APPENDIX A
VARIANT SMILES

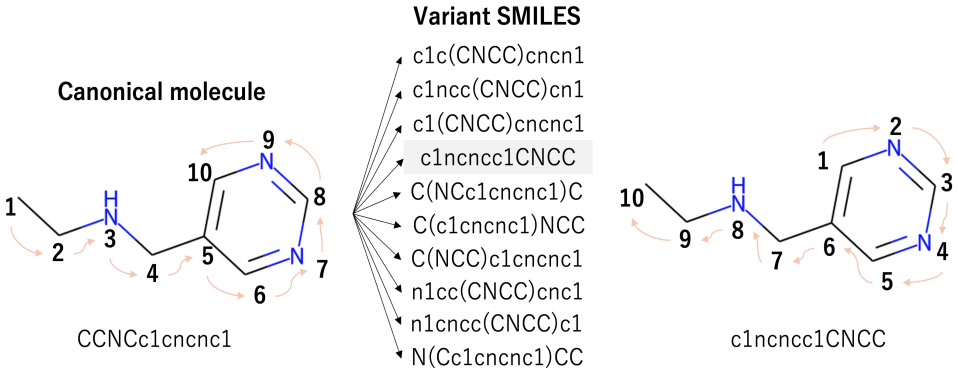


Fig. A.1. Example of producing variant SMILES strings.

Figure A.1 demonstrates an example of producing variant SMILES strings. Generally, a molecular graph has a canonical SMILES string (left figure). The numbers and arrows in the figure denote the traversal order of the atoms. Unlike NLP, the canonical molecule has various SMILES representations (middle figure) according to different traversal orders (right figure) called variant SMILES. However, the same molecular graph can be represented using these ten variant SMILES strings. Therefore, variant SMILES strings can be produced to improve the pretraining of the generator and prevent it from learning only a single semantic and syntactic feature. Several representations of the same molecule can be used to fully train the generator. This also alleviates the problem of mode collapse caused by a "perfect discriminator."

APPENDIX B
TRANSFORMER ENCODER

The input representations consist of three parts: the query matrix $\mathbf{Q}$, key matrix $\mathbf{K}$, and value matrix $\mathbf{V}$. The self-attention sublayer utilizing $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ is expressed as follows:

$$\text{self-attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_{model}}})\mathbf{V}, \tag{8}$$

where d_{model} denotes the dimension of the transformer encoder. Usually, the self-attention can be refined into multi-head self-attention so that the information of different representation subspaces from different positions can be jointly considered:

$$\text{multi-head}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{concat}(\text{head}_1,\cdots,\text{head}_H)\mathbf{W}, \tag{9}$$

and

$$\text{head}_i = \text{self-attention}(\mathbf{Q}\mathbf{W}^Q,\mathbf{K}\mathbf{W}^K,\mathbf{V}\mathbf{W}^V), \tag{10}$$

where $\mathbf{W} \in \mathbb{R}^{d_v \times d_{model}}$, $\mathbf{W}^Q \in \mathbb{R}^{d_{model} \times d_k}$, $\mathbf{W}^K \in \mathbb{R}^{d_{model} \times d_k}$, and $\mathbf{W}^V \in \mathbb{R}^{d_{model} \times d_v}$ are the parameter matrices, and

$$d_k = d_v = \frac{d_{model}}{h}, \tag{11}$$

where h is the number of heads. Then, the output of the multi-head self-attention layer is fed into a position-wise feedforward network to generate the final source representation $H_{1:T} = \{h_1,\cdots,h_T\}$:

$$H_{1:T} = \text{feed-forward}(\text{multi-head}(\mathbf{Q},\mathbf{K},\mathbf{V})) + \mathbf{V}. \tag{12}$$

Note that the multi-head attention is masked to avoid exposing the generator to future information during the sampling process. A padding mask is used to ensure that the transformer encoders do not pay any attention to the padding tokens in the discriminator.

APPENDIX C
PROPERTY CALCULATION

A. Diversity

The diversity was calculated from the Tanimoto similarity between the Morgan fingerprints of any two molecules in the generated set. Let V_i and V_j be the Morgan fingerprints of two arbitrary generated molecules. The Tanimoto similarity is then defined as

$$\text{Sim}(V_i, V_j) = \frac{|V_i \& V_j|}{|V_i| + |V_j| - |V_i \& V_j|},$$

where $|\cdot|$ represents the number of bits set in the fingerprints, and $\&$ is the common bits in the two fingerprints. The diversity is then calculated as

$$\text{Div}(\mathcal{D}_z) = 1 - \frac{1}{|\mathcal{D}_z|} \sum_{V_i, V_j \in \mathcal{D}_z} \text{Sim}(V_i, V_j).$$

B. Drug-likeness

Drug-likeness is evaluated by the QED scores. We generally assign different weights to eight molecular descriptors: the molecular weight (MW), octanol-water partition coefficient (ALOGP), number of hydrogen bond donors (HBDs), number of hydrogen bond acceptors (HBAs), molecular polar surface area (PSA), number of rotatable bonds (ROTBs), number of aromatic rings (AROMs), and number of structural alerts (ALERTS). The calculation is as follows:

$$\text{QED} = \exp(\frac{\sum_{i=1}^{8} W_i \ln d_i)}{\sum_{i=1}^{8} W_i},$$

where d_i and W_i represent the desirability function and weight of the i-th descriptor, respectively. Usually, the weights of the eight molecular descriptors are obtained through chemical experiments. In practice, the QED score is calculated by a function in the RDKit tool. The larger the QED score, the more drug-like the molecule.

