Supplementary Materials

Sphere Stacking

The initial graph used in the graph translation module for topology generation is derived using a face-centered cubic (FCC) stacking model. This model is employed to fill a defined space, connecting the first and second nearest neighbors to simulate potential molecular structures. Given the fixed positioning and edge angles in this model, questions arise regarding its capability to accurately reflect the geometric diversity of molecules, which exhibit a wide range of bond lengths and angles.

To ascertain the suitability of FCC stacking as an approximation for molecular geometry, we designed a molecular graph restoration task. We selected a sample of 20,000 molecules with 3D conformations from the PubChem dataset, representing their shapes using van der Waals (VDW) radii. These shapes were then filled using FCC stacking, incorporating random orientations and slight positional perturbations. The assignment of atoms to their nearest spheres was evaluated to determine if bonds could be accurately formed with the first or second nearest neighbors. This procedure was repeated ten times for each stacking instance. A restoration was deemed successful if, in at least ten attempts, a molecular graph could be accurately reconstructed from the stacked spheres.

Out of the 20,000 molecules analyzed, only 15 failed this restoration task, providing strong evidence that FCC stacking is a sufficiently accurate method for generating basic patterns in graph translation.

Model Architecture

The core of TopMT-GAN's efficiency lies in its distinctive approach to both topology generation and molecular assignment, where each module is trained through adversarial loss exclusively. The foundational elements of this architecture are the Node-Edge Co-evolution Translation (NECT) layers. These NECT blocks are designed to not only aggregate node (atom) information but also integrate edge (bond) information, facilitating a comprehensive representation of the relationship between atoms and bonds within a molecule.

For the topology generation module, a deeper generator architecture is employed to encapsulate global molecular structures effectively. This depth is crucial for accurately predicting the overarching shape and connectivity of potential ligands, ensuring that the generated topologies are viable for further refinement. Conversely, the molecular assignment module utilizes shallower layers in its generator design. This strategic simplification prioritizes the identification and incorporation of local substructures, aiming to foster diversity within the generated molecules. The topology generation module is depicted in Figure S1, while the molecular assignment module can be seen in Figure S2.

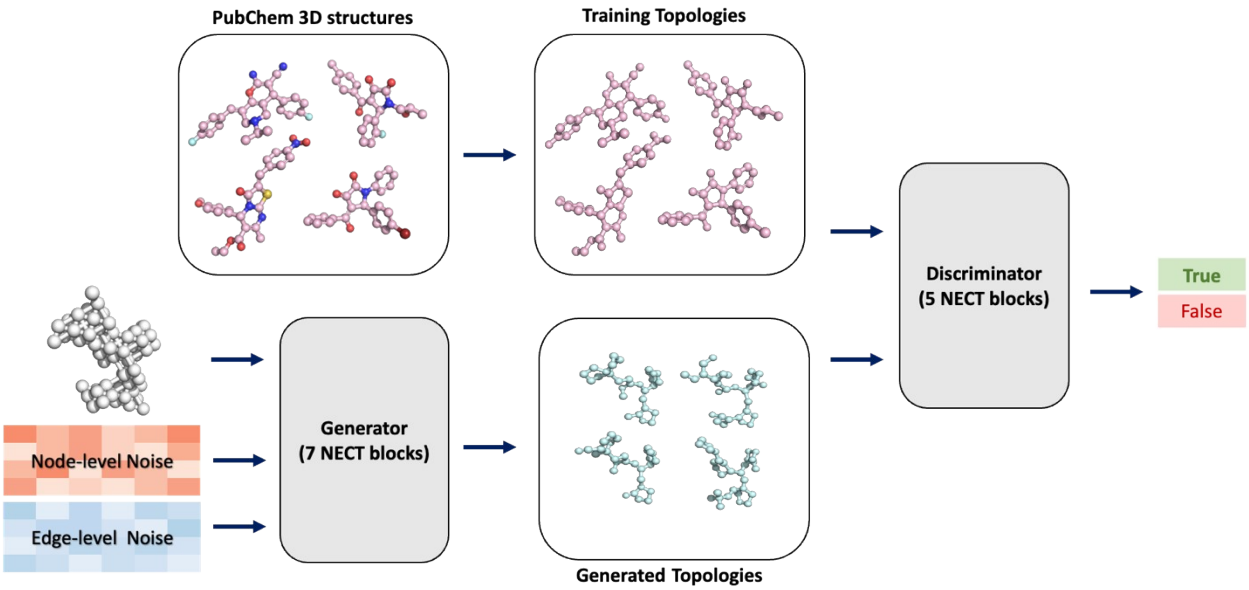

Figure S1. Model architecture of topology generation GAN.

