Inhibitor_Mol_VAE: A Variational Autoencoder Approach for Generating Corrosion Inhibitor Molecules

Haiyan Gong[1,3,4], Zhongheng Fu[1,2], Lingwei Ma[1,2,3,4*], Dawei Zhang[1,2,3,4*]

1. National Materials Corrosion and Protection Data Center, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China
2. Institute of Materials Intelligent Technology, Liaoning Academy of Materials, Shenyang 110004, China
3. Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing, 100083, China
4. Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China

Corresponding author:

Dawei Zhang, Email: dzhang@ustb.edu.cn; Lingwei Ma, Email: mlw1215@ustb.edu.cn

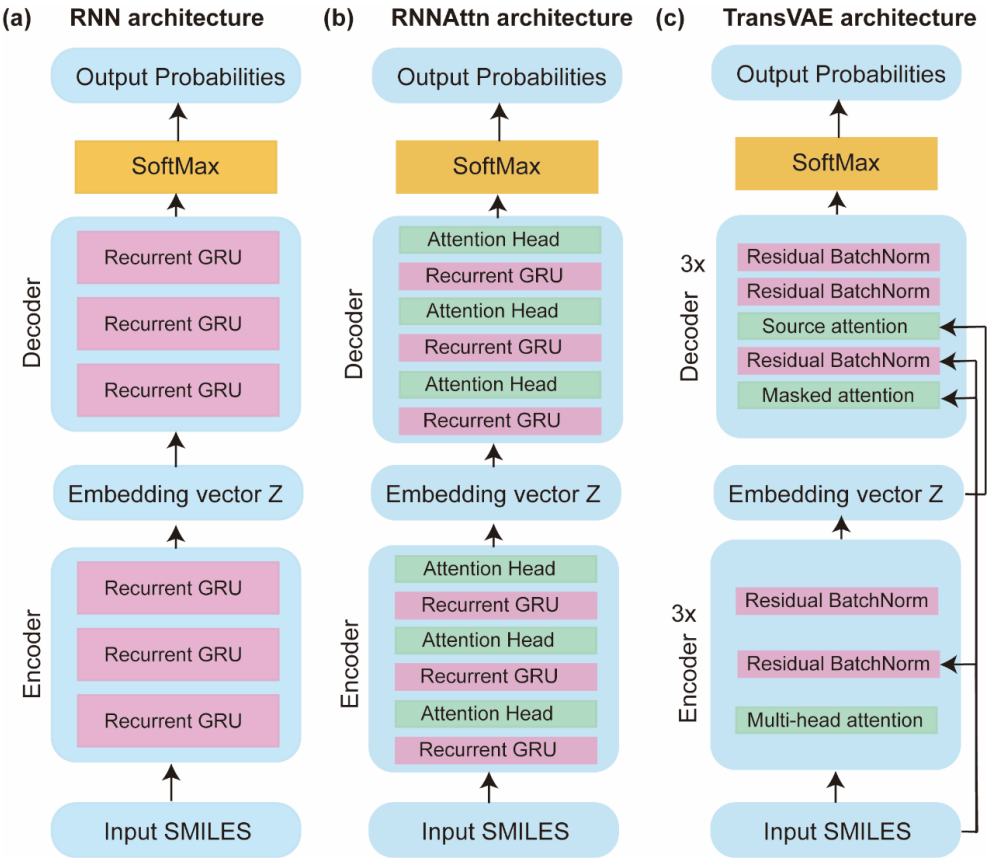

Fig.S1 Schematic diagram of models. (a) RNN, (b) RNNAttn, and (c) TransVAE.