C. Synthesizability

Synthesizability is evaluated by the SA score defined as

$$\text{SA} = r_s - \sum_{i=1}^{5} p_i,$$

where r_s indicates the "synthetic knowledge" gained by analyzing the features of synthetic molecules. r_s is the ratio of the summed contributions from all fragments to the number of fragments in the molecule. In this work, we calculated r_s from the experimental results [42]. p_i ($i \in \{1, \cdots, 5\}$) represents the ring complexity, stereo complexity, macrocycle penalty, size penalty, and bridge penalty, which were calculated using the RDKit tool. The larger the SA score, the easier the synthesis of the molecule.

D. Solubility

In the physical sciences, solubility is quantified by logP, where P is the partition coefficient (defined as the ratio of concentrations of a molecule in a mixture of two immiscible solvents at equilibrium). The logP is calculated as

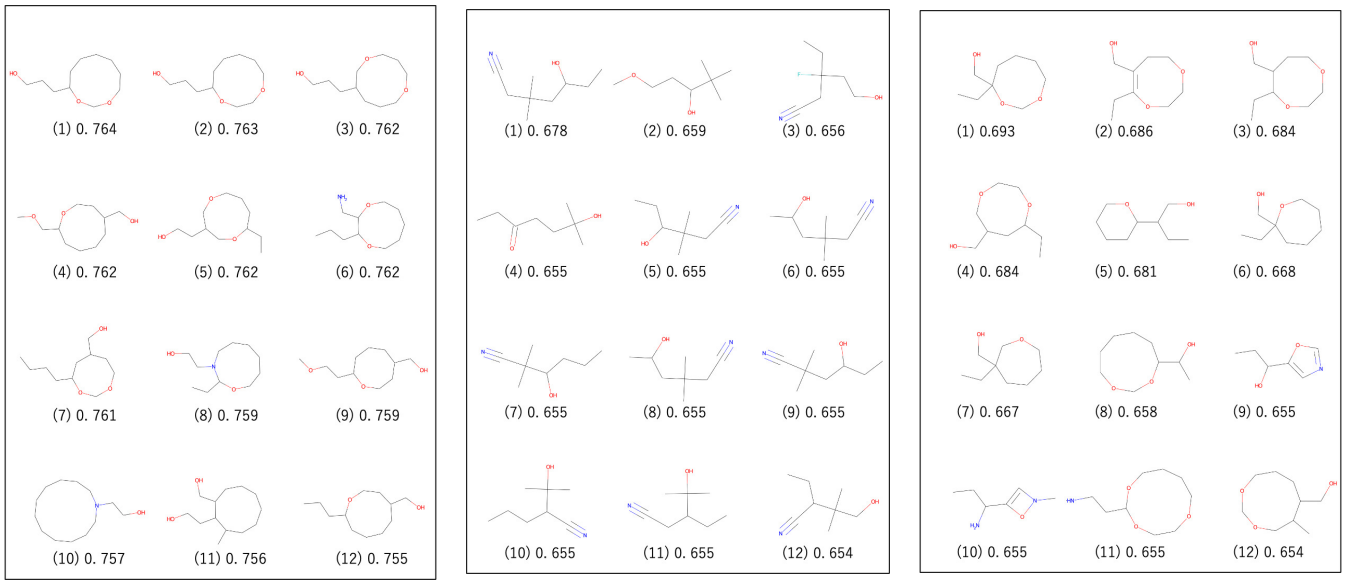
$$\text{logP} = \log \frac{c_o}{c_w},$$

where c_o and c_w indicate the substance activity in the organic and water phases, respectively. In practice, we calculate the logP of a molecule using the RDKit tool. The larger the logP value, the higher the lipophilicity of the molecule to the organic phase.

APPENDIX D
EXPERIMENTAL RESULTS ON QM9 DATASET

A. Molecule Structures

Figure C.1 shows the top-12 molecule structures by QED score for the following molecular generative models on the QM9 dataset: Naïve RL, ORGAN, and OR(W)GAN. In organic chemistry, molecular structures follow certain rules, in which existing cyclic organic compounds generally have five- or six-member rings because of their high stability. The molecules generated by Naïve RL [Fig. C.1 (a)] have ring structures consisting of more than seven atoms, which are not stable. Additionally, most of them violate Hückel's rule, which is a fundamental rule in organic chemistry. Meanwhile, all of the top-12 compounds generated by ORGAN [Fig. C.1 (b)] are acyclic and the QED scores are lower than those for the compounds generated by Ten(W)GAN. Similar to Naïve RL, most of the compounds generated by OR(W)GAN [Fig. C.1 (c)] have ring structures consisting of more than seven atoms. A four-member ring compound (10) that does not satisfy H$'$uckel's rule was also generated.