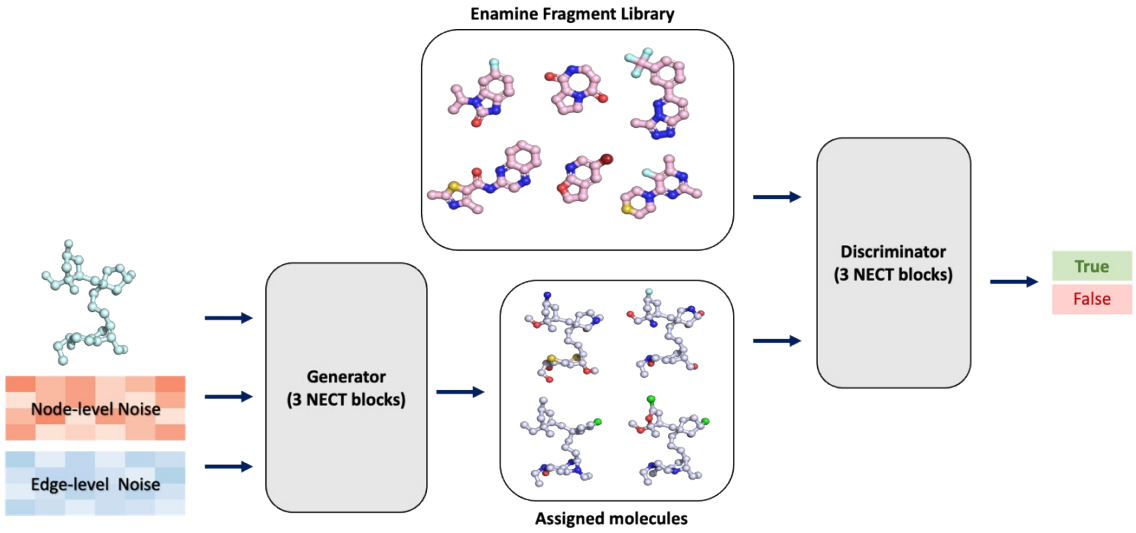

Figure S2. Model architecture of molecular assignment GAN.

Evaluations

The five protein targets chosen for evaluation with TopMT-GAN are detailed in Table S1. The pockets in PDB IDs 7d3i, 1e3g, and 7wf5 are orthosteric pockets, whereas those in 3jvs and 5vew are identified as allosteric. Analysis based on the ratio of solvent-accessible surface area (SASA) to volume reveals that allosteric pockets tend to be shallower. The selection of 7wf5 and 3jvs, both kinases with very similar structures, was deliberate, aimed at assessing the robustness of TopMT-GAN across different pocket types. The original ligands and detected pockets are shown in (7d3i: Figure 5; 1e3g, 5vew: Figure S3; 7wf5, 3jvs: Figure 3).

Known actives for these targets were collected from the Binding Database. It is important to note that these molecules might not bind to identical pocket locations. While many of these ligands don't have crystal structure data, they sufficiently serve as a baseline for comparative analysis. The vina score histogram of generated molecules, enamine HTS and known actives are shown in Figure S4. The vina score distributions compared with other models are shown in Figure S6.

LogP and SAS properties are compared with the enamine HTS library, known actives and 1 million molecules randomly selected from PubChem database and their distributions are shown in Figure S5. The distributions of maximum similarities to known actives are shown in Figure S7.

Target	PDB ID	SASA (A^2)	Volume(A^3)	Ratio	# Actives
3C-like protease	7d3i	468	365	1.28	101
Androgen Receptor	1e3g	336	227	1.48	108
Kinase (Orthosteric)	7wf5	591	404	1.46	5,565
Kinase (Allosteric)	3jvs	360	322	1.12	3,725
GLP-1 Receptor	5vew	464	400	1.16	796

Table S1. Summary of selected protein targets.