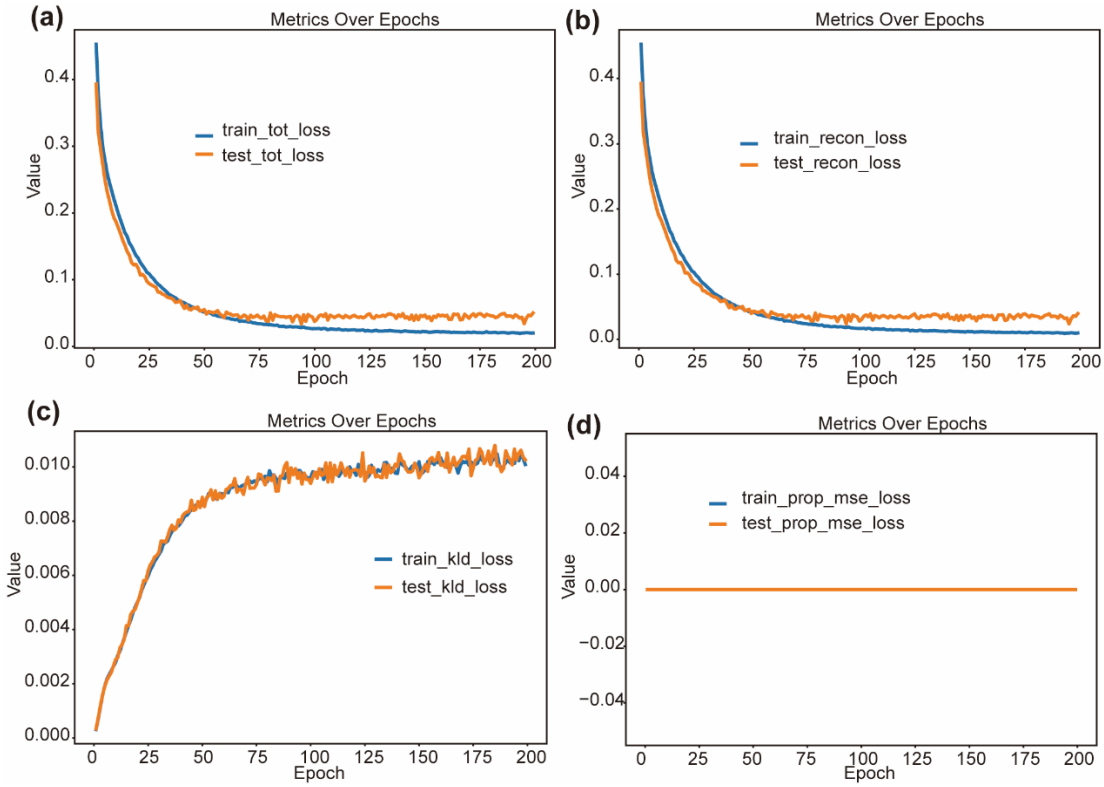

Fig.S2. Comparison of loss metrics over epochs when training the Inhibitor1368_data_0 dataset using RNN model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

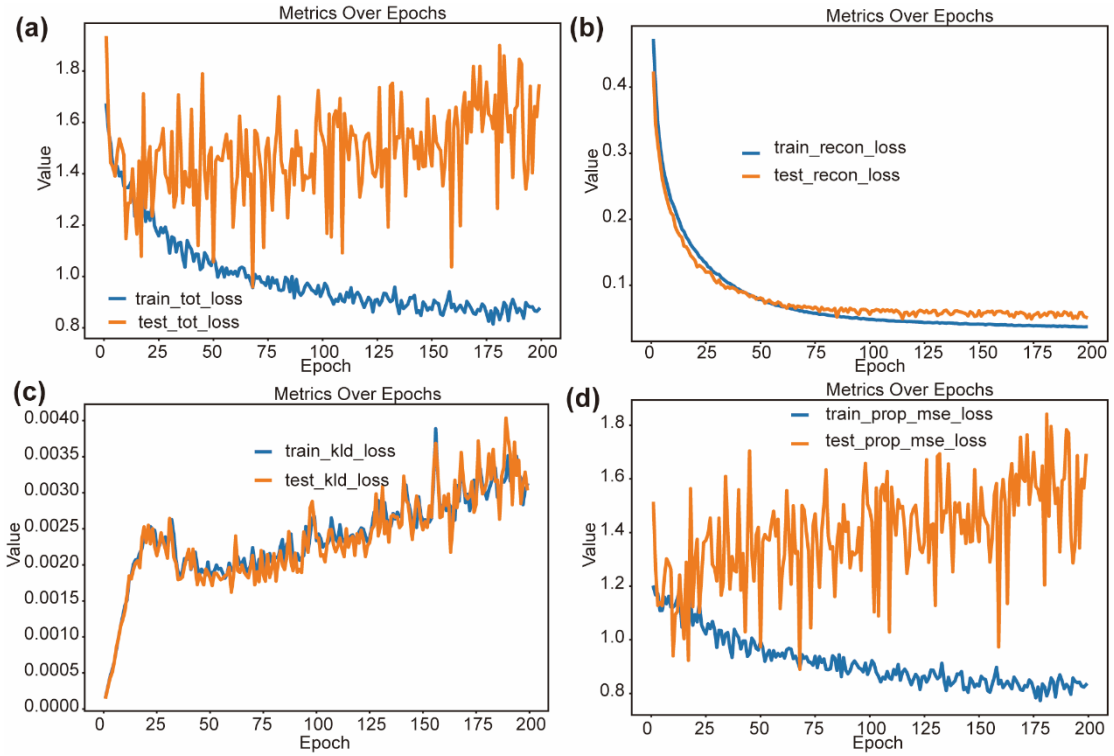

Fig.S3. Comparison of loss metrics over epochs when training the Inhibitor1368_data_2 dataset using RNN model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of

properties.

