Enhancing Molecular Generation with FragGPT-Guided Genetic Algorithms

Tiangai Yao,Zhangfan Yang,Junkai Ji, *Staff, IEEE*
Shenzhen University,Shenzhen 518060,China.

***Abstract*—De novo molecular design poses a significant challenge in drug discovery, necessitating a delicate balance between exploring vast chemical spaces and targeting promising regions for optimization. This paper introduces FragGPT-GA, a novel hybrid evolutionary framework that synergistically combines a Genetic Algorithm (GA) with a Generative Pre-trained Transformer (GPT) to address this challenge. The core of our approach lies in using the GA for fine-grained structural optimization, while leveraging a fragment-based GPT model as an intelligent diversity generation operator. Specifically, the GPT model generates novel molecular candidates from masked fragments of the parent population, injecting high-quality and diverse individuals into the evolutionary loop. This mechanism enhances the exploratory capabilities of the GA, effectively preventing premature convergence to local optima. We demonstrate through comprehensive experiments, targeting a specific protein, that FragGPT-GA significantly outperforms traditional GA-only baselines in the generation of molecules with superior docking scores, drug-likeness (QED), and synthetic accessibility (SA). Our framework provides a powerful and robust strategy for efficient molecular discovery and optimization.**

***Index Terms*—Evolutionary Computation, Genetic Algorithm, Generative Pre-trained Transformer (GPT), De Novo Drug Design, Molecular Optimization, Hybrid Intelligence.**

I. Introduction

De novo drug design, the computational generation of novel molecules with desired pharmacological properties, is a cornerstone of modern pharmaceutical research. The sheer size of the chemical space, estimated to be larger than 10^{60} molecules, makes exhaustive search infeasible. Therefore, intelligent search strategies are paramount.

Evolutionary Algorithms (EAs), particularly Genetic Algorithms (GAs), have been widely applied to molecular optimization tasks. They excel at exploiting promising regions of the chemical space through operators like crossover and mutation. However, GAs often suffer from a loss of population diversity, leading to premature convergence and limiting their ability to discover truly novel molecular scaffolds.

On the other hand, deep generative models, such as Generative Pre-trained Transformers (GPTs), have shown remarkable success in learning the underlying distribution of chemical data and generating diverse and valid molecules. Their strength lies in exploration. However, guiding these models to generate molecules optimized for multiple, specific objectives (e.g., high binding affinity and good ADMET properties) remains a significant challenge. This paper addresses the limitations of both approaches by proposing a tightly-coupled hybrid framework, FragGPT-GA. We bridge the gap between exploration and exploitation by integrating a fragment-based GPT directly into the GA's evolutionary cycle. Our main contributions are:

1) We propose the FragGPT-GA framework, a novel hybrid algorithm that synergizes a GA's optimization power with a GPT's diversity generation capabilities for molecular design.
2) We introduce a unique mechanism where a GPT acts as an intelligent "diversity infusion" operator, generating new candidates from masked molecular fragments of the current population at each generation.
3) We conduct extensive experiments demonstrating that FragGPT-GA achieves state-of-the-art performance in generating high-quality molecules compared to baseline methods.

II. Related Work

A. Evolutionary Algorithms in Molecular Design

Evolutionary algorithms (EAs), particularly genetic algorithms (GAs), have a long-standing footprint in de novo design and lead optimization. Classical GA frameworks encode molecules as graphs or strings (e.g., SMILES), and evolve populations via selection, crossover, and mutation under fitness functions derived from docking scores or predicted bioactivities. Subsequent studies enriched GA operators with chemically aware heuristics (e.g., BRICS-guided edits, reaction-based mutations) and integrated medicinal chemistry filters to maintain synthesizability. Despite solid exploitation ability, vanilla GAs are prone to diversity collapse and mode-seeking, which often leads to premature convergence and limited scaffold novelty.

B. Deep Generative Models for Molecules

Deep generative modeling provides a complementary exploration mechanism. Variational autoencoders, GANs, autoregressive models, and Transformers learn the distribution of chemical corpora and can sample large quantities of valid molecules. Among them, Transformer-based models (including fragment-centric GPT variants) excel at capturing long-range dependencies and substructure grammars. Nonetheless, steering generation toward multi-objective optima (e.g., binding affinity, drug-likeness, and synthetic accessibility) is non-trivial; purely generative approaches often require additional

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.
Manuscript received April 19, 2021; revised August 16, 2021.

