Supplementary Section 1: The meanings of the molecular physicochemical features

ALERTS: ALERTS (ALERTS for structure) is a set of rules used to identify structural patterns in molecules that may indicate potential issues. These rules are commonly employed in drug screening and chemoinformatics applications to identify structures that might exhibit chemical instability, toxicity, or other adverse properties. By recognizing and excluding these problematic structures, researchers can focus on compounds with better drug potential.

FractionCSP3: FractionCSP3 (Fraction of sp3 Carbon-Sp3 Carbon bonds) is a molecular descriptor that describes the saturation level of the molecular skeleton. It represents the ratio of sp3 hybridized carbon-carbon bonds (single bonds) to all carbon-carbon bonds (including single, double, and triple bonds) in the molecule. A higher FractionCSP3 value indicates a higher proportion of sp3 carbon-carbon bonds, while a lower value indicates a higher proportion of unsaturated bonds (such as double and triple bonds). The FractionCSP3 value is often associated with the stereochemical complexity and drug-like properties of molecules.

MW: MW (Molecular Weight) is the sum of the masses of all atoms in a molecule. Molecular weight is closely related to physical properties such as size, sedimentation coefficient, diffusion coefficient, and so on. In the process of drug screening, molecular weight is often used as a filtering criterion because larger molecules may have more difficulty in penetrating cell membranes or exhibit unfavorable pharmacokinetic properties.

ALOGP: ALOGP (ALOGP value of a molecule) is the logarithm of the partition coefficient between water and an organic phase (typically n-octanol). It is a descriptor that characterizes the lipophilicity of a molecule and is often associated with the absorption, distribution, metabolism, and excretion (ADME) properties of a molecule in the body. A higher ALOGP value indicates a molecule's higher stability in the organic phase, while a lower value indicates a more hydrophilic molecule. In the process of drug screening and optimization, ALOGP values are frequently used to assess the drug-like properties and bioavailability of molecules.

Supplementary Section 2: MoleculeNet datasets

SIDER: The Side Effect Resource (SIDER) is a database that contains information on marketed drugs and their adverse drug reactions (ADRs).

BBBP: The Blood-Brain Barrier Penetration (BBBP) dataset is based on recent research on modeling and predicting barrier permeability. The blood-brain barrier is a membrane that separates the circulating blood from the brain extracellular fluid, preventing most drugs, hormones, and neurotransmitters from passing through. This dataset includes binary labels indicating the permeability characteristics of over 2000 compounds.

ToxCast: ToxCast provides toxicology data from a large compound library screened using high-throughput in vitro methods, similar to the Tox21 project. The subset processed in MoleculeNet includes qualitative results from over 600 experiments for 8615 compounds.

Tox21: The Toxicology in the 21st Century (Tox21) project created a public database for measuring the toxicity of compounds, which was used in the 2014 Tox21 Data Challenge [42].

ClinTox: The ClinTox dataset compares FDA-approved drugs with drugs that failed clinical trials due to toxicity reasons. It includes two classification tasks for 1491 known drug compounds: (1) clinical trial toxicity (or non-toxicity) and (2) FDA approval status [42].

FreeSolv: The Free Solvation database provides experimental and calculated hydration free energies of small molecules in water.

ESOL: The ESOL dataset contains water solubility data for 1128 compounds. It has been used to train models for predicting solubility based on the chemical structures encoded in SMILES (Simplified Molecular Input Line Entry System) strings. These structures do not include 3D coordinates since the solubility is a molecular property, not a specific conformational property.

qm7: The dataset is a subset of the GDB-13 database, where the 3D Cartesian coordinates of each molecule are determined using binary density functional theory (PBE0/tier2 basis set) to obtain the most stable conformations and electronic properties (atomic energies, HOMO/LUMO eigenvalues, *etc.*). Learning methods based on the qm7 benchmark are responsible for predicting these electronic properties using the most stable conformation coordinates.

qm8: The dataset originates from recent research on quantum mechanical calculations of electronic spectra and modeling of small molecule excited-state energies. It is a subset of GDB-17 and contains four excited-state properties computed using three different methods on 22 thousand samples.