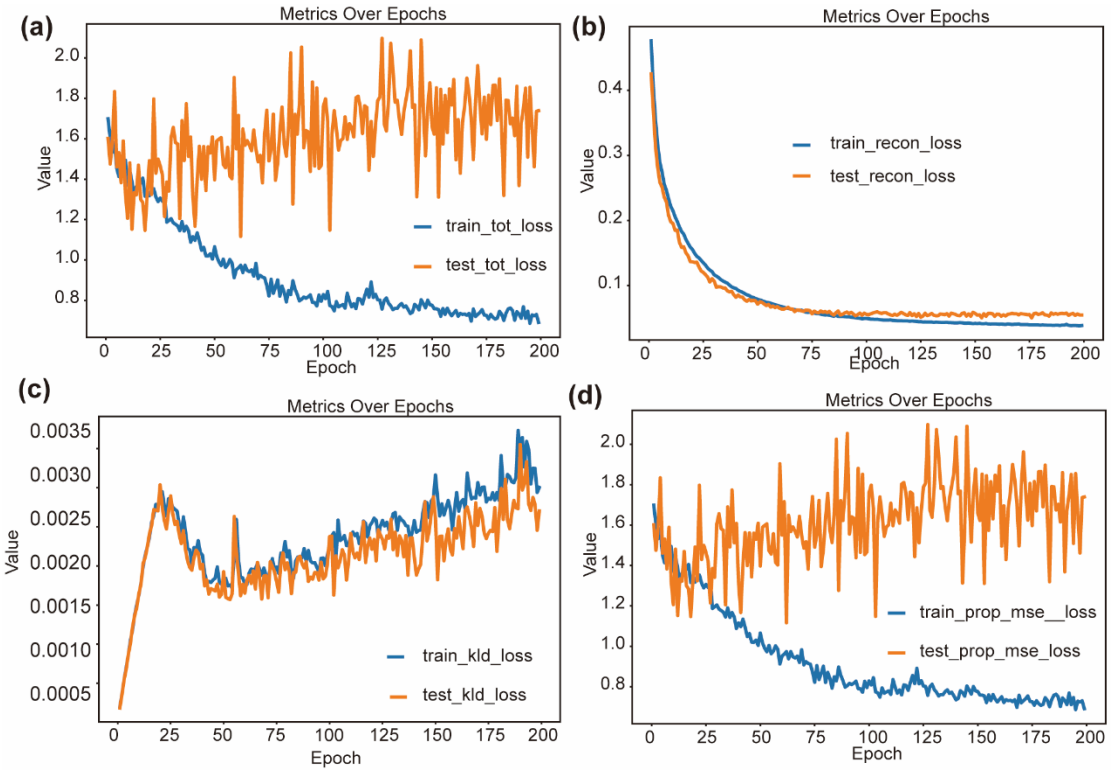

Fig.S4. Comparison of loss metrics over epochs when training the Inhibitor1368_data_9 dataset using RNN model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

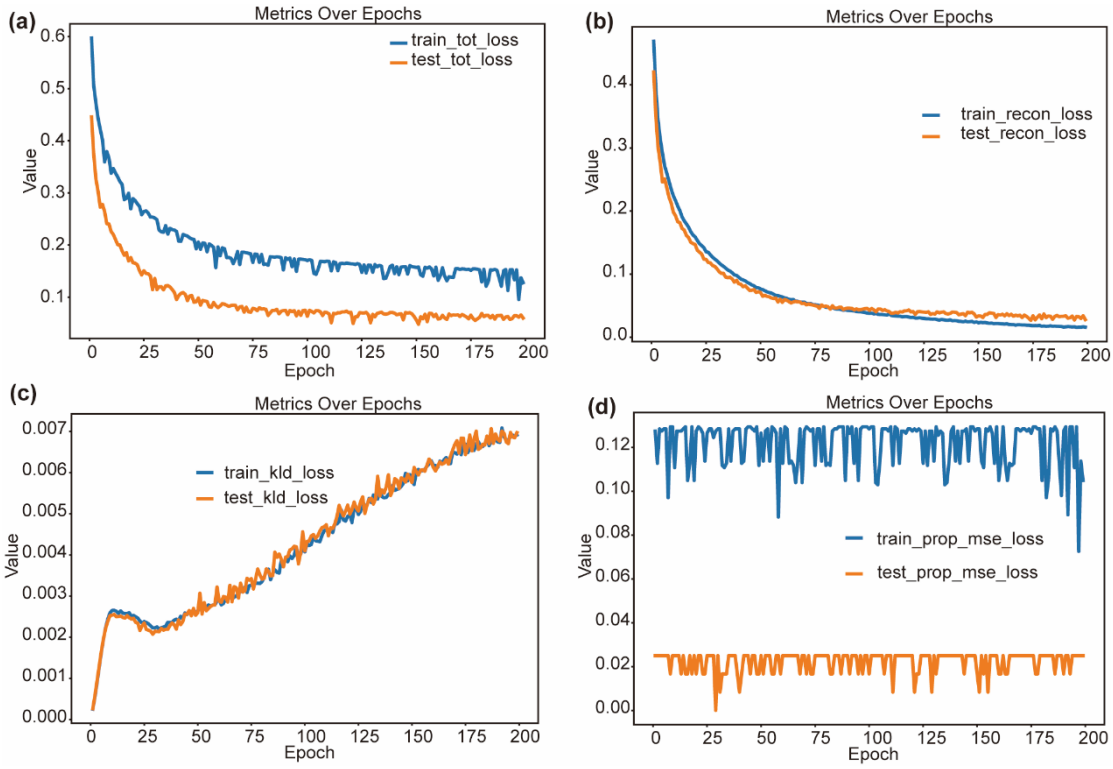

Fig.S5. Comparison of loss metrics over epochs when training the Inhibitor1368-13 dataset using

