A Experimental Details

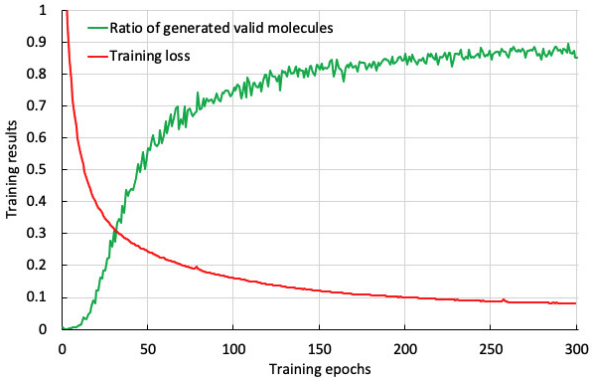

Fig. A.1. Training loss and the ratio of valid molecules generated by the proposed Gx2Mol. The red curve indicates the training loss value of the LSTM with the training epochs. The green curve denotes the ratio of valid molecules generated by the LSTM with the training epochs. Note that the valid molecules are examined by RDKit.

Validity denotes the ratio of valid molecules to the total number of training SMILES strings. In practice, this measure is calculated using RDKit tool.

B Calculation of Molecular Diversity

To evaluate the structural diversity among the newly generated molecules, we computed pairwise Tanimoto distances based on circular molecular fingerprints. Specifically, each molecule was first converted into a binary fingerprint using the Morgan algorithm (radius = 2, 2048 bits) as implemented in RDKit. For each pair of molecules, the Tanimoto similarity between their fingerprints was calculated, and the corresponding Tanimoto distance was defined as

$$\text{Diversity} = \frac{2}{N(N-1)} \sum_{i<j} (1 - \text{Tanimoto}(i,j)) \tag{6}$$

The diversity of a set of molecules was then measured by the average of all pairwise distances. In addition, we report the maximum, minimum, and standard deviation of these distances to provide a comprehensive view of diversity distribution within each generated set.

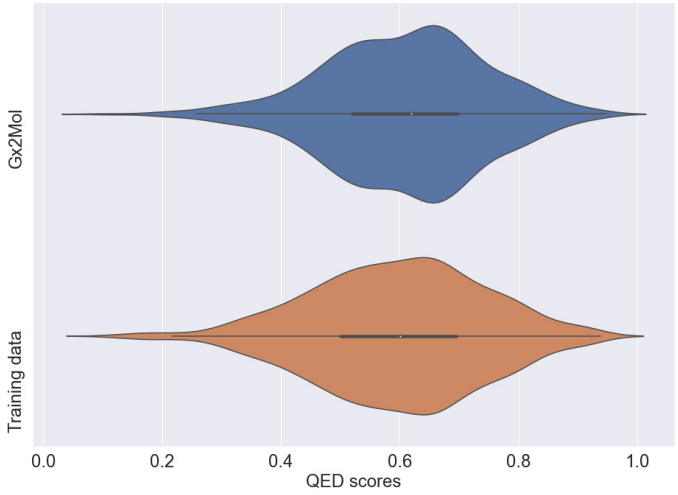

Fig. A.2. Violin plots of QED scores for molecules from the training dataset and Gx2Mol.

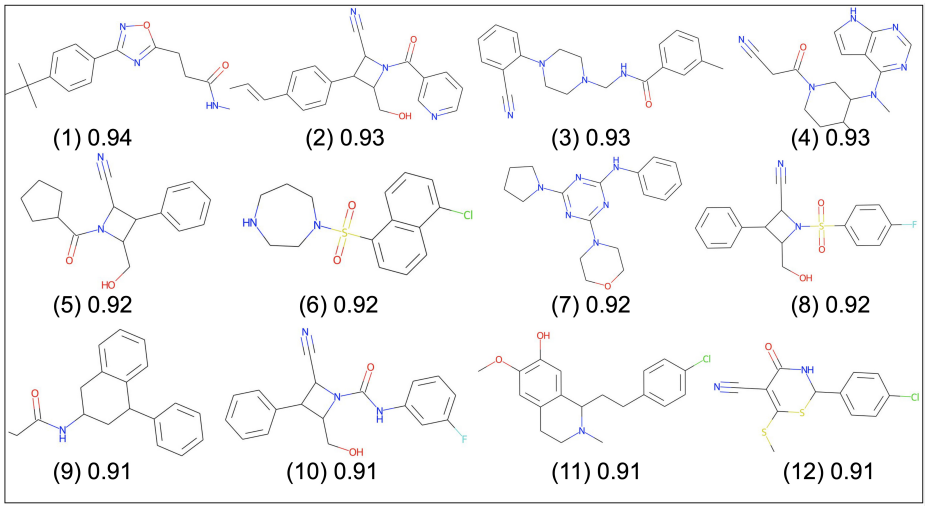

Fig. A.3. Top-12 molecular structures and QED scores for molecules in the training dataset.

