\pi\sigma^2(s_i)}} \exp\left\{-\frac{\left(x-\mu(s_i)\right)^2}{\sigma^2(s_i)}\right\} dx \qquad (1)$$

where $\mu(s_i)$ and $\sigma^2(s_i)$ are the predicted mean and variance of the docking score, respectively; and $y_{\min}$ is the highest docking score observed in the training data. The sampling probability for selecting molecules for the docking simulation is defined as follows:

$$p(s_i) = \frac{PI(s_i)^{\beta_k}}{\sum PI(s_i)^{\beta_k}} \quad (2)$$

where β_k is the sampling parameter for cycle k. A β_k value of 0 indicates the sampling is uniform; larger β_k values bias the selection toward molecules with higher $PI(s_i)$. β_k was assigned a value of 0 and 1 for $k \leq 30$ and $k > 50$, respectively, and was increased exponentially from 31 to 50 with NumPy's logspace function.

2.3 Analysis

The generated structures were dimension-reduced using Uniform Manifold Approximation and Projection (UMAP) [25]. The metric was set to Jaccard. Clustering was performed using the *K*-means method, with the clusters divided into six groups. The four structures closest to the center of gravity of each cluster were selected as representative structures.

The synthetic accessibility (SA) scores [26] were used as indicators of synthetic difficulty, ranging from 1 (easy to synthesize) to 10 (very difficult); lower values indicated easier synthesis. The quantitative estimate of drug-likeness (QED) score [27] is a composite desirability index over multiple molecular properties: molecular weight, predicted LogP [28], number of hydrogen bond acceptors, number of hydrogen bond donors, polar surface area, number of rotatable bonds, and AROM (number of aromatic rings). The QED values range from 0 to 1, with larger values reflecting a higher degree of drug likeness. Protein and ligand structures were visualized using PyMOL [29] and LigPlot+ ver. 2.2.8 [30].

Results

The 956 virtual molecular structures generated after 60 cycles were projected onto three dimensions using UMAP (Fig. 2(a)). Figure 2(b) shows representative structures of the six clusters. The structures of Clusters 1, 2, and 3 were similar to those of goshuyu-derived quinolone alkaloids. Cluster 3 tended to have larger side chains than Clusters 1 and 2. Clusters 4, 5, and 6 had a quinolone alkaloid scaffold as the core with long carbon chains linked to bicyclic aromatic moieties.

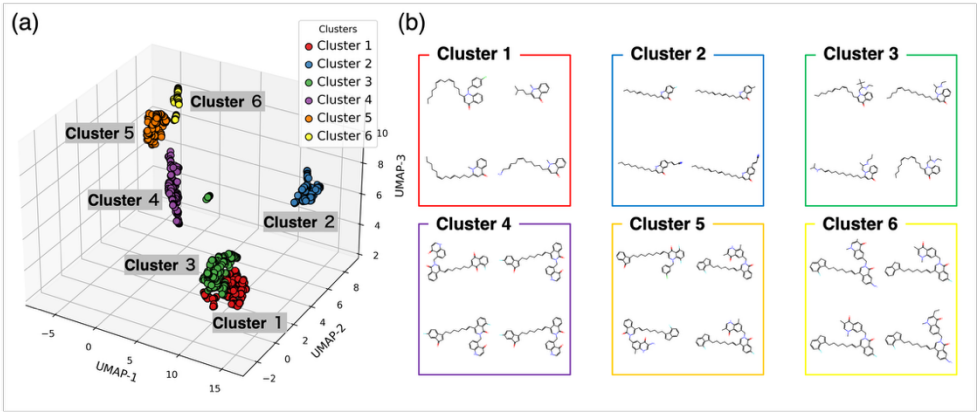

Fig. 2. Generated virtual molecular structures. (a) Embedding of six clusters in 3D space using UMAP; (b) representative structures of each cluster.

Figure 3 shows the docking simulation results for the virtual structures sampled in each cycle. We started with structures with an initial docking score of approximately -8 kcal/mol and finally obtained a molecular structure with a score of -13.5 kcal/mol over 60 cycles of iterative molecular design and screening. The clustering of the output structures for each cycle showed that the structural groups produced clear transition patterns as the cycles progressed. Structures belonging to Clusters 1–3 were generated during the initial stages, and were structurally diverse. However, all structures were replaced by those of Cluster 4 in the 41st cycle after structures from Cluster 4 appeared in the 37th cycle. This transition showed that molecular structures similar to those in Cluster 4, which had the lowest docking score, were easily sampled when the β value was sufficiently large, leading to convergence toward optimization. Subsequently, the structures gradually transitioned to Clusters 5 and 6.