Supplementary Section 3: Implementation Details

Encoder

Our VAE architecture, inspired by the JTVAE, involves a two-part encoder and decoder system designed specifically for molecular graphs. The encoder consists of two separate components:

1. **Graph Encoder:** This component encodes the original molecular graph into a latent representation. It uses a graph message passing network where each vertex and edge has feature vectors representing atom and bond types respectively. The final graph representation is aggregated from the latent vectors of all vertices. The dimensionality of the graph encoder output is designed to capture the fine-grained connectivity of the molecular graph.
2. **Tree Encoder:** This component encodes a junction tree representation of the molecule. The junction tree is formed by clusters that represent subgraphs such as rings or bonds. The tree encoder uses a message passing network specifically adapted for trees, where each node or cluster in the tree is encoded into a latent representation. This representation captures the higher-level structure of the molecule.

The latent embedding (z) from our VAE consists of two parts: z_T from the tree encoder and z_G from the graph encoder. Each part of z is sampled from a Gaussian distribution derived from the respective encoder outputs.

Decoder

The decoder also has two components corresponding to the two encoders:

Table S1. The hyper-parameters for VAE training.

Hyper-parameter	Value	Description
batch_size	32	The input batch size
hidden_size	450	The size of the hidden layers in the model
latent_size	64	The dimensionality of the latent space
depthG	3	The number of GNN layers in the model
lr	0.001	Initial learning rate
clip_norm	50.0	Maximum norm for gradient clipping
max_beta	1.0	Maximum value for beta in KL annealing
warmup	40000	Number of steps for learning rate warmup
epoch	20	Total number of training epochs
anneal_rate	0.9	Annealing rate for learning rate adjustment
anneal_iter	40000	Iterations over which to anneal the learning rate

Table S2. Atom and Bond features.

	Features	Size	Description
Atom	Atom type	23	The atom type (e.g., C, N, O), by atomic number
	Number of H	6	The number of bonded hydrogen atoms
	Charge	5	The formal charge of the atom
	Chirality	4	The chiral-tag of the atom
	Is-aromatic	1	Whether the atom is part of an aromatic system or not
Bond	Bond type	5	The bond type (e.g., single, double, triple et al.)
	Stereo	6	The stereo-configuration of the bond

1. **Tree Decoder:** This decodes the latent tree representation z_T back into a junction tree structure. It operates in a top-down fashion, predicting the structure and labels of the nodes in the tree.
2. **Graph Decoder:** After reconstructing the junction tree, this component predicts the detailed connectivity between the clusters based on z_G. It ensures that the final output is a chemically valid molecular graph that corresponds to the encoded molecule.

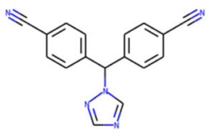
Input to the Model

The input to our model is a molecular graph, where each atom is represented as a vertex and each bond as an edge. Feature vectors for vertices and edges include atom type, valence, bond type, and other chemical properties.

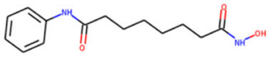
GNN Architecture

The GNN used in both the graph encoder and decoder employs a message passing mechanism where messages are exchanged in a loopy belief propagation fashion, iterating through multiple steps to refine the encoding of graph structure.

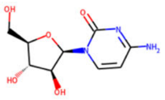
Supplementary Section 4: Supplementary Experimental Results