(a) Samples of the Naïve RL. (b) Samples of the ORGAN. (c) Samples of the OR(W)GAN.

Fig. C.1. Top-12 molecule structures with the QED scores (drug-likeness) generated by the baseline models.

B. Property Optimization of Ten(W)GAN

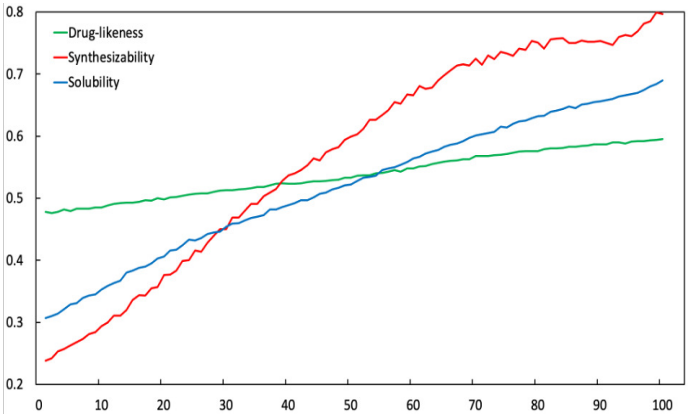
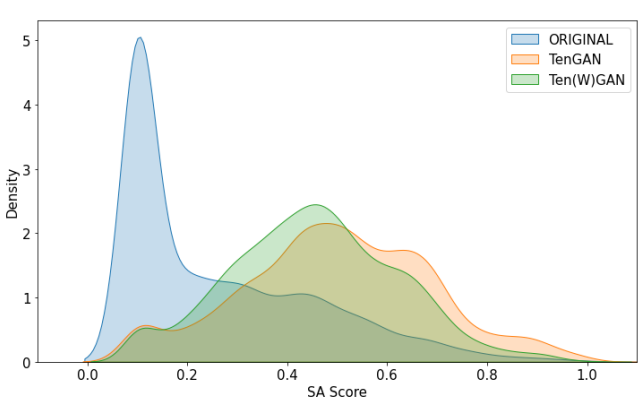


Fig. D.1. Changes in the property scores according to the training epoch for Ten(W)GAN on the QM9 dataset.

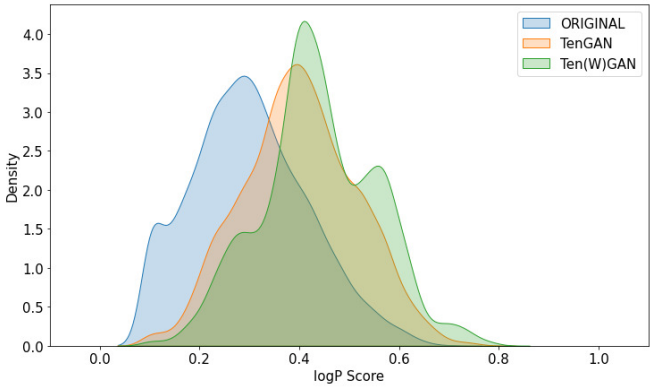
Figure D.1 shows the changes in the property scores according to the training epoch for Ten(W)GAN on the QM9 dataset. Similar to TenGAN, the optimized property scores increased with the training epoch for Ten(W)GAN. When the optimized property was drug-likeness, the curve of Ten(W)GAN increased smoothly and slowly. The curves for the synthesizability and solubility increased faster than the drug-likeness because more semantic and syntactic features were required to generate molecules with high drug-likeness. Fewer semantic and syntactic features were required to generate molecules with high synthesizability because simple molecules are easier to synthesize. For solubility, molecules with higher logP values tend to have longer carbon chains. Only a few semantic and syntactic features are required to generate molecules with a high logP. Therefore, the curve for the drug-likeness increased slowly, whereas the curves for the synthesizability and solubility increased rapidly. In summary, Ten(W)GAN can also generate molecules with desired chemical properties.

C. Property Distributions

Figure D.2 shows the SA and logP distributions of the generated molecules when the drug-likeness was the optimized property for TenGAN and Ten(W)GAN on the QM9 dataset. Figure D.2 (a) shows that not only the QED distribution but also the SA distribution was significantly improved with the proposed models. The SA distribution of the training set (ORIGINAL) had a peak at 0.1. TenGAN and Ten(W)GAN shifted the peaks to the right. Although the two distributions of the proposed



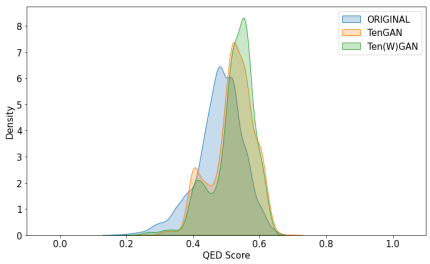
(a) Distribution of SA scores (synthesizability).