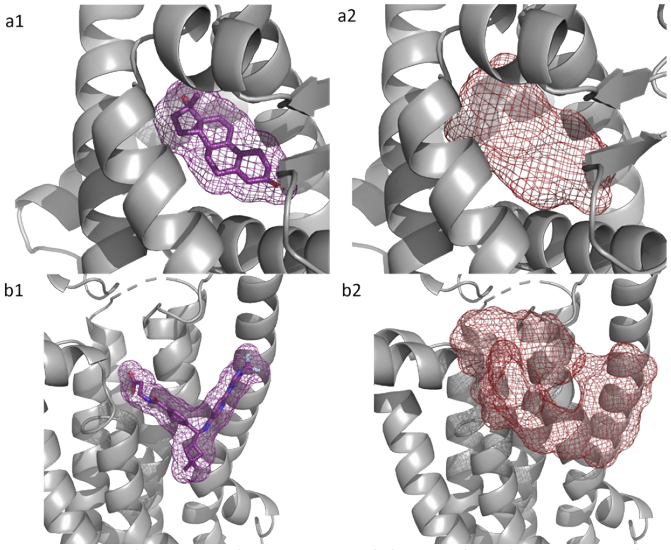

Figure S3. Ligands in crystal structure and detected pocket. a: 1e3g; b: 5vew

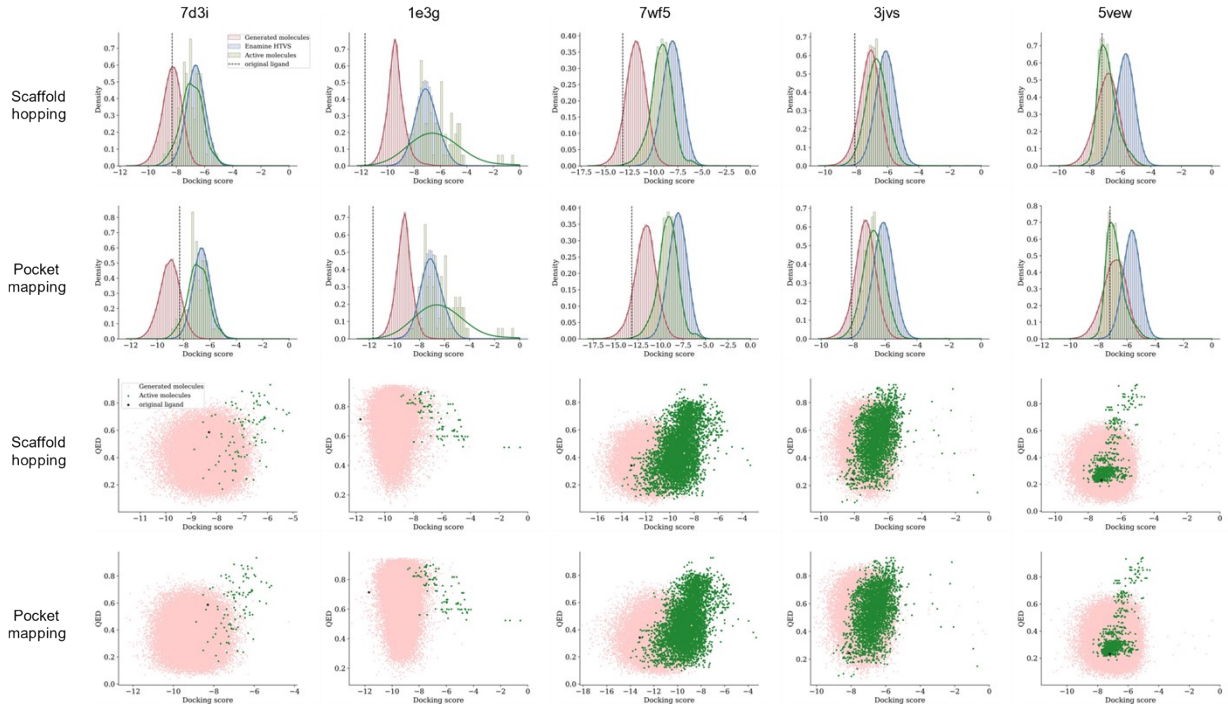

Figure S4. vina score distributions and QED-vina score scattering of molecules generated by TopMT-GAN compared to known active and PubChem molecules.