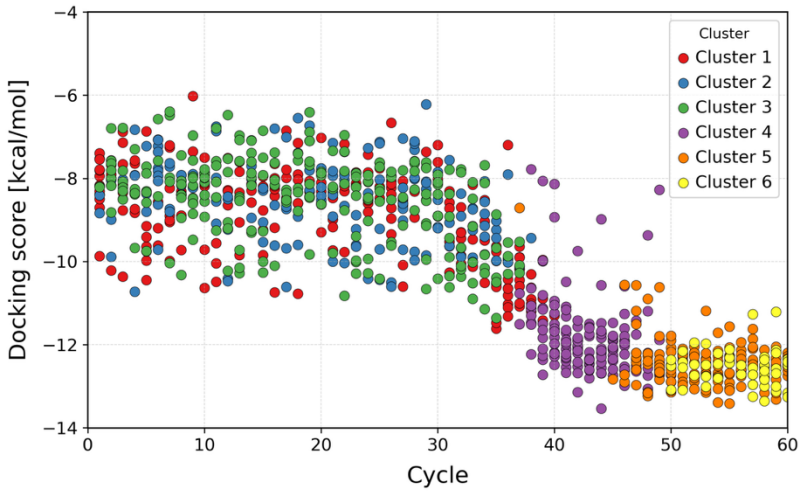

Fig. 3. Changes in docking scores with cycles. The simulations generated 100 poses for each virtual molecular structure sampled in a given docking cycle, from which the lowest (most favorable) docking score was plotted as the representative value. The color corresponds to the cluster assignment of each structure.

Figure 4(a) shows the SA and QED scores of the generated molecules. This result indicated that Clusters 4–6 had good (more negative) docking scores and low QED values. A breakdown of the QED scores is shown in Fig. 4(b). Min–max scaling was applied because the scale of each feature widely varied. Clusters with high docking scores likely had low QED values because of their structures containing large number of aromatic rings, which resulted in a large molecular weight. The SA scores indicate synthetic difficulty and fell in the range of 3–4, which do not necessarily suggest synthetic difficulty. However, no similar compounds were found in any database, such as PubChem or ChEMBL, suggesting that their synthesis may be challenging.

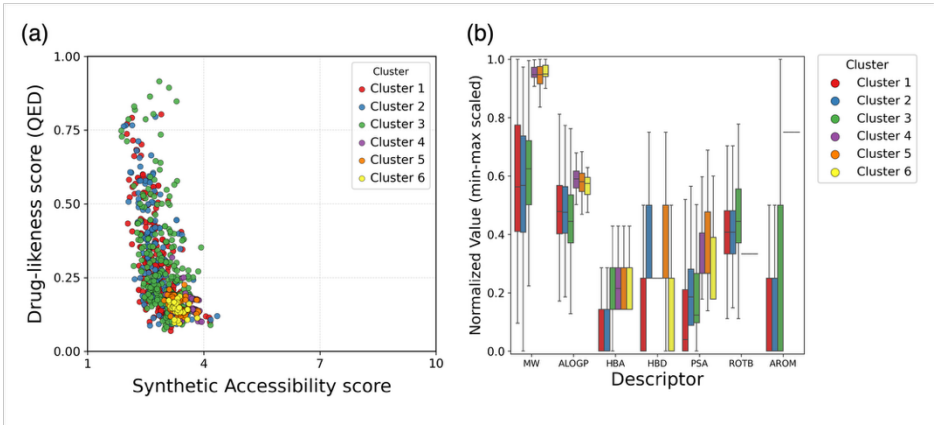

Fig. 4. Synthetic accessibility (SA) and quantitative estimate of drug-likeness (QED of each cluster. (a) Distributions of SA and QED scores. (b) Box plots of QED score features for each cluster. Zero variance resulted in values being displayed as "–" for some features, e.g., number of rotatable bonds and number of aromatic rings in Clusters 4–6.

The virtual structure belonging to Cluster 4 had the lowest docking score of -13.536 kcal/mol. Figure 5 shows the docking pose and interaction analyses. The compound possesses a quinolone alkaloid scaffold with a long conjugated alkenyl chain linked to a polycyclic aromatic moiety and fits well into the docking pocket (Fig. 5(a)). Figure 5(b) shows that the side chain of the catalytic residue Ser152 is located in close proximity to the fluorine atom of the ligand, whereas the side chain of Arg256 forms a hydrogen-bond-like interaction with its carbonyl oxygen. The pocket is lined with residues: Phe77, Ile78, Asp79, Tyr114, His151, Leu153, Ala178, Pro180, Phe215, Leu264, and Tyr267.