RNN model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

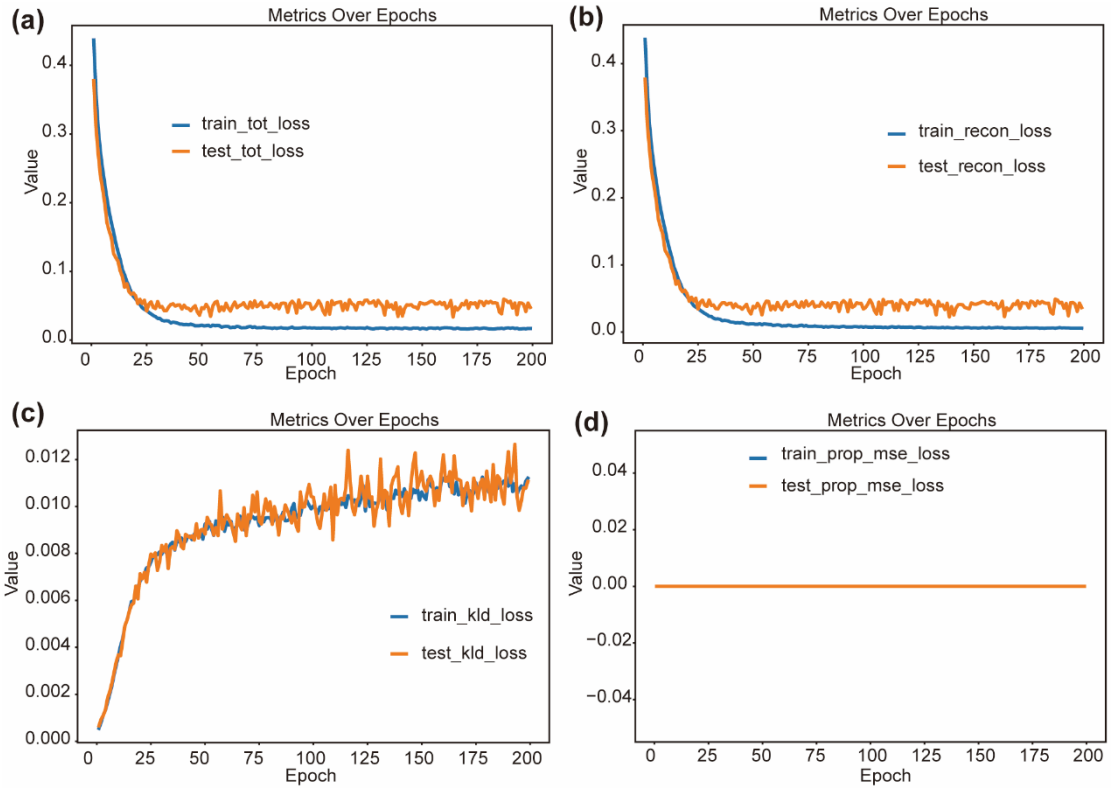

Fig.S6. Comparison of loss metrics over epochs when training the Inhibitor1368_data_0 dataset using RNNAttn model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

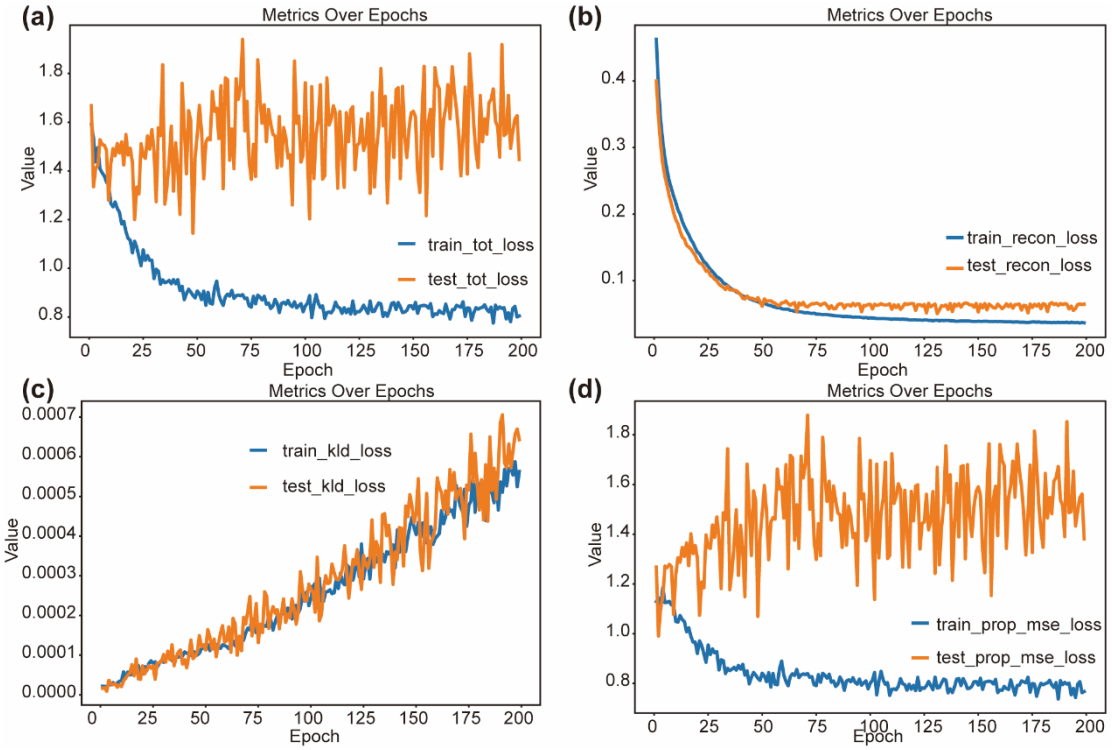

Fig.S7. Comparison of loss metrics over epochs when training the Inhibitor1368_data_2 dataset using RNNAttn model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of

properties.