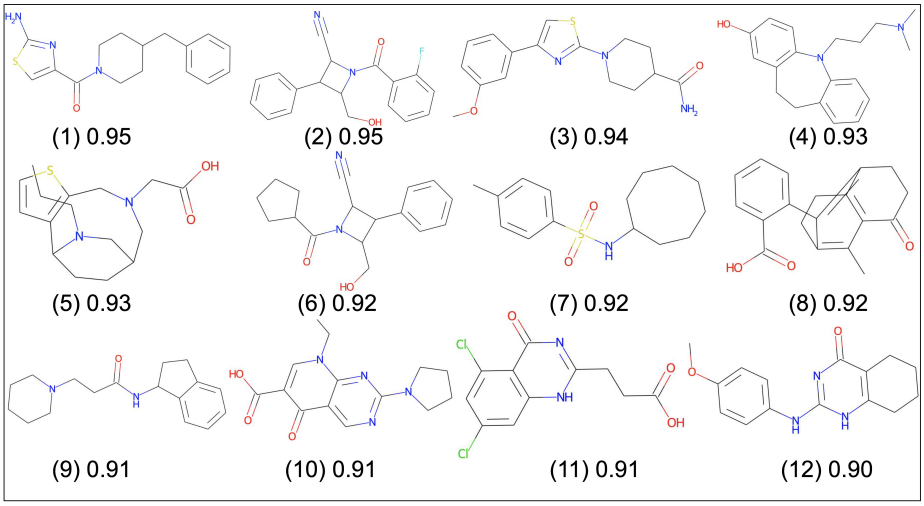

Fig. A.4. Top-12 molecular structures with QED scores for molecules generated by Gx2Mol.

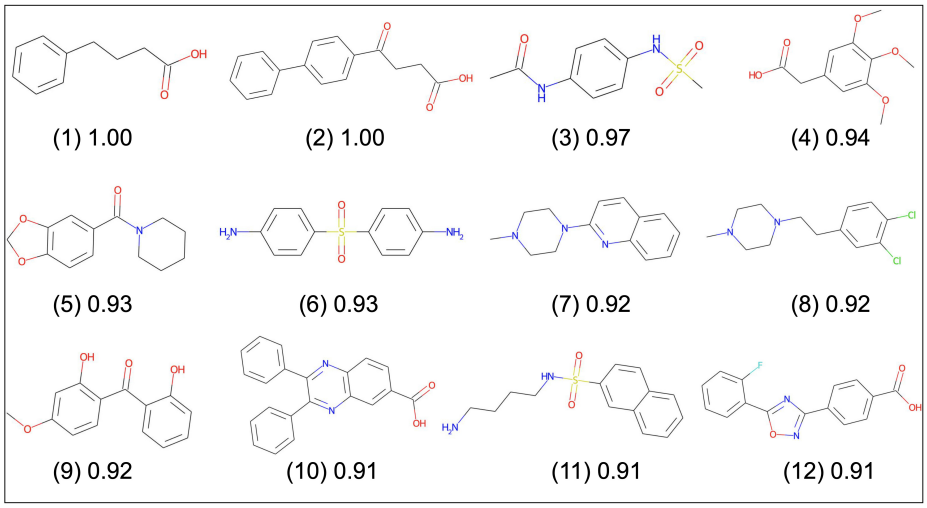

Fig. A.5. Top-12 molecular structures and their SA scores for molecules in the training dataset.

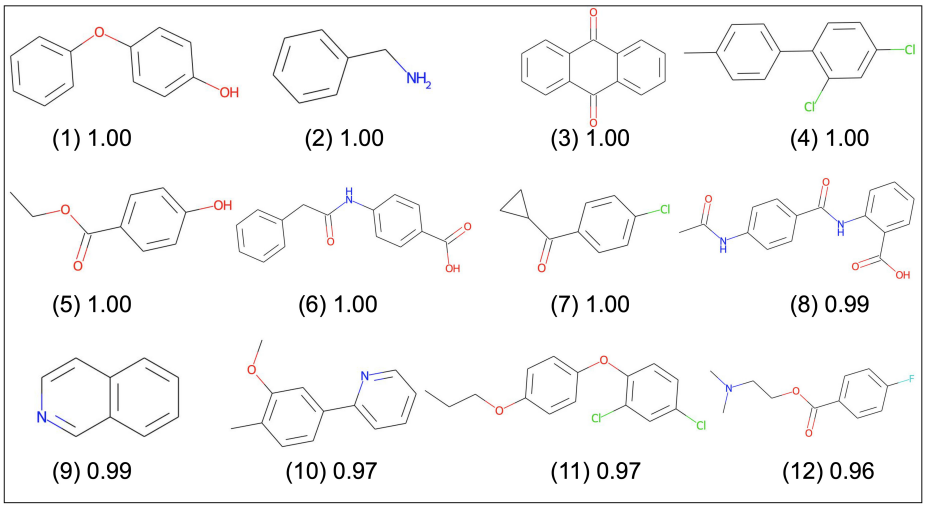

Fig. A.6. Top-12 molecular structures with their SA scores for molecules generated by Gx2Mol.

Table B.1. Statistical results of diversity metrics for newly generated molecules by Gx2Mol.

Therapeutic target protein	Max	Mean	Std
AKT1	1.00	0.87	0.09
AKT2	1.00	0.83	0.10
AURKB	1.00	0.87	0.09
CTSK	1.00	0.90	0.07
EGFR	1.00	0.86	0.10
HDAC1	1.00	0.86	0.09
MTOR	1.00	0.89	0.07
PIK3CA	1.00	0.82	0.11
SMAD3	1.00	0.87	0.11
TP53	1.00	0.85	0.09

⋆ Max, Mean, and Std denote the maximum, average, and standard deviation of diversity among the generated molecules.