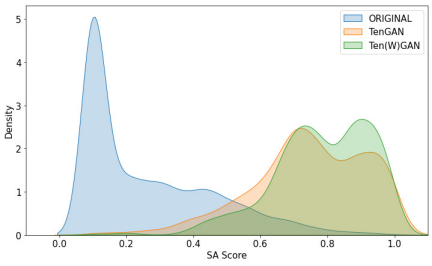
(b) Distribution of logP scores (solubility).

Fig. D.2. Distributions of SA and logP with drug-likeness as the optimized property on the QM9 dataset.

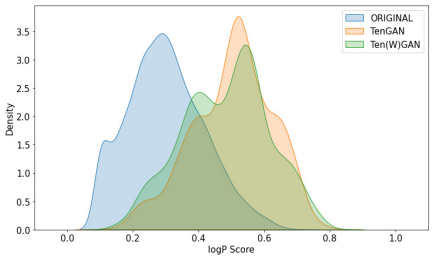
model became more scattered with the distribution of the original training set, the mean value of the SA increased overall. Similarly, Fig. D.2 (b) shows that the logP distributions improved with the proposed models. The difference between the SA and logP scores is that the QED score could not reach the maximum value of one.



(a) Distribution of QED (drug-likeness).

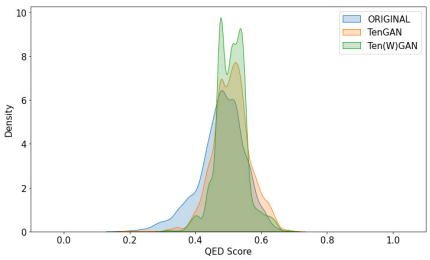


(b) Distribution of SA (synthesizability).

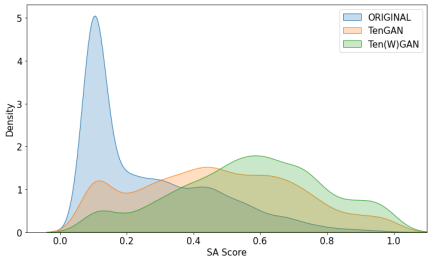


(c) Distribution of logP (solubility).

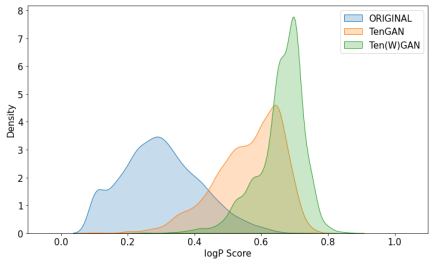
Fig. D.3. Distributions of QED, SA and logP with synthesizability as the optimized property on the QM9 dataset.



(a) Distribution of QED (drug-likeness).



(b) Distribution of SA (synthesizability).



(c) Distribution of logP (solubility).

Fig. D.4. Distributions of QED, SA and logP with solubility as the optimized property on the QM9 dataset.

When the property of synthesizability was optimized, TenGAN and Ten(W)GAN significantly improved the distributions of the generated molecules, and the SA score reached the maximum value of one. Figure D.3 (b) shows that the distribution peak of the training set (0.1) shifted to the right to around 0.75. Therefore, TenGAN and Ten(W)GAN generated more molecules that are easy to synthesize when the optimized property was synthesizability. Meanwhile, solubility is improved by generating carbon chain molecules, which are also easy to synthesize. Therefore, Fig. D.3 (c) shows that the proposed models improved the logP distribution. Unlike synthesizability and solubility, drug-likeness generated high-quality molecules. A molecule that is easy to synthesize and has a high logP value and generally has low drug-likeness. Figure D.3 (a) shows that the QED distributions of the proposed models did not improve much when synthesizability was the optimized property.

Similarly, the distributions of QED scores were not much improved when solubility was the optimized property. Figure D.4 (b) and Fig. D.4 (c) show that TenGAN and Ten(W)GAN effectively improved the SA and logP distributions of the generated

molecules, whereas the QED distribution remained the same [Fig. D.4] (a) because the solubility and drug-likeness properties are in conflict when the length of the SMILES string is small.

APPENDIX E
EXPERIMENTAL RESULTS ON ZINC DATASET

A. Comparison with RNN-based Methods

TABLE E.1
EVALUATION RESULTS WITH CHEMICAL PROPERTIES COMPARED WITH THE CLOSEST RELATED WORKS ON THE ZINC DATASET.