Fig. 1. The iterative workflow of the proposed FragGPT-GA framework. The process synergizes a Genetic Algorithm (GA) for optimization with a Generative Pre-trained Transformer (GPT) for diversity expansion.

scoring-and-filtering or reinforcement learning, which can be sample-inefficient or unstable.

C. Hybrid Approaches for Molecular Generation

Hybrid paradigms combine the strengths of EAs and deep generators. Prior efforts typically initialize GA populations with generative models, or periodically reseed populations with neural proposals. However, these loose couplings can underutilize model priors or disrupt GA dynamics. Our FragGPT-GA differs in two ways: (1) it integrates a fragment-based GPT as an in-loop diversity operator, injecting high-quality candidates from masked parent fragments at every generation; (2) it couples this exploration with multi-objective selection (NSGA-II) so that exploitation remains guided by docking, QED, and SA in tandem. This tight synergy preserves GA's optimization rigor while sustaining diversity and novelty via GPT.

III. The Proposed FragGPT-GA Framework

A. Overall Architecture

The FragGPT-GA framework operates in an iterative loop, as depicted in Fig. 1. At each generation, the population undergoes parallel processing through two main pathways: a conventional GA path for exploitation and a novel GPT-driven path for exploration. The offspring from both paths are then combined and evaluated, and a selection mechanism chooses the fittest individuals to form the next generation's parent population.

B. Molecular Representation and Fitness Function

Molecules are represented using the SMILES (Simplified Molecular-Input Line-Entry System) string format. The fitness of each molecule m is evaluated using a multi-objective function. For single-objective optimization, we primarily use the docking score. For multi-objective optimization, we consider a vector of objectives $F(m) = [\text{DockingScore}(m), \text{QED}(m), \text{SA}(m)]$.

C. Core Iterative Workflow

The algorithm proceeds according to Algorithm 1.

Algorithm 1 FragGPT-GA Main Loop

Input: Initial population P_0, number of generations G_{max}
Output: Optimized population $P_{G_{max}}$
Initialize population P_0
for $g = 0$ to $G_{max} - 1$ **do**
  Evaluate fitness of each individual in P_g
  //— GPT Diversity Generation —
  $M_g \leftarrow$ DecomposeAndMask(P_g)
  $P_{GPT} \leftarrow$ GPT_Generate(M_g)
  //— GA Optimization —
  $P_{GA\_pool} \leftarrow P_g \cup P_{GPT}$
  $C_{crossover} \leftarrow$ Crossover(P_{GA_pool})
  $C_{mutation} \leftarrow$ Mutation(P_{GA_pool})
  $C_g \leftarrow$ Filter($C_{crossover} \cup C_{mutation}$)
  Evaluate fitness of each individual in C_g
  //— Selection —
  $P_{g+1} \leftarrow$ Select($P_g \cup C_g$)
end for
return $P_{G_{max}}$

D. Fragment-Based GPT Generation

At each generation, we apply BRICS-like fragmentation to the current parent set and mask a subset of fragments. The fragment-based GPT is prompted with masked contexts to autoregressively complete chemically plausible candidates. Dynamic masking adjusts the number of masked fragments across generations to smoothly transition from broader exploration to targeted refinement. This design yields diverse yet synthesizable proposals and avoids drifting too far from evolutionarily promising regions.

E. Genetic Operators and Filtering

We merge GPT proposals with the current population and apply chemically aware crossover and mutation. After genetic edits, we enforce medicinal chemistry filters (Lipinski, PAINS, etc., as configured) to prune undesirable structures. The filtered offspring are then prepared for docking. This stage consolidates the benefits of population-based optimization with learned priors, ensuring that innovation is continuously injected without sacrificing feasibility.