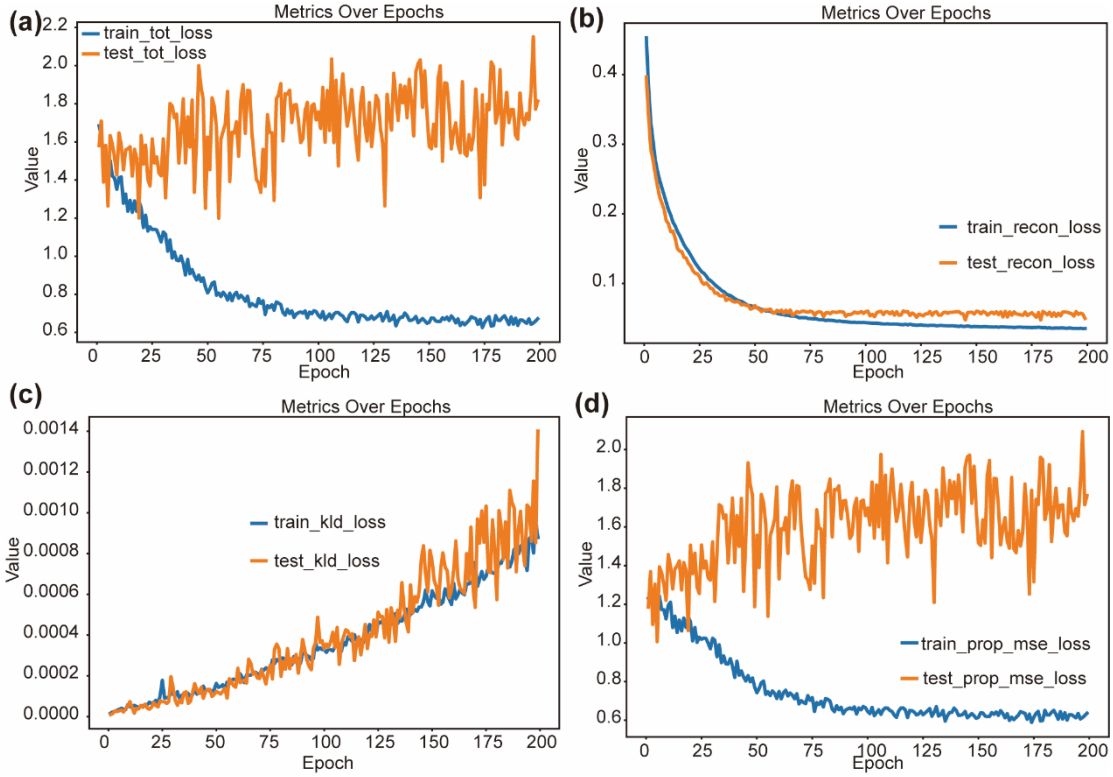

Fig.S8. Comparison of loss metrics over epochs when training the Inhibitor1368_data_9 dataset using RNNAttn model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

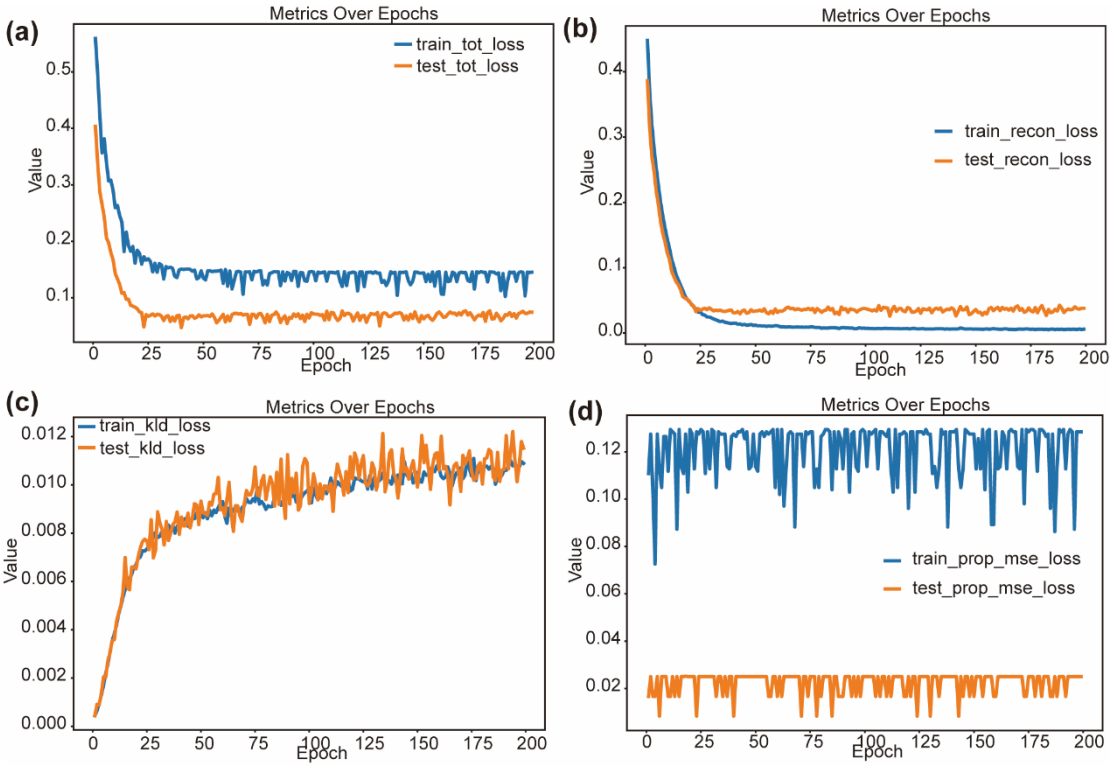

Fig.S9. Comparison of loss metrics over epochs when training the Inhibitor1368-13 dataset using