Property	Algorithm	Validity	Uniqueness	Novelty	QED	SA	logP	Diversity	Time (h)
	LSTM	51.9%	97.6%	98.6%	0.76	0.78	0.65	0.89	0.92
	TransEn	85.2%	97.8%	97.6%	0.79	0.78	0.63	0.89	0.30
Drug-likeness	Naïve RL	89.2%	65.5%	**98.5%**	0.81	0.53	0.64	0.86	12.79
	ORGAN	91.7%	54.9%	98.4%	0.80	0.48	**0.66**	0.85	12.40
	OR(W)GAN	91.5%	54.1%	98.2%	0.79	0.48	**0.66**	0.85	11.58
	TenGAN	**95.3%**	80.3%	96.2%	**0.84**	0.87	0.64	0.86	**5.21**
	Ten(W)GAN	**95.3%**	**81.2%**	96.5%	**0.84**	**0.88**	0.65	**0.87**	5.31
Synthesizability	Naïve RL	88.5%	39.3%	98.7%	0.79	0.91	0.66	0.86	13.22
	ORGAN	81.7%	44.0%	98.9%	0.77	0.87	0.61	0.86	13.29
	OR(W)GAN	76.1%	26.6%	**99.3%**	0.74	0.89	**0.75**	0.84	13.00
	TenGAN	**91.7%**	85.4%	96.5%	**0.81**	0.90	0.67	**0.86**	**3.25**
	Ten(W)GAN	90.5%	**86.5%**	96.3%	0.80	**0.91**	0.69	**0.86**	3.70
Solubility	Naïve RL	84.2%	76.4%	98.9%	0.74	**0.85**	0.74	0.86	13.58
	ORGAN	68.1%	44.2%	99.3%	0.78	0.82	0.67	0.84	13.55
	OR(W)GAN	83.8%	64.6%	98.3%	0.78	0.53	0.71	0.85	13.38
	TenGAN	**93.0%**	**80.7%**	**99.6%**	0.62	0.82	**0.92**	**0.86**	**4.96**
	Ten(W)GAN	91.7%	79.2%	**99.6%**	0.63	0.83	**0.92**	0.85	5.06

* Red boxes indicate the directly optimized properties. Bold values indicate the highest scores among different methods. The values in gray cells indicate that TenGAN / Ten(W)GAN outperformed the corresponding baseline.

Table E.1 compares the results of TenGAN and Ten(W)GAN with Naïve RL, ORGAN, and OR(W)GAN for the ZINC dataset. The results shows that the proposed models performed better than the baselines on most measures on the ZINC dataset.

For the drug-likeness, TenGAN improved the validity and uniqueness to 95.3% and 80.3%, respectively. Ten(W)GAN improved to 95.3% and 81.2%, respectively. Both TenGAN and Ten(W)GAN had the novelty scores larger than 96.0%. The QED scores were improved to 0.84. Additionally, TenGAN and Ten(W)GAN performed significantly better in terms of diversity and computational cost. When the optimized properties were synthesizability and solubility, TenGAN and Ten(W)GAN also worked better than the other models.

B. Top-12 Molecule Structures

Figures. E.1, E.2 and E.3 provide examles of the top-12 molecules of TenGAN and Ten(W)GAN for the ZINC dataset. For example, TenGAN and Ten(W)GAN in Fig. E.1 generated drug-like chemical structures.

C. Property Optimization of Ten(W)GAN

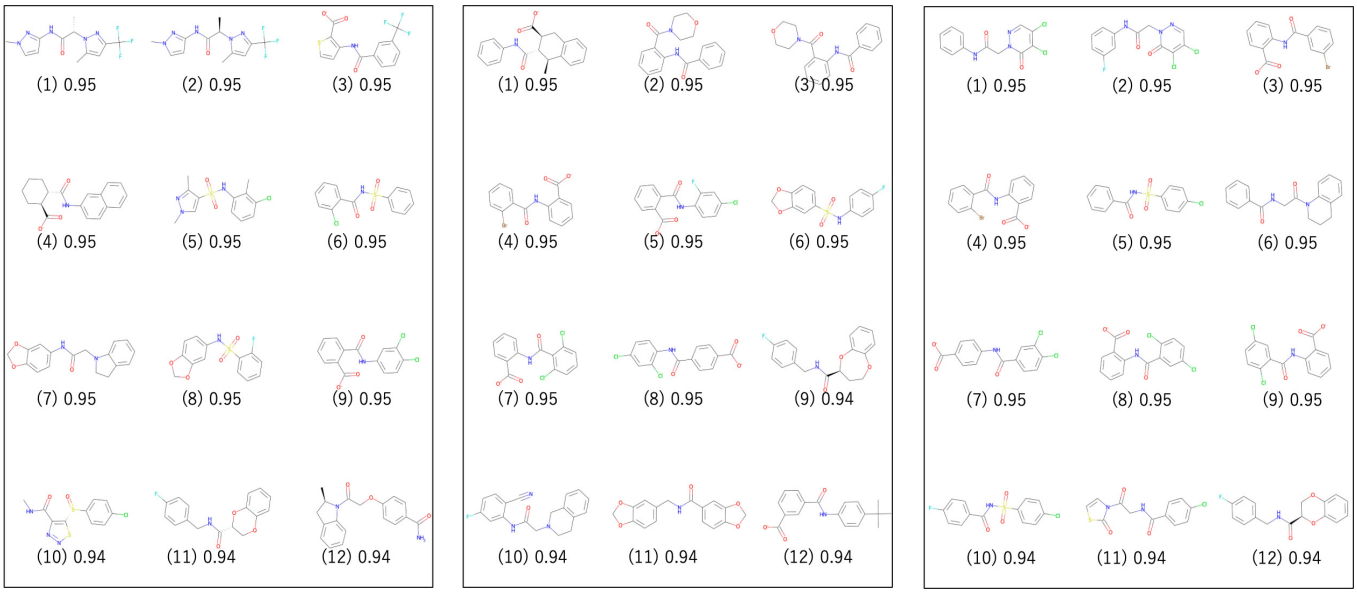
Figures E.4 shows the two curves of the property scores according to the training epochs for TenGAN and Ten(W)GAN on the ZINC dataset. The increase in the property scores with the training epochs indicated that TenGAN and Ten(W)GAN generated molecules with the desired chemical properties.