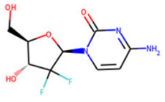
0



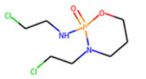
1



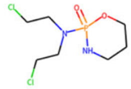
2



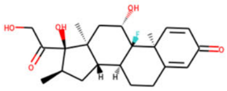
3



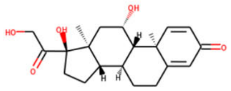
4



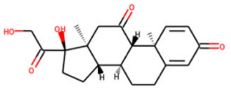
5



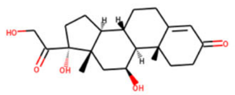
6



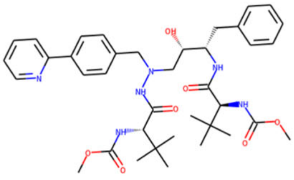
7



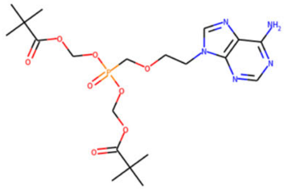
8



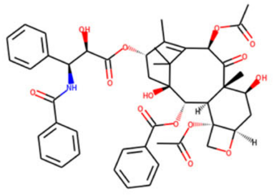
9



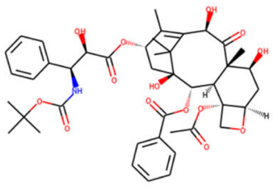
10



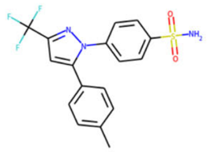
11



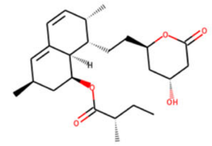
12



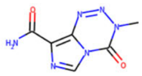
13



14



15



16

Fig. S1. The structures of 17 FDA-approved but toxic molecules.

Table S3. Metabolic half-lives of small molecules in mouse and human liver microsomes