Figure 5(c) shows the spatial relationship between the aromatic rings of the ligand and the surrounding aromatic residues. Tyr114 and Phe215 are positioned nearly parallel to ring A (C18–C19–C20–C25–C26–C27) and ring B (C20–C21–C22–C23–C24–C25) of the ligand, respectively. The centroid–centroid distances are 3.8 Å (ring A, Fig. 5(c) yellow dashed line) and 4.0 Å (ring B, Fig. 5(c) magenta dashed line) for Tyr114 as well as 4.1 Å (ring A, Fig. 5(c) cyan dashed line) and 3.8 Å (ring B, Fig. 5(c) orange dashed line) for Phe215. The in-plane offsets are approximately 1.2–2.0 Å, consistent with the geometric criteria characteristic of parallel-displaced π–π stacking [31].

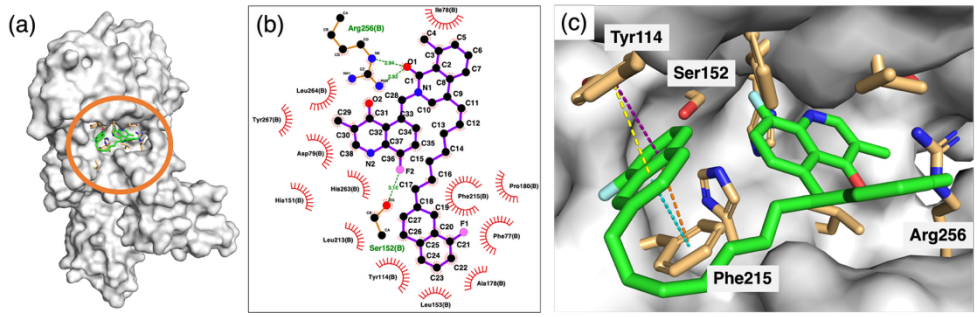

Fig. 5. The docking pose of the top-scoring structure. (a) Overview of the docking pose. The orange circle, green, and brown indicate the docking region, ligands, and flexible residues, respectively. (b) Residues interacting with the ligand were visualized with LigPlot+. (c) Spatial arrangement of the ligand aromatic rings and surrounding aromatic residues. Dashed lines indicate centroid connections: yellow, ring A–Tyr114; cyan, ring A–Phe215; purple, ring B–Tyr114; orange, ring B–Phe215.

Discussion

Our results of docking suggests an entrance-blocking mechanism dominated by aromatic contacts in the portal region, narrowing access to the catalytic triad (Ser152, Asp176, and His263). This interaction pattern qualitatively agrees with those reported for pancreatic lipase inhibitors that exploit Phe77, Tyr114, and Phe215 [8, 15]. We also observed polar approach near Ser152, where the fluorinated terminus of the ligand is positioned in close proximity to the Oγ atom of Ser152.

We started our docking analyses with simple quinolone alkaloids such as evocarpine. The generator sampled chemotypes qualitatively reminiscent of reported lipase inhibitors, namely, a quinolone-like core extended with a long aliphatic chain with shape complementarity. An additional simulation with a different random seed similarly converged to a closely related chemotype with the highest docking score of -11.5 kcal/mol, suggesting a reproducible outcome despite different initializations. The initial docking runs favored relatively simple quinolone derivatives with short side chains; subsequent optimization consistently favored bulkier aromatic chemotypes. This trend is consistent with the long-chain preferences of native triglyceride substrates and several reported inhibitors [6, 14, 15].

In future studies, we plan to conduct more systematic structure–activity relationship analyses of the virtual chemotypes we identified to further understand the docking-derived trends. Our findings are based on static docking into a single X-ray structure (ligand-bound conformation), and AutoDock Vina scores provide only approximate estimates of binding free energies. Therefore, any quantitative conclusions should be drawn with caution. Additional validation of our results is required through, for example, examining docking to MD-relaxed receptors or using approximate free-energy estimation with methods such as MM/GBSA. The final confirmation of the suitability of the structure for blocking pancreatic lipase requires in vitro inhibition assays. We are considering replacing the current generator with a molecular design framework that explicitly accounts for synthetic accessibility to further increase the practical relevance of the structure selection for subsequent experimental validation [32].