RNNAttn model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

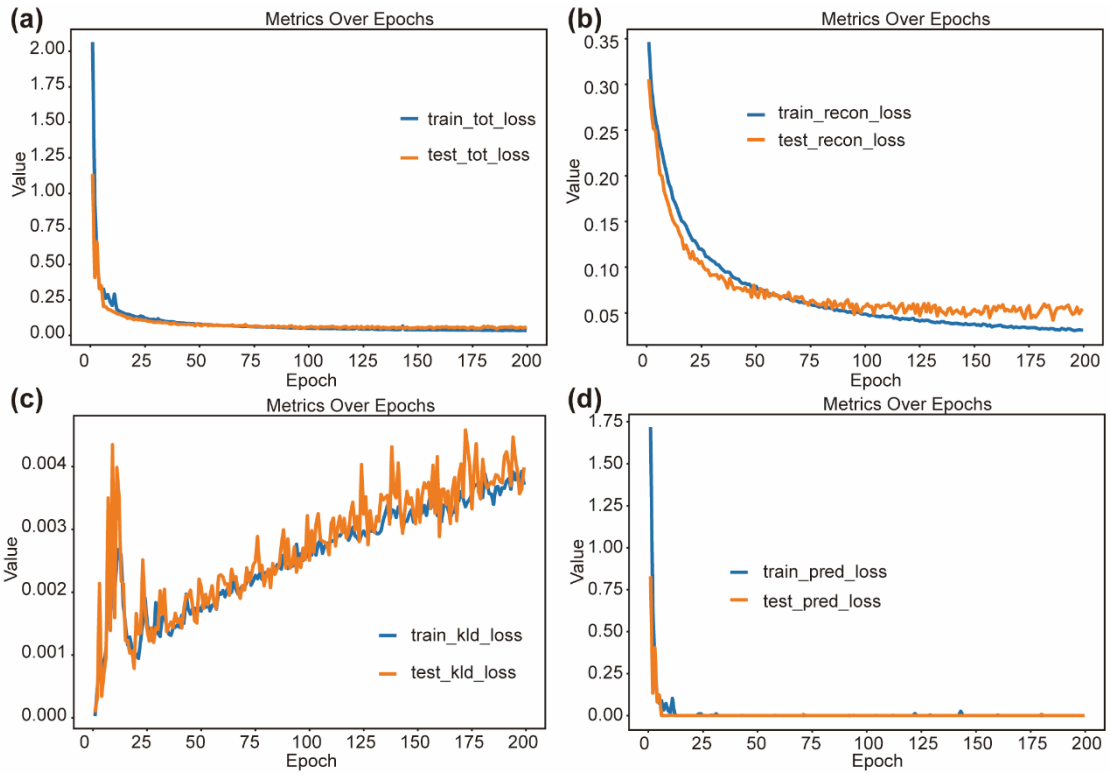

Fig.S10. Comparison of loss metrics over epochs when training the Inhibitor1368_data_0 dataset using TransVAE model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of predicting properties.

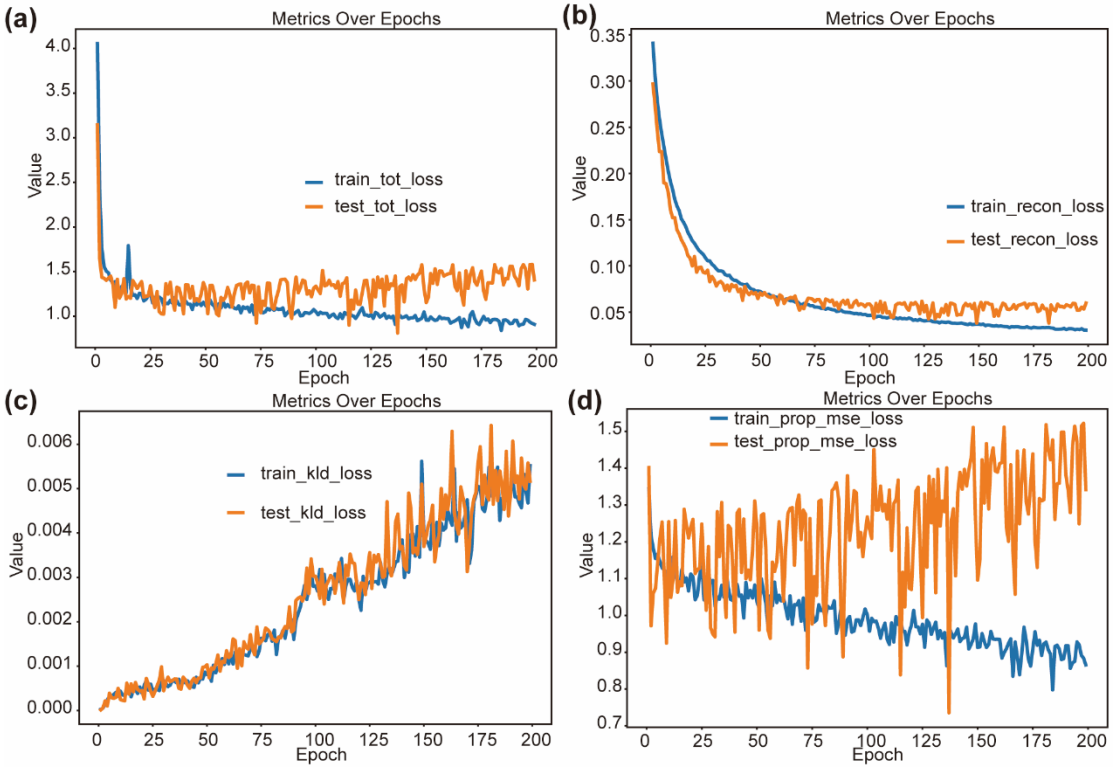

Fig.S11. Comparison of loss metrics over epochs when training the Inhibitor1368_data_2 dataset