SMILES	Mouse	Human
ClC1=CC(NN=C2NC3CCN(S(=O)(C)=O)CC3)=C2C=C1C4CCN(C)CC4	120	120
O=C(C1COC1)NC2=CC=C(C=C2)C3=CSC4=CN=C(NC5CCN(S(=O)(C)=O)CC5)N=C43	8	45.6
O=S(N(CC1)CCC1NC2=NC=C3C(C(C4=CC=C(NC(OC)=O)C=C4)=CS3)=N2)(C)=O	16	45.3
O=S(N(CC1)CCC1NC2=NC=C3C(C(C4=CC=C(C(NCC5(F)COC5)=O)C=C4)=CS3)=N2)(C)=O	1.95	9
O=C(OC)NC1=CC=C(C2=CSC3=CN=C(NC4CCN(S(=O)(C)=O)CC4)N=C32)C=C1	21.5	120
O=C(OC)NC1=CC=C(C(F)=C1)C2=CSC3=CN=C(NC4CCN(S(=O)(C)=O)CC4)N=C32	31.1	56.8
O=C(OC1COC1)NC2=CC=C(C=C2)C3=CSC4=CN=C(NC5CCN(S(=O)(C)=O)CC5)N=C43	7.73	31.4
O=C(OC)NC1=CC=C(C(F)=C1)C2=CSC3=CN=C(NC4CCN(S(=O)(CCN)=O)CC4)N=C32	38.3	120
NCCS(N(CC1)CCC1NC2=NC=C3C(C(C4=C(F)C=C(C(NCC5COCC5)=O)C=C4)=CS3)=N2)(=O)=O	58.2	120
O=C(OC)NC1=CC=C(C(C)=C1)C2=CSC3=CN=C(NC4CCN(S(=O)(CCN)=O)CC4)N=C32	31.6	120
O=C(OC(C)C)NC1=CC=C(C(F)=C1)C2=CSC3=CN=C(NC4CCN(S(=O)(CCN)=O)CC4)N=C32	51	120
O=C(OCC(F)(F)F)NC1=CC=C(C(F)=C1)C2=CSC3=CN=C(NC4CCN(S(=O)(CCN)=O)CC4)N=C32	30.9	120
O=C(OC)NC1=CC=C(C(Cl)=C1)C2=CSC3=CN=C(NC4CCN(S(=O)(CCN)=O)CC4)N=C32	26.1	120
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4OCCCCC5(CCNCC5)CN3)N2N=C1	15	27
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4OCCCCCC5(CCNCC5)CN3)N2N=C1	24.3	10.9
CC(C)C1=C2N=C3N=C(NCC4=CC(F)=CC=C4OCCCCC5(CCNCC5)CN3)N2N=C1	16.5	34.5
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4OCC@HCCCC5(CCNCC5)CN3)N2N=C1	48.5	21.3
CC(C)C1=C2N=C3N=C(NCC4=CC(C(F)(F)F)=CC=C4OCCCCC5(CCNCC5)CN3)N2N=C1	114	92.4
CC(C)C1=C2N=C3N=C(NCC4=CC(C(F)(F)F)=CC=C4OCCCCCC5(CCNCC5)CN3)N2N=C1	316	65.4
CC(C)C1=C2N=C3N=C(NCC4=CC=CC(N4CCCCC5(CCNCC5)CN3)=O)N2N=C1	4.77	43.9
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4OCC(O)CCCC5(CCNCC5)CN3)N2N=C1	38.5	34.1
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4OCCCCCC5(CCNCC5)CN3)N2N=C1	24.3	10.9
CC(C)C1=C2N=C3N=C(NCC4=CC(OC)=CC=C4OCCCCCC5(CCNCC5)CN3)N2N=C1	39.8	20.7
CC(C)C1=C2N=C3N=C(NCC4=CC=NC=C4OCCCCCC5(CCNCC5)CN3)N2N=C1	5.1	5.07
CC(C)C1=C2N=C3N=C(NCC4=CC=C(C)N=C4OCCCCCC5(CCNCC5)CN3)N2N=C1	15.7	46.8
CC(C)C1=C2N=C3N=C(NCC4=CC=CC(N4CCCCCCC5(CCNCC5)CN3)=O)N2N=C1	13.2	63
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4COCCCCC5(CCNCC5)CN3)N2N=C1	39.1	31.2
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4COC/C=C/CC5(CCNCC5)CN3)N2N=C1	14.4	24.1
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4OCC@HCCCC5(CCNCC5)CN3)N2N=C1	59.7	46.2
CC(C)C1=C2N=C3N=C(NCC4=CC=CC=C4OCCCCCN3CC5CCNCC5)N2N=C1	4.99	84.5
CC1=C(CP(C)(C)=O)C2=C(C=C1)C(C3=NC(N[C@H]4CCCNC4)=NC5=C3SC=C5)=CN2	120	8.6
CP(C1=CC=CC2=C1NC=C2C3=NC(N[C@H]4CCC@@HNC4)=NC5=C3SC=C5)(C)=O	120	10.6
CP(C1=CC=CC2=C1NC=C2C3=NC(N[C@H]4CCNC4)=NC5=C3SC=C5)(C)=O	——	10.7
O=P(C)C1=CC=CC2=C1NC=C2C3=NC(N[C@H]4CC@@HNC4)=NC5=C3SC=C5	——	10.6
CP(C1=CC=CC2=C1NC=C2C3=NC(N[C@H]4C@@HCNC4)=NC5=C3SC=C5)(C)=O	——	6.2
CP(C1=CC=CC2=C1NC=C2C3=NC(NC4C5C4CNC5)=NC6=C3SC=C6)(C)=O	——	6.6
CP(C1=CC=CC2=C1NC=C2C3=NC(N[C@@H]4CC@@HCC4)=NC5=C3SC=C5)(C)=O	120	48.8
CP(C1=C(C#N)C=CC2=C1NC=C2C3=NC(N[C@H]4CCCNC4)=NC5=C3SC=C5)(C)=O	120	23.3
CP(C1=CC(F)=CC2=C1NC=C2C3=NC(N[C@H]4CCCNC4)=NC5=C3SC=C5)(C)=O	120	14.5
CP(C1=CC=C(F)C2=C1NC=C2C3=NC(N[C@H]4CCCNC4)=NC5=C3SC=C5)(C)=O	120	14.1
CP(C1=CC(C(F)(F)F)=CC2=C1NC=C2C3=NC(N[C@H]4CCCNC4)=NC5=C3SC=C5)(C)=O	——	85.6
CP(C1=CC(C#N)=CC2=C1NC=C2C3=NC(N[C@H]4CCCNC4)=NC5=C3SC=C5)(C)=O	120	106
CP(C1=NC=CC2=C1NC=C2C3=NC(N[C@H]4CCCNC4)=NC5=C3SC=C5)(C)=O	120	17.9
CP(C1=C(C#N)C=CC2=C1NC=C2C3=NC(N[C@H]4CCCN(C)C4)=NC5=C3SC=C5)(C)=O	——	12.6
CP(C1=C(C#N)C=CC2=C1NC=C2C3=NC(N[C@@H]C)=NC4=C3SC=C4)(C)=O	——	120
O=P(C)(C)C1=C(C#N)C=CC2=C1NC=C2C3=NC(N[C@H]4CC@@HCNC4)=NC5=C3SC=C5	120	120
CP(C1=C(C#N)C=CC2=C1NC=C2C3=NC(N[C@H]4CC@HCNC4)=NC5=C3SC=C5)(C)=O	72.2	50.6
CP(C1=C(C#N)C=CC2=C1NC=C2C3=NC(NC4CC(F)(F)CNC4)=NC5=C3SC=C5)(C)=O	22.6	20.7
O=P(C)(C)C1=C(C#N)C=CC2=C1NC=C2C3=NC4=C(SC=C4)C(N[C@H]5CCCNC5)=N3	120	103