Conclusion

We identified a structure with a docking score for the active site of the target human pancreatic lipase that decreased from -8 to -13 kcal/mol over 60 iterations of molecular design and virtual screening, demonstrating the efficacy of the sequential learning and sampling strategies used in this study. Furthermore, the analysis of the interactions of the lowest-scoring structure revealed interaction patterns similar to those for known pancreatic lipase inhibitors. However, all structures with best (most negative) docking scores had low QED scores, indicating a need for further study. Although the SA scores were not high, a high SA score does not necessarily indicate high synthetic feasibility. Therefore, a molecular generator that explicitly accounts for synthetic accessibility should be introduced in future in vitro validations.

This study was supported by JSPS KAKENHI, grant number 24K20895 (Grant-in-Aid for Early-Career Scientists). H.S. and Y.N. conceived and designed this study. H.S. and C.Y. developed the methodology; H.S., Y.N., and K.W. implemented the software. Validation was performed by H.S., Y.N., and M.N.; formal analysis was conducted by H.S.; Y.N. performed the investigation. M.T. and R.M. provided the resources. H.S., C.Y., and K.W curated the data. The original draft was prepared by Y.N. and M.T., and all authors contributed to the review and editing of the manuscript. Figures and visualizations were generated by Y.N. and H.S. Supervision was provided by Y.M., R.M., and M.N.

Project administration and funding acquisition were managed by Y.N. and M.T. All authors have read and approved the final manuscript.

References

1. F. Ren, A. Aliper, J. Chen, H. Zhao, S. Rao, C. Kuppe, I. V. Ozerov, M. Zhang, K. Witte, C. Kruse, V. Aladinskiy, Y. Ivanenkov, D. Polykovskiy, Y. Fu, E. Babin, J. Qiao, X. Liang, Z. Mou, H. Wang, F.W. Pun, P. Torres-Ayuso, A. Veviorskiy, D. Song, S. Liu, B. Zhang, V. Naumov, X. Ding, A. Kukharenko, E. Izumchenko, A. Zhavoronkov, A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat. Biotechnol.* **43**, 63 (2025). https://doi.org/10.1038/s41587-024-02143-0
2. GBD 2021 Risk Factors Collaborators, Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **403**, 2162 (2024). https://doi.org/10.1016/S0140-6736(24)00933-4
3. GBD 2021 Adult BMI Collaborators, Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. *Lancet* **405**, 813 (2025). https://doi.org/10.1016/S0140-6736(25)00355-1
4. Z. Zheng, Y. Zong, Y. Ma, Y. Tian, Y. Pang, C. Zhang, J. Gao, Glucagon-like peptide-1 receptor: mechanisms and advances in therapy. *Signal Transduct. Target. Ther.* **9**, 234 (2024). https://doi.org/10.1038/s41392-024-01931-z
5. M.-P. Egloff, F. Marguet, G. Buono, R. Verger, C. Cambillau, H. van Tilbeurgh, The 2.46 Å resolution structure of the pancreatic lipase–colipase complex inhibited by a C11 alkyl phosphonate. *Biochemistry* **34**, 275 (1995). https://doi.org/10.1021/bi00009a003
6. A. M. Heck, J. A. Yanovski, K. A. Calis, Orlistat, a new lipase inhibitor for the management of obesity. *Pharmacotherapy* **20**, 270 (2000). https://doi.org/10.1592/phco.20.4.270.34882
7. Y. Zhai, K. Wang, Z. Yu, S. Zhou, J. Fan, Pancreatic lipase inhibitors: virtual screening and mechanistic analysis. *Int. J. Biol. Macromol.* **310**, 143128 (2025). https://doi.org/10.1016/j.ijbiomac.2025.143128
8. A. Gholami, D. Minai-Tehrani, L.A. Eriksson, Combining kinetics and in silico approaches to evaluate bromhexine as an anti-pancreatic lipase agent for obesity management. *Sci. Rep.* **15**, 18420 (2025). https://doi.org/10.1038/s41598-025-02625-4
9. Y. Matsuo, T. Nozaki, Y. Kamewada, M. Nakagawa, Y. Nakamura, N. Inaba, Y. Mimaki, New quinolone alkaloids from Euodia fruit, and their pancreatic lipase inhibitory and PPAR-γ ligand-binding activities. *Fitoterapia* **180**, 106322 (2025). https://doi.org/10.1016/j.fitote.2024.106322
10. S. Tripathi, R. H. Cote and H. Vashisth, Coupling of conformational dynamics and inhibitor binding in the phosphodiesterase-5 family, *Protein Sci.* **32**, e4720 (2023). https://doi.org/10.1002/pro.4720
11. A. Kobayashi, M. Nakajima, Y. Noguchi, R. Morikawa, Y. Matsuo, and M. Takasu, Molecular Dynamics Simulation of the Complex of PDE5 and Evodiamine. *Life* **13**(2), 578 (2023). https://doi.org/10.3390/life13020578
12. S. Nanjo, Arifin, H. Maeda, Y. Hayashi, K. Hatakeyama-Sato, R. Himeno, T. Hayakawa, R. Yoshida, SPACIER: on-demand polymer design with fully automated

all-atom classical molecular dynamics integrated into machine learning pipelines. *NPJ Comput. Mater.* **11**, 16 (2025). https://doi.org/10.1038/s41524-024-01492-3