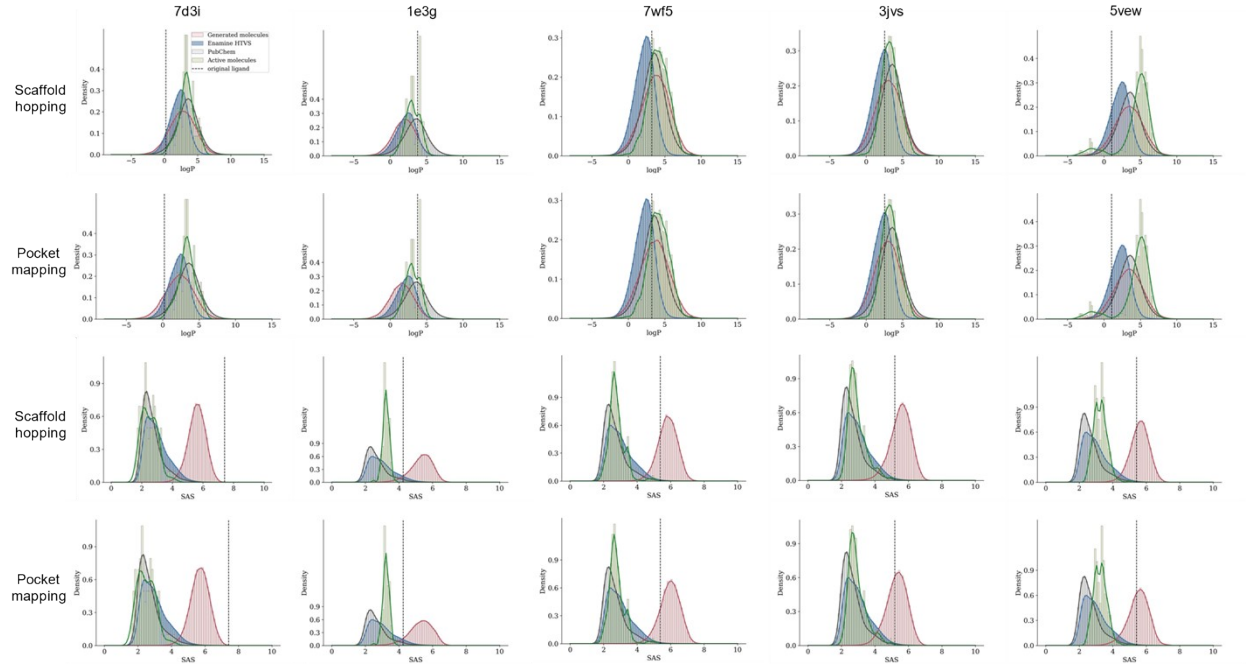

Figure S5. LogP and SAS distributions of molecules generated by TopMT-GAN compared to known active and PubChem molecules.

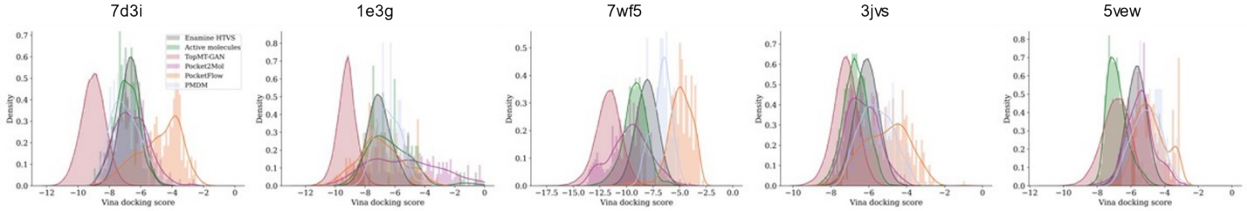

Figure S6. Comparison of Vina docking score distributions across different 3D structure-based generative models.