Table S4. Molecular Count in Datasets for Training and Prediction Tasks in Drug-likeness Studies

Task	Dataset	Molecular Count
Train/Drug-likene ss prediction	CD	4527
	ChEMBL	943683
	ZINC	249451
	GDB	1048547
Drug-likeness prediction	Anticancer drugs[39]	195
	SIDER	1427
	BBBP	2039
	ToxCast	8576
	Tox21[42]	7831
	ClinTox[42]	1478
	FreeSolv	642
	ESOL	1128
	qm7	6830
	qm8	21786
	WITHDRAW[44]	240

Table S5. Pros and Cons of Classical QED, Current QED (ADMET-score), and DrugMetric in Drug-likeness Evaluation

Feature / Method	Classical QED (Bickerton et al., 2012)	Current QED (ADMET-score, 2019)	DrugMetric
Core Concept	Evaluates drug-likeness based on a desirability function over molecular properties.	Evaluates drug-likeness based on 18 ADMET properties.	Introduces an unsupervised learning framework combining VAE and GMM to enhance precision and reliability in drug-likeness evaluation.
Pros	Straightforward and intuitive. Transparent scoring system. Easily implemented and integrated.	Comprehensive evaluation of ADMET properties. Aligned with pharmacokinetic considerations. Useful for ADMET predictions.	Advanced AI techniques (VAEs, GMM) enhance distinction capabilities. Effectively utilizes unlabeled data to overcome traditional method limitations. Proven robustness and higher accuracy across various datasets, improving drug discovery processes.
Cons	May miss biological aspects of drug-likeness. Limited to static molecular descriptors.	Complex model interpretation. - Requires extensive data for accuracy.	Higher computational resources initially required. Steeper learning curve for understanding model intricacies.
Data Requirements	Low; basic molecular descriptors.	High; requires accurate ADMET property models.	Moderate; leverages unlabeled as well as labeled data effectively.
Computational Complexity	Low; relies on simple calculations of molecular properties.	Moderate; involves complex predictions based on multiple ADMET properties.	Moderate to High; employs advanced AI but optimized for efficiency. Model Parameters: 17,426K. Training Time: 20 hours on a single NVIDIA 3090 GPU.

Table S6. Criteria for Various Rules Used in Drug Screening

Rule	Drug Screening Criteria
Lipinski's Rule of Five	Molecular weight $\leq$ 500 Da, LogP $\leq$ 5, hydrogen bond donors $\leq$ 5, hydrogen bond acceptors $\leq$ 10, violations $\leq$ 1.
Pfizer Rule	Molecular weight $\leq$ 480 Da, CLogP $\leq$ 5, hydrogen bond donors $\leq$ 5, hydrogen bond acceptors $\leq$ 10, polar surface area $\leq$ 140 Å^2, violations $\leq$ 1.
GSK Rule	Molecular weight $\leq$ 500 Da, CLogP $\leq$ 5, hydrogen bond donors $\leq$ 3, hydrogen bond acceptors $\leq$ 3, rotatable bonds $\leq$ 10.
Golden Triangle Rule	Molecular weight ranging from 200-600 Da, LogP ranging from -0.4 to 5.6, with a negative correlation between molecular weight and LogP.
QED	Overall score: 0 to 1, a higher QED score indicates greater drug-likeness, approaching 1.