13. S. Rocha, C. Proença, A.N. Araújo, M. Freitas, I. Rufino, N. Aniceto, A.M.S. Silva, F. Carvalho, R.C. Guedes, E. Fernandes, Flavonoids as potential modulators of pancreatic lipase catalytic activity. *Pharmaceutics* **17**, 163 (2025). https://doi.org/10.3390/pharmaceutics17020163
14. P. T. V. Nguyen, H. A. Huynh, D. V. Truong, T.-D. Tran, C.-V. T. Vo, Exploring aurone derivatives as potential human pancreatic lipase inhibitors through molecular docking and molecular dynamics simulations. *Molecules* **25**, 4657 (2020). https://doi.org/10.3390/molecules25204657
15. C.-T. Vo, T. T. Nguyen, T. N. Danga, M. Q. Dao, V. T. Vo, O. T. Tran, L. T. Vu, T.-D. Tran, Design, synthesis, biological evaluation and molecular docking of alkoxyaurones as potent pancreatic lipase inhibitors, *Bioorg. Med. Chem. Lett*. **98** 129574 (2024). https://doi.org/10.1016/j.bmcl.2023.129574
16. 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**, 3891 (2021). https://doi.org/10.1021/acs.jcim.1c00203
17. E. F. Pettersen, T. D. Goddard, 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**, 1605 (2004). https://doi.org/10.1002/jcc.20084
18. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox. *J. Cheminform*. **3**, 33 (2011). https://doi.org/10.1186/1758-2946-3-33
19. S. Wu, G. Lambard, C. Liu, H. Yamada, R. Yoshida, iQSPR in XenonPy: a Bayesian molecular design algorithm. *Mol. Inf*. **39**, 1900107 (2020). https://doi.org/10.1002/minf.201900107
20. H. Ikebata, K. Hongo, T. Isomura, R. Maezono, R. Yoshida, Bayesian molecular design with a chemical language model. *J. Comput. Aided Mol. Des*. **31**, 379–391 (2017). https://doi.org/10.1007/s10822-016-0008-z
21. A. Kumar, S. Chauhan, Pancreatic lipase inhibitors: the road voyaged and successes. *Life Sci*. **271**, 119115 (2021). https://doi.org/10.1016/j.lfs.2021.119115
22. S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, PubChem 2023 update. *Nucleic Acids Res.* **51**, D1373(2023). https://doi.org/10.1093/nar/gkac956
23. D. Rogers, M. Hahn, Extended-connectivity fingerprints. *J. Chem. Inf. Model*. **50**, 742–754 (2010). https://doi.org/10.1021/ci100050t
24. J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse, Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci*. **42**, 1273 (2002). https://doi.org/10.1021/ci010132r
25. L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426v3 (2018). https://doi.org/10.48550/arXiv.1802.03426
26. P. Ertl, A. Schuffenhauer, Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform*. **1**, 8 (2009). https://doi.org/10.1186/1758-2946-1-8

27. G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A.L. Hopkins, Quantifying the chemical beauty of drugs. *Nat. Chem*. **4**, 90 (2012). https://doi.org/10.1038/nchem.1243
28. A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of ALOGP and CLOGP methods. *J. Phys. Chem. A* **102**, 3762 (1998). https://doi.org/10.1021/jp980230o
29. Schrödinger, L., W. DeLano, PyMOL. (2020). http://www.pymol.org/pymol
30. R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model*. **51**, 2778 (2011). https://doi.org/10.1021/ci200227u
31. C.A. Hunter, J.K.M. Sanders, The Nature of π–π Interactions, *J. Am. Chem. Soc.* 112, 5525 (1990). https://doi.org/10.1021/ja00170a016
32. S. Nakamura, N. Yasuo, M. Sekijima, Molecular optimization using a conditional transformer for reaction-aware compound exploration with reinforcement learning, *Commun. Chem.* **8**, 40 (2025). https://doi.org/10.1038/s42004-025-01437-x