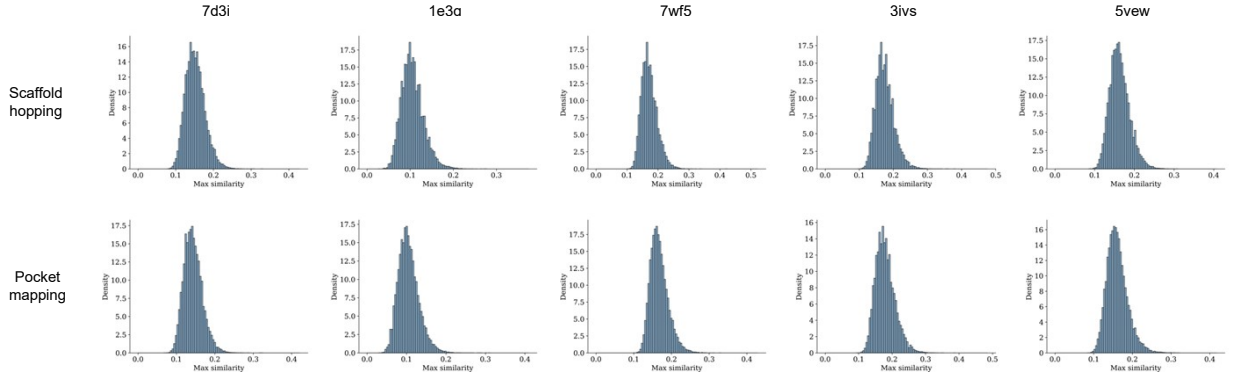

Figure S7. Maximum similarity distributions of molecules generated by TopMT-GAN to known actives.

Models	QED	SAS	LogP	Internal Diversity
TopMT-GAN	0.45 ± 0.19	5.6 ± 0.7	2.7 ± 2.0	0.88
PMDM	0.26 ± 0.15	5.6 ± 0.8	-0.8 ± 2.1	0.90
PocketFlow	0.46 ± 0.17	3.4 ± 1.1	1.6 ± 2.6	0.91
Pocket2Mol	0.65 ± 0.17	3.1 ± 1.3	3.0 ± 1.7	0.88

Table S2. Comparison of molecular properties across different 3D structure-based generative models.

PoseCheck was used to evaluate the quality of generated poses across different structure-based generative models. We assessed both steric clashes and strain energy for molecules generated by TopMT-GAN, PMDM, PocketFlow, and Pocket2Mol. For each model, we analyzed 100 generated molecules per target across five protein targets (total 500 molecules) from our benchmark dataset. The results are summarized in Table S3.

Redock scores compared with generated pose scores are summarized in Table S4. If the difference in docking score is less than 0.5 kcal/mol, we assume the scores are similar or equally good. For each system, the average percentages of redocked poses that were equally good or better than the generated poses are 94.5% for 3C-Like Protease, 78% for Androgen Receptor, 82% for c-SRC kinase, 88.5% for CHK1 kinase, and 94% for GLP-1 receptor. Overall, an average of 87.4% of redocked poses across all systems were either equally good or better than the generated poses.

Models	Average Clashes	Max Clashes	Average Strain Energy (kcal/mol)	Max Strain Energy (kcal/mol)
TopMT-GAN	27	62	1,170	6,385
PMDM	263	514	30,822	903,268
PocketFlow	32	193	6,939	824,928
Pocket2Mol	15	63	91	1,984

Table S3. Comparison of Steric Clashes and Strain Energy Across Different Structure-Based Molecular Generative models.

Target	Mode	Redock better Ratio	Redock Similar $\pm 0.5\ kcal/mol$	Redock Worse Ratio
3C-Like Protease PDB: 7d3i	Scaffold-hopping	0.50	0.44	0.06
	Pocket-mapping	0.62	0.33	0.05
Androgen Receptor PDB: 1e3g	Scaffold-hopping	0.07	0.68	0.25
	Pocket-mapping	0.13	0.68	0.19
c-SRC kinase PDB: 7wf5	Scaffold-hopping	0.47	0.35	0.18
	Pocket-mapping	0.50	0.32	0.18
CHK1 kinase PDB: 3jvs	Scaffold-hopping	0.46	0.48	0.06
	Pocket-mapping	0.24	0.59	0.17
GLP-1 receptor PDB: 5vew	Scaffold-hopping	0.79	0.20	0.01
	Pocket-mapping	0.47	0.42	0.11

Table S4. Redock scores compared with vina scores of generated poses.