Table S7. Molecular properties of FDA-approved but clinically toxic compounds.

Index	SMILES	QED	DrugMetric
0	C1=CC(=CC=C1C#N)C(C2=CC=C(C=C2)C#N)N3C=NC=N3	74.07	90.13
1	C1=CC=C(C=C1)NC(=O)CCCCCCC(=O)NO	38.32	86.03
2	C1=CN(C(=O)N=C1N)[C@H]2[C@H]([C@@H]([C@H](O2)CO)O)O	44.89	92.04
3	C1=CN(C(=O)N=C1N)[C@H]2C(C@@HO)(F)F	61.21	91.33
4	C1CN(P(=O)(OC1)NCCCl)CCCl	60.57	84.77
5	C1CNP(=O)(OC1)N(CCCl)CCCl	60.57	79.47
6	C[C@@H]1C[C@H]2[C@@H]3CCC4=CC(=O)C=C[C@@]4([C@]3([C@H](C[C@@]2([C@]1(C(=O)CO)O)C)O)F)C	66.72	90.77
7	C[C@]12C[C@@H]([C@H]3[C@H]([C@@H]1CC[C@@]2(C(=O)CO)O)CCC4=CC(=O)C=C[C@]34C)O	69.46	88.36
8	C[C@]12CC(=O)[C@H]3[C@H]([C@@H]1CC[C@@]2(C(=O)CO)O)CCC4=CC(=O)C=C[C@]34C	78.48	88.69
9	C[C@]12CCC(=O)C=C1CC[C@@H]3[C@@H]2[C@H](C[C@]4([C@H]3CC[C@@]4(C(=O)CO)O)C)O	69.6	84.86
10	CC(C)(C)[C@@H](C(=O)N[C@@H](CC1=CC=CC=C1)[C@H](CN(CC2=CC=C(C=C2)C3=CC=CC=N3)NC(=O)[C@H](C(C)(C)C)NC(=O)OC)O)NC(=O)OC	15.43	84.88
11	CC(C)(C)C(=O)OCOP(=O)(COCCN1C=NC2=C1N=CN=C2N)OCOC(=O)C(C)(C)C	20.72	81.18
12	CC1=C2[C@H](C(=O)[C@@]3([C@H](C[C@@H]4[C@]([C@H]3[C@@H]([C@@](C2(C)C)(C[C@@H]1OC(=O)[C@@H]([C@H](C5=CC=CC=C5)NC(=O)C6=CC=CC=C6)O)O)OC(=O)C7=CC=CC=C7)(CO4)OC(=O)C)O)C)OC(=O)C	12.98	74.99
13	CC1=C2[C@H](C(=O)[C@@]3([C@H](C[C@@H]4[C@]([C@H]3[C@@H]([C@@](C2(C)C)(C[C@@H]1OC(=O)[C@@H]([C@H](C5=CC=CC=C5)NC(=O)OC(C)(C)C)O)O)OC(=O)C6=CC=CC=C6)(CO4)OC(=O)C)O)C)O	14.68	85.81
14	CC1=CC=C(C=C1)C2=CC(=NN2C3=CC=C(C=C3)S(=O)(=O)N)C(F)(F)F	75.41	84.23
15	CC[C@H](C)C(=O)O[C@H]1C[C@H](C=C2[C@H]1[C@H]([C@H](C=C2)C)CC[C@@H]3C[C@H](CC(=O)O3)O)C	67.2	89.99
16	CN1C(=O)N2C=NC(=C2N=N1)C(=O)N	56.01	86.33

Table S8. Comparison of DrugMetric and QED Scores for Selected Drugs

Drug	DrugMetric Score	QED Score
Bortezomib	77.94	46.30
Thalidomide	90.39	72.34
Warfarin	88.06	74.76