D. Property Distributions

Figure E.5 shows the distributions with drug-likeness, synthesizability, and solubility as the optimized properties on the ZINC dataset. From the figures, it was observed that TenGAN and Ten(W)GAN effectively improved the QED, SA, and logP distributions of the generated molecules.

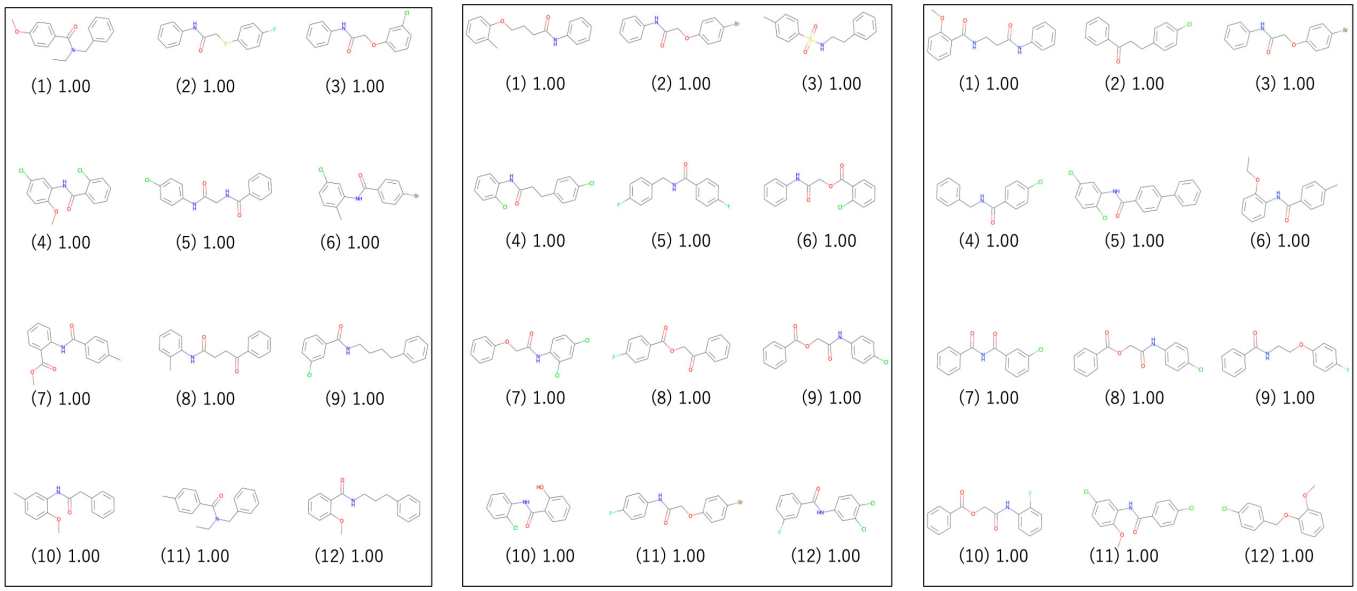
E. Comparison with Graph and VAE Methods

Table E.2 shows that although JTVAE and GraphAF realized the highest validity (i.e., 100%), the uniqueness and logP score were much lower than TenGAN. Similarly, CharVAE and GramVAE had much lower novelty and logP scores than TenGAN. Although TransVAE combined a VAE with a transformer, it had the smallest validity among these baseline models (i.e., 25.4%). In summary, TenGAN performed best in terms of Total and logP value while maintaining high validity, uniqueness, and novelty.



(a) Samples of the ZINC dataset. (b) Samples of the TenGAN. (c) Samples of the Ten(W)GAN.

Fig. E.1. Top-12 molecule structures with the QED scores (drug-likeness) on the ZINC dataset.