Molecular generation efficiency of TopMT-GAN is summarized in Table S5, and its comparison to other models is shown in Table S6.

The hit rates from virtual high-throughput screening for each target are summarized in Table S7 for reference.

PDB	Mode	Topology Sampling (GPU-time/topology)	Molecule assignment (CPU-core-time/mol)	Scoring (CPU-core-time/mol)	Total Sample time Topology + assign & score (days)
7d3i	Scaffold-hopping	1.67s	0.37s	1.83s	0.97+0.17=1.14
	Pocket-mapping	0.36s	0.40s	2.30s	0.21+0.21=0.42
1e3g	Scaffold-hopping	0.19s	0.33s	1.64s	0.11+0.14=0.25
	Pocket-mapping	1.70s	0.32s	2.08s	0.99+0.18=1.17
7wf5	Scaffold-hopping	2.06s	0.43s	2.25s	1.20+0.20=1.40
	Pocket-mapping	2.90s	0.39s	2.33s	1.68+0.21=1.89
3jvs	Scaffold-hopping	0.12s	0.37s	1.68s	0.08+0.17=0.25
	Pocket-mapping	2.87s	0.33s	1.61s	1.66+0.16=1.82
5vew	Scaffold-hopping	0.49s	0.40s	2.26s	0.28+0.20=0.48
	Pocket-mapping	2.04s	0.36s	2.16s	1.21+0.20=1.41

Table S5. TopMT-GAN generation speed.

Model	#Gen Mols	#Systems	Baseline	Efficiency	Hardware
PMDM	100 for each target	100 CrossDock	NA	906s /100 mols	Tesla V100
PocketFlow	10,000 for each target & 100,000 for two cases	10 CrossDock 2 case studies	NA	~345s/100 mols	2*Tesla P100
ResGen	Not mentioned	100 CrossDock	200 random BindingDB actives	NA	NA
Lingo3DMol	~100 for each target	101 DUD-E	100 ligands CrossDock	874s/100 mols	Tesla V100
SurfGen	100 for each target	18 CrossDock	200 random BindingDB actives	NA	NA
TargetDiff	100 for each target	100 CrossDock	NA	3428s/100 mols	GTX 3090
DESERT	100 for each target	10 CrossDock	NA	~4 hours/100 mols	Tesla V100
GraphBP	100 for each target	10 CrossDock	NA	NA	NA
Pocket2Mol	100 for each target	10 CrossDock	NA	2503s/100 mols	Tesla V100
DeepLigBuilder	19,014 (MCTS)	1 (SARS-Cov-2 main protease)	NA	3-4hours/100 mols	RTX 3070
SBDD	100 for each target	10 CrossDock	NA	NA	NA
TopMT	50,000 for each target	5 diverse protein	**1.3 million Enamine HTVS 100-5000 BindingDB actives**	**~330s/100 mols**	Mixed Tesla A5000 & CPUs

Table S6. Performance comparison of TopMT-GAN with other 3D structure-based generative models (values are obtained from original papers).

Target	Redock score Kcal/mol	HR < redock	HR < -8	HR < -9	HR < -10	HR < -11	HR < -12
3C-Like Protease PDB: 7d3i	-8.3	0.72%	2.1%	0.036%	0.00023%	N/A	N/A
Androgen Receptor PDB: 1e3g	-11.7	0..000075%	11.0%	0.63%	0.012%	0.00015%	N/A
c-SRC kinase PDB: 7wf5	-13.2	0.00023%	52.8%	18.8%	3.3%	0.28%	0.013%
CHK1 kinase PDB: 3jvs	-8.1	0.15%	0.23%	0.0019%	N/A	N/A	N/A
GLP-1 receptor PDB: 5vew	-7.2	0.86%	0.030%	0.00038%	N/A	N/A	N/A

Table S7. Hit Rate of HTVS for the 5 benchmark systems at different vina score levels.