using TransVAE model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

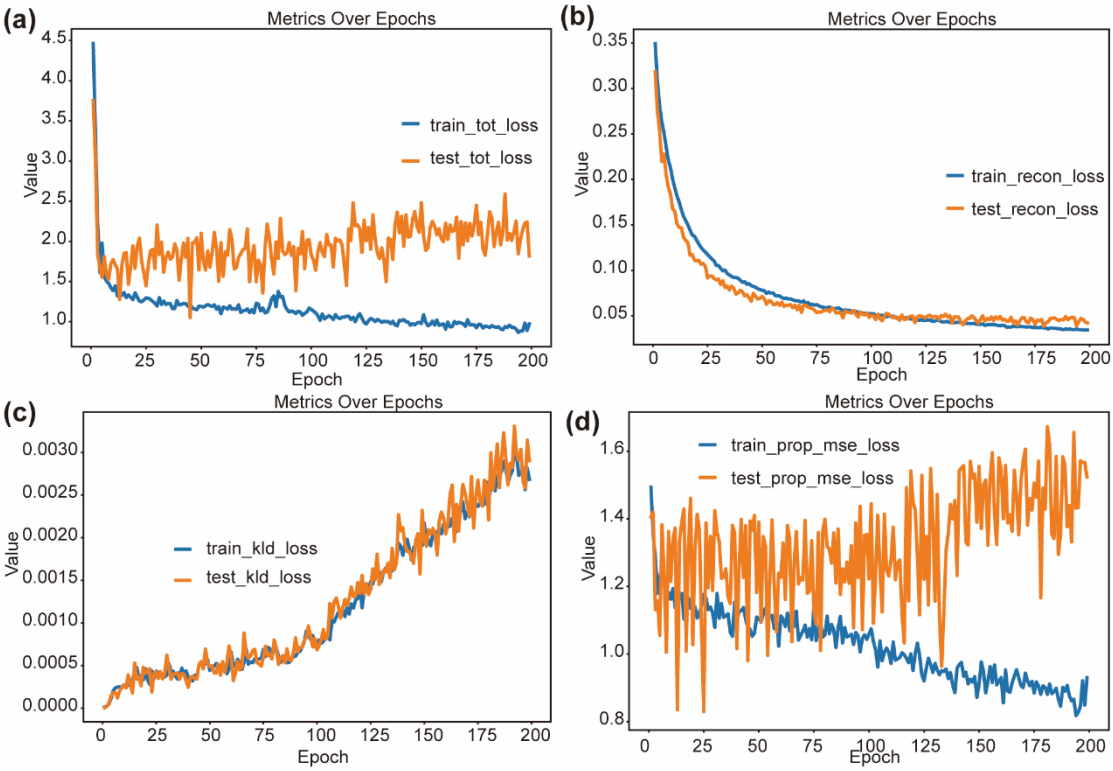

Fig.S12. Comparison of loss metrics over epochs when training the Inhibitor1368_data_9 dataset using TransVAE model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

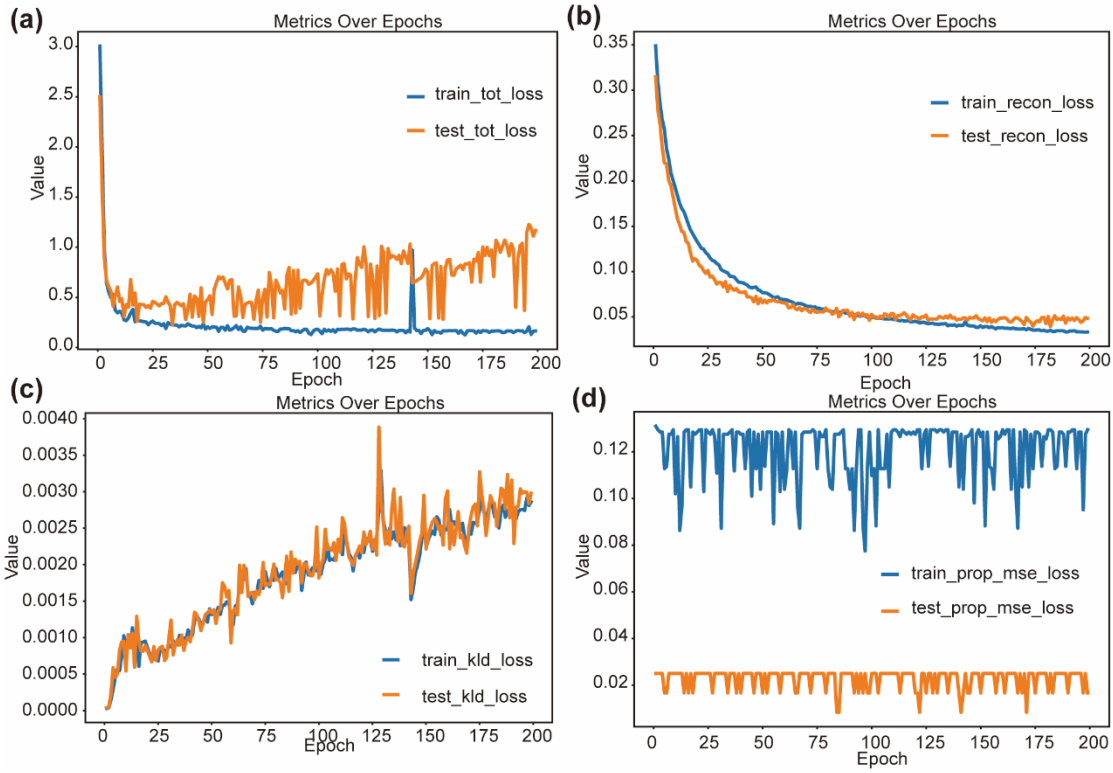

Fig.S13. Comparison of loss metrics over epochs when training the Inhibitor1368-13 dataset using TransVAE model. (a) Total loss, (b) reconstruction loss, (c) KLD loss, and (d) MSE loss of properties.