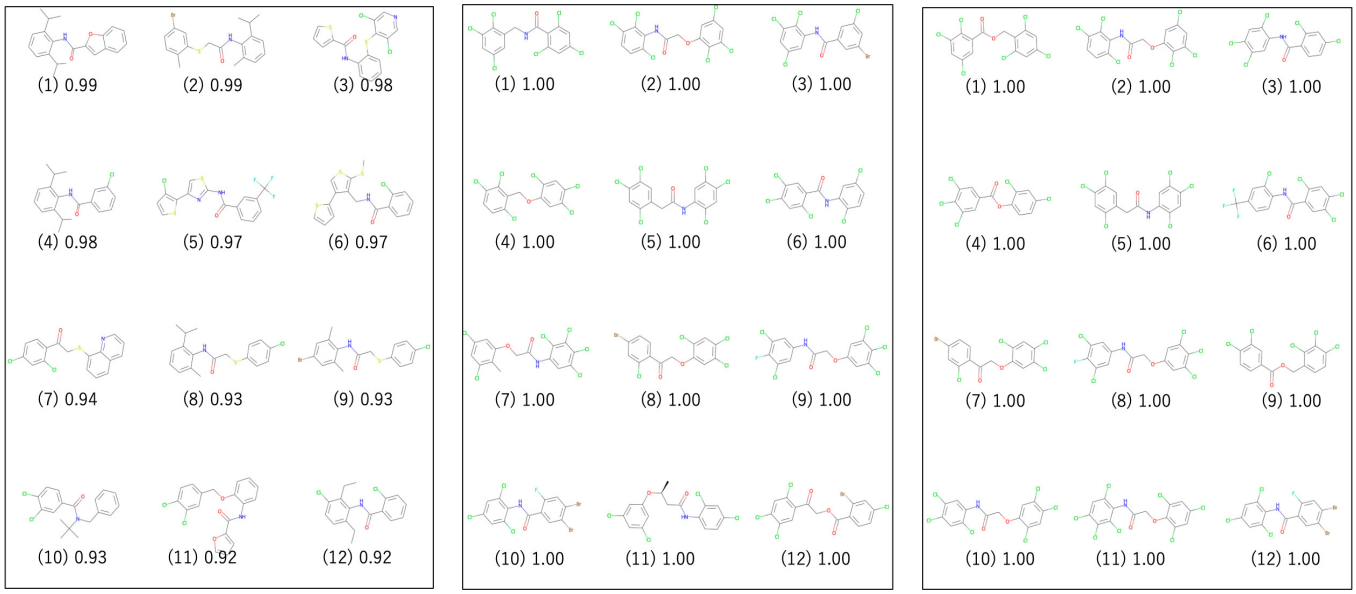


(a) Samples of the ZINC dataset. (b) Samples of the TenGAN. (c) Samples of the Ten(W)GAN.

Fig. E.2. Top-12 molecule structures with the SA scores (synthesizability) on the ZINC dataset.

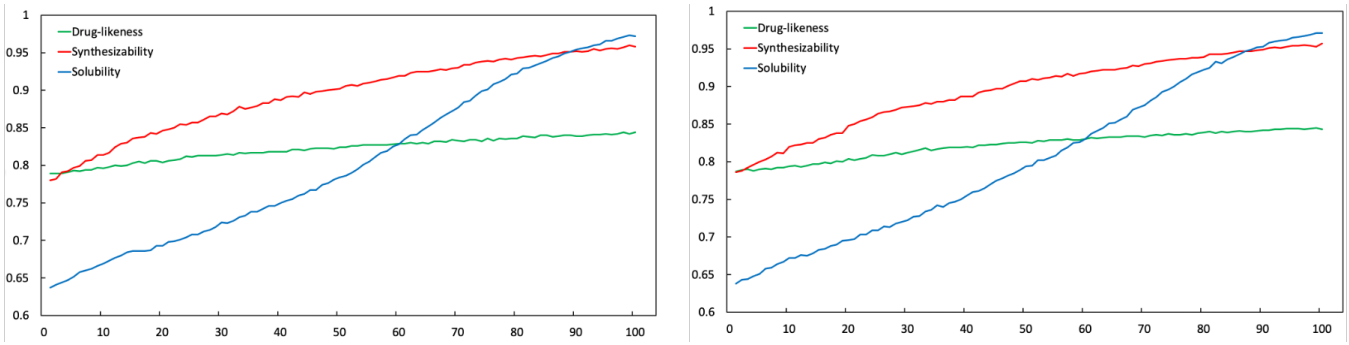
APPENDIX F
ABLATION STUDIES

Effect of λ. λ represents the tradeoff between the GAN and RL. Different $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ were used on the drug-likeness to explore the effect of λ on the performance of TenGAN. The results are presented in Table F.1. Increasing λ decreased the QED score, validity, and novelty while increasing the uniqueness and diversity. These results are reasonable because RL mainly controls the training process when λ is small. Particularly, the training is fully executed by RL when $\lambda = 0$. Because RL updates the generator by improving the average reward of generated molecules and invalid molecules receive zero rewards, the generator tends to generate mostly valid molecules in this case. However, a higher validity and QED score mean that the model must use less data to learn strategies (i.e., the overlap between the training dataset distribution and the generated data distribution shown in Fig. 4). Therefore, when RL dominates the training process, the uniqueness and diversity decrease.