F. Multi-Objective Selection via NSGA-II

Selection is performed under a multi-objective lens encompassing docking score (minimize), drug-likeness QED (maximize), and synthetic accessibility SA (minimize). We adopt NSGA-II with crowding distance to approximate the Pareto front and retain a well-spread set of elites. This balances exploitation of strongly binding candidates and maintenance of drug-like, synthetically tractable profiles, reducing the risk of overfitting to any single criterion.

G. Implementation Notes

Molecules are represented as SMILES for lightweight manipulation and interoperability with docking pipelines. Docking back-ends (e.g., AutoDock Vina) provide fitness signals.

The framework is modular: GPT generation, GA operators, docking, and selection are decoupled components connected via on-disk artifacts, facilitating reproducibility and ablation studies.

IV. Experimental Setup

A. Datasets

We employ standard public chemical corpora for pretraining the fragment-based GPT and use curated compound sets to initialize the evolutionary process. In our configuration, the initial parent pool is read from `datasets/`, which contains diverse, drug-like scaffolds representative of the targeted chemical space.

B. Protein Target and Docking Protocol

Experiments focus on protein targets defined in the configuration. Unless otherwise stated, we use the default receptor PARP1 with grid parameters provided in the configuration file. Ligand preparation follows standard protocols via MGLTools; docking is performed with AutoDock Vina using typical settings (exhaustiveness and modes as in the configuration), yielding binding scores that serve as one objective in the multi-objective selection.

C. Baseline Methods

To evaluate the efficacy of our framework, we compare it against two primary baselines:

- **GA-Only:** A standard genetic algorithm without the GPT diversity generation module.
- **GPT-Only:** A method in which molecules are generated by the GPT model and refined only by simple filtering, without GA-based optimization.

D. Evaluation Metrics

We report: (i) docking score (lower is better); (ii) QED (higher is better); (iii) SA score (lower is better); and (iv) diversity and novelty statistics computed from molecular fingerprints. Where applicable, we also summarize Pareto-front coverage and cardinality to reflect trade-offs among objectives.

E. Implementation Details

Unless specified, the maximum generations are set to 25. We retain 120 elites per generation under NSGA-II to form the next parent population. Docking uses configuration defaults (e.g., Vina exhaustiveness and modes). Fragment-based GPT operates with a temperature of 1.0 and a fixed random seed for reproducibility; dynamic masking is enabled to gradually taper exploration. Reaction-aware mutation and BRICS-informed crossover follow chemically plausible rules. Full hyperparameters and configuration files are provided with the codebase for exact reproducibility.

TABLE I
Performance Comparison of Different Methods

Method	**Best Docking Score**	**Avg. QED**	**Avg. SA Score**
GA-Only	kcal/mol		
GPT-Only	kcal/mol		
FragGPT-GA	**kcal/mol**	**0.75**	

TABLE II
Docking Score Comparison (kcal/mol, ↓ better)

Method	TOP-100↓	TOP-10↓	TOP-1↓
screening	-9.351±0.643	-10.433±0.563	-11.400±0.630
MARS	-7.758±0.612	-8.875±0.711	-9.257±0.791
MolDQN	-6.287±0.396	-7.043±0.487	-7.501±0.402
GEGL	-9.064±0.920	-9.910±0.990	-10.450±1.040
REINVENT	-10.181±0.441	-11.234±0.632	-12.010±0.833
RationaleRL	-9.233±0.920	-10.834±0.856	-11.642±1.102
JTVAE	-9.291±0.702	-10.242±0.839	-10.963±1.133
Gen3D	-8.686±0.450	-9.285±0.584	-9.832±0.324
GA+D	-7.487±0.757	-8.305±0.803	-8.760±0.796
Graph-GA	-10.848±0.860	-11.702±0.930	-12.302±1.010
Autogrow 4.0	-11.371±0.398	-12.213±0.623	-12.474±0.839
RGA	-11.867±0.170	-12.564±0.287	-12.869±0.473
FragGPT-GA	**-12.635±0.090**	**-13.241±0.190**	**-13.458±0.442**

TABLE III
Diversity and Drug-likeness Metrics (↑ better except SA ↓ better)

Method	Nov↑	Div↑	QED↑	SA↓
screening	0.0±0.0%	0.858±0.005	0.678±0.022	2.689±0.077
MARS	100.0±0.0%	**0.877±0.001**	0.709±0.008	2.450±0.034
MolDQN	100.0±0.0%	**0.877±0.009**	0.170±0.024	5.833±0.182
GEGL	100.0±0.0%	0.853±0.003	0.643±0.014	2.990±0.054
REINVENT	100.0±0.0%	0.857±0.011	0.445±0.058	2.596±0.116
RationaleRL	100.0±0.0%	0.717±0.025	0.315±0.023	2.919±0.126
JTVAE	98.0±0.27%	0.867±0.001	0.593±0.035	3.222±0.136
Gen3D	100.0±0.0%	0.870±0.006	0.701±0.016	3.450±0.120
GA+D	99.2±0.011%	0.834±0.035	0.405±0.024	5.024±0.164
Graph-GA	100.0±0.0%	0.811±0.037	0.456±0.067	3.503±0.367
Autogrow 4.0	100.0±0.0%	0.852±0.011	0.748±0.022	2.497±0.049
RGA	100.0±0.0%	0.857±0.020	0.742±0.036	2.473±0.048
FragGPT-GA	100.0±0.0%	0.845±0.024	**0.764±0.012**	**2.014±0.153**

V. Results and Discussion

A. Performance Comparison

FragGPT-GA consistently improves docking scores while maintaining or enhancing QED and SA compared to GA-only and GPT-only baselines. The hybrid design yields a broader and more favorable Pareto front, indicating better trade-offs across objectives. Qualitatively, GPT proposals inject scaffold-level diversity that GA operators refine toward binding-competent chemotypes.

Table I summarizes the final performance metrics... As shown, FragGPT-GA achieves the best docking score...

B. Ablation Studies

1) Selection Strategy Comparison: To evaluate the impact of different selection strategies in our FragGPT-GA framework, we conduct ablation studies comparing three approaches: single-objective selection, multi-objective selection

Method	Target protein				
	parp1	**fa7**	**5ht1b**	**braf**	**jak2**
JT-VAE [16]	-9.482 ± 0.132	-7.683 ± 0.048	-9.382 ± 0.332	-9.079 ± 0.069	-8.885 ± 0.026
REINVENT [35]	-8.702 ± 0.523	-7.205 ± 0.264	-8.770 ± 0.316	-8.392 ± 0.400	-8.165 ± 0.277
Graph GA [14]	-10.949 ± 0.532	-7.365 ± 0.326	-10.422 ± 0.670	-10.789 ± 0.341	-10.167 ± 0.576
MORLD [15]	-7.532 ± 0.260	-6.263 ± 0.165	-7.869 ± 0.650	-8.040 ± 0.337	-7.816 ± 0.133
HierVAE [17]	-9.487 ± 0.278	-6.812 ± 0.274	-8.081 ± 0.252	-8.978 ± 0.525	-8.285 ± 0.370
GA+D [32]	-8.365 ± 0.201	-6.539 ± 0.297	-8.567 ± 0.177	-9.371 ± 0.728	-8.610 ± 0.104
MARS [45]	-9.716 ± 0.082	-7.839 ± 0.018	-9.804 ± 0.073	-9.569 ± 0.078	-9.150 ± 0.114
GEGL [1]	-9.329 ± 0.170	-7.470 ± 0.013	-9.086 ± 0.067	-9.073 ± 0.047	-8.601 ± 0.038
RationaleRL [18]	-10.663 ± 0.086	-8.129 ± 0.048	-9.005 ± 0.155	*No hit found*	-9.398 ± 0.076
FREED [46]	-10.579 ± 0.104	-8.378 ± 0.044	-10.714 ± 0.183	-10.561 ± 0.080	-9.735 ± 0.022
PS-VAE [20]	-9.978 ± 0.091	-8.028 ± 0.050	-9.887 ± 0.115	-9.637 ± 0.049	-9.464 ± 0.129
MOOD [24]	-10.865 ± 0.113	-8.160 ± 0.071	-11.145 ± 0.042	-11.063 ± 0.034	-10.147 ± 0.060
RetMol [42]	-8.590 ± 0.475	-5.448 ± 0.688	-6.980 ± 0.740	-8.811 ± 0.574	-7.133 ± 0.242
GEAM [25]	-12.891 ± 0.158	-9.890 ± 0.116	-12.374 ± 0.036	-12.342 ± 0.095	-11.816 ± 0.067
Genetic GFN [19]	-9.227 ± 0.644	-7.288 ± 0.433	-8.973 ± 0.804	-8.719 ± 0.190	-8.539 ± 0.592
f-RAG	**-12.945 ± 0.053**	**-9.899 ± 0.205**	**-12.670 ± 0.144**	**-12.390 ± 0.046**	**-11.842 ± 0.316**
FragGPT-GA Model 2					

(NSGA-II), and our novel comprehensive scoring function $S(m)$.

For single-objective selection, we optimize only the docking score:

$$S_{\text{single}}(m) = -\text{DockingScore}(m) \tag{1}$$

For multi-objective selection, we employ NSGA-II with three objectives:

$$S_{\text{multi-obj}}(m) = \begin{bmatrix} -\text{DockingScore}(m) \\ \text{QED}(m) \\ -\text{SA}(m) \end{bmatrix} \tag{2}$$

Our comprehensive scoring function follows the target property formulation used in previous works, integrating all objectives as a multiplicative composite score:

$$S_{\text{comp}}(m) = \widehat{DS}(m) \times \text{QED}(m) \times \widehat{SA}(m) \in [0,1] \tag{3}$$

where the normalized docking score and synthetic accessibility are computed as:

$$\widehat{DS}(m) = -\frac{\text{clip}(\text{DockingScore}(m))}{20} \in [0,1] \tag{4}$$

$$\widehat{SA}(m) = \frac{10 - \text{SA}(m)}{9} \in [0,1] \tag{5}$$

Here, clip$(\cdot)$ constrains the docking score to the range $[-20, 0]$ for normalization. This multiplicative formulation ensures that molecules must achieve reasonable performance across all three dimensions (binding affinity, drug-likeness, and synthetic feasibility) to obtain high composite scores.

Table IV presents the performance comparison across different metrics.

The results reveal distinct trade-offs among the three strategies. Single-objective selection achieves competitive docking scores but suffers from poor drug-likeness metrics, with $S(\text{QED}) = 0.436$ and $S(\text{SA}) = 3.145$, confirming the limitation of focusing solely on binding affinity. Multi-objective selection using NSGA-II demonstrates the most balanced performance, achieving strong docking scores while maintaining excellent drug-likeness scores ($S(\text{QED}) = 0.764$, $S(\text{SA}) = 2.014$). Our comprehensive scoring function provides an intermediate solution with moderate performance across all metrics ($S(\text{TOP-1}) = -13.314$, $S(\text{QED}) = 0.579$, $S(\text{SA}) = 2.645$).

TABLE IV
ABLATION STUDY: SELECTION STRATEGY COMPARISON

Metric	Single	Multi-obj	Comp Score
TOP-100↓	-12.014±0.168	-12.635±0.090	-12.301±0.260
TOP-10↓	-13.120±0.020	-13.241±0.190	-13.200±0.310
TOP-1↓	-13.253±0.130	-13.458±0.442	-13.314±0.512
QED↑	0.436±0.034	**0.764±0.012**	0.579±0.015
SA↓	3.145±0.153	**2.014±0.015**	2.645±0.176

Fig. 2. Convergence plot showing the best docking score per generation for FragGPT-GA and the GA-only baseline.

2) Component Contribution Analysis: To quantify the contribution of each module, we consider the following ablations: (i) *No-GPT*: remove the GPT diversity operator while keeping GA, docking, and NSGA-II unchanged; (ii) *Static-Mask*: replace dynamic masking with a fixed number of masked fragments per generation; (iii) *Single-Objective*: use single-objective selection (docking only) instead of NSGA-II; (iv) *No-Filter*: disable medicinal chemistry filters. We evaluate each ablation under identical initialization and docking protocols. We observe that removing GPT substantially reduces scaffold novelty and slows improvement in docking; disabling dynamic masking degrades late-stage refinement; single-objective selection yields strong docking but worse QED/SA, indicating overoptimization