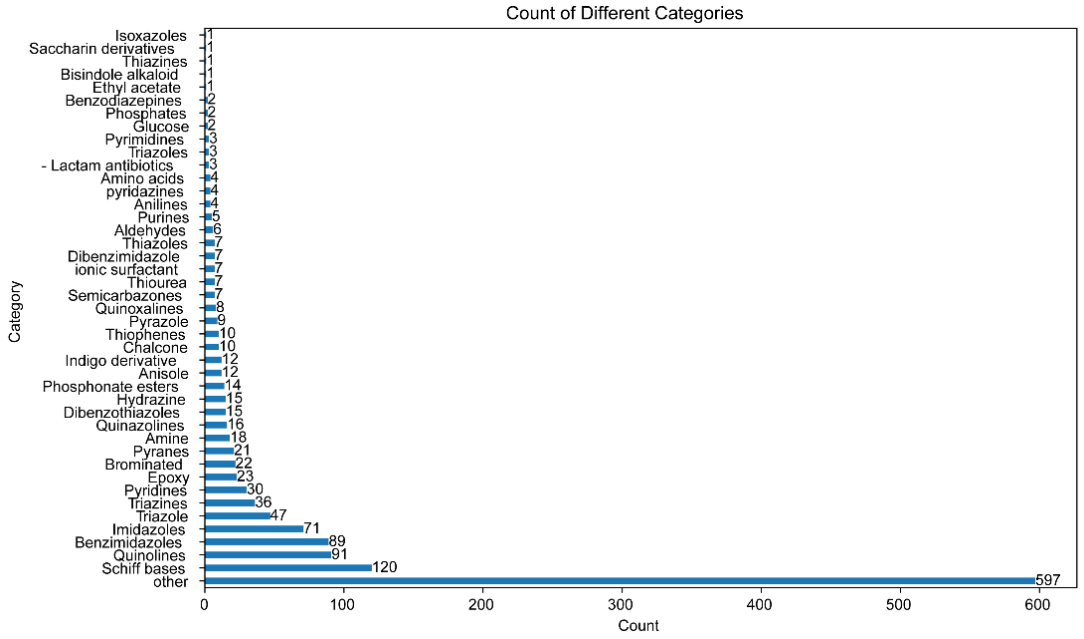

Fig.S14. Count of corrosion inhibitor molecules categories in Inhibitor1368 dataset.

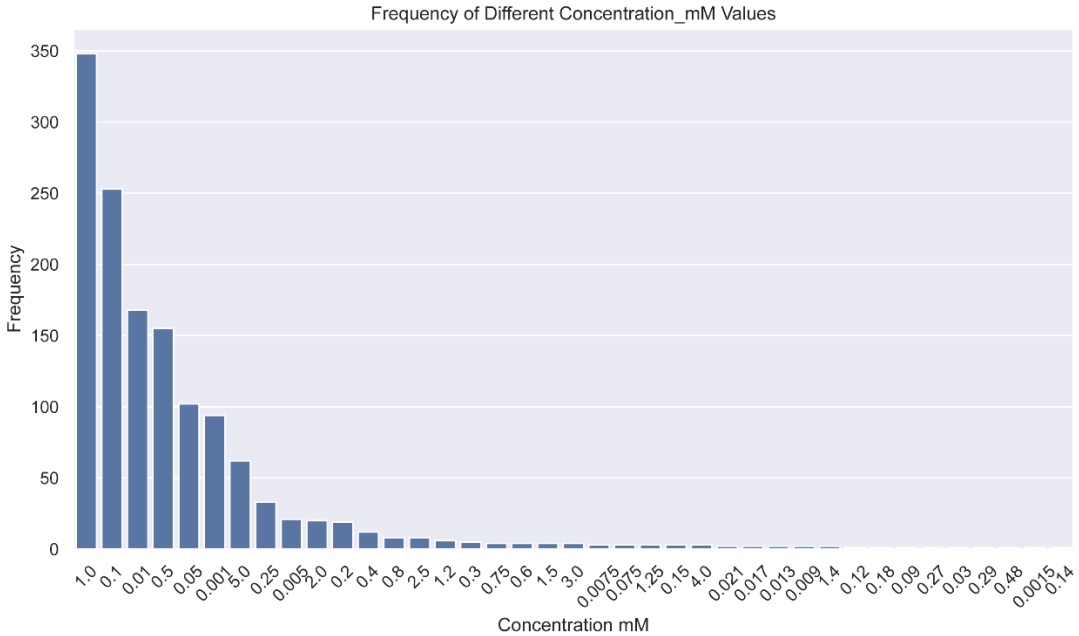

Fig.S15. Distribution of the Concentration_mM values of corrosion inhibitor.