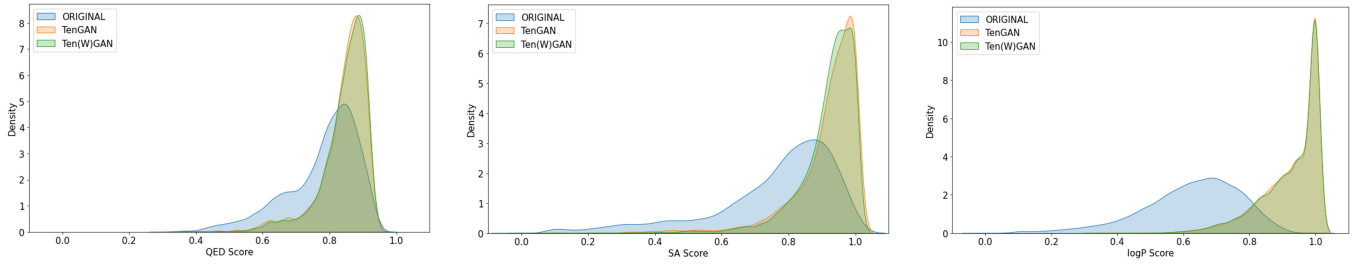
(a) Samples of the ZINC dataset. (b) Samples of the TenGAN. (c) Samples of the Ten(W)GAN.

Fig. E.3. Top-12 molecule structures with the logP scores (solubility) on the ZINC dataset.



(a) Property change curve for TenGAN. (b) Property change curve for Ten(W)GAN.

Fig. E.4. Changes in the property scores according to the training epoch on the ZINC dataset.



(a) Distribution of QED (drug-likeness). (b) Distribution of SA (synthesizability). (c) Distribution of logP (solubility).

Fig. E.5. Distributions with drug-likeness, synthesizability, and solubility as the optimized properties on the ZINC dataset.

Effect of N. The rollout times N controls the number of MC searches, and it also affects the performance of TenGAN. Intuitively, when N is small, the randomness of the strings generated based on substrings and the variance of rewards are relatively large, which may result in model instability. Table F.2 shows the effect of N on the performance of TenGAN with regard to drug-likeness (QED score). Increasing N stabilizes the validity, uniqueness, and novelty, but the computational time increased as N increases.

TABLE E.2
COMPARISON OF TENGAN WITH GRAPH-BASED AND VAE-BASED ALGORITHMS ON THE ZINC DATASET.

Alogrithm	Validity	Uniqueness	Novelty	Total	logP
JTVAE	**100.0%**	19.8%	99.8%	19.8%	0.51
GraphAF	**100.0%**	83.2%	100.0%	83.2%	0.61
CharVAE	86.7%	81.2%	26.4%	18.6%	0.55
GramVAE	91.9%	77.2%	11.9%	8.5%	0.58
TransVAE	25.4%	**100.0%**	**100.0%**	25.4%	0.14
MoFlow	26.0%	**100.0%**	**100.0%**	26.0%	0.71
MolGAN	95.3%	4.3%	**100.0%**	4.1%	0.62
TenGAN	91.6%	95.4%	97.5%	**85.2%**	**0.73**
Ten(W)GAN	91.2%	95.7%	97.2%	84.8%	0.71

* Total is the product of validity, uniqueness, and novelty.

TABLE F.1
EFFECT OF DIFFERENT λ ON THE QM9 DATASET.

λ	QED	Validity	Uniqueness	Novelty	Diversity
0	0.59	98.1%	82.6%	**99.8%**	0.90
0.1	**0.60**	97.8%	82.6%	**99.8%**	0.89
0.3	0.57	**98.7%**	60.2%	96.7%	0.89
0.5	0.57	97.8%	70.7%	98.0%	0.90
0.7	0.52	96.8%	71.5%	92.9%	0.92
0.9	0.47	91.3%	**91.6%**	86.6%	**0.93**
1.0	0.46	82.0%	90.2%	85.8%	**0.93**

TABLE F.2
EFFECT OF DIFFERENT N ON THE QM9 DATASET.

N	QED	Validity	Uniqueness	Novelty	Time (h)
0	0.48	89.5%	**93.9%**	82.7%	0.33
2	**0.57**	98.1%	77.9%	**99.4%**	2.16
4	0.55	98.2%	60.2%	94.5%	2.99
8	**0.57**	97.8%	62.0%	98.7%	4.73
16	**0.57**	97.8%	70.7%	98.0%	5.06
32	0.56	**98.6%**	71.9%	97.4%	11.62

TABLE F.3
EFFECT OF THE VARIANT SMILES ON THE ZINC DATASET.

Algorithm	QED	Validity	Uniqueness	Novelty
TenGAN w/o	0.83	93.6%	69.2%	96.1%
TenGAN	**0.84**	**95.3%**	**80.3%**	**96.2%**
Ten(W)GAN w/o	0.84	93.7%	74.2%	95.9%
Ten(W)GAN	**0.84**	**95.3%**	**81.2%**	**96.5%**

* "w/o" indicates model training without the variant SMILES.

Effect of Variant SMILES. We also evaluated the effect of the variant SMILES on the performance of TenGAN and Ten(W)GAN, respectively. The evaluation results are shown in Table F.3. All metrics (i.e., validity, uniqueness, novelty, and QED scores) of TenGAN and Ten(W)GAN improved after using variant SMILES. Therefore, variant SMILES can help TenGAN and Ten(W)GAN to fully train and learn more syntactic and semantic rules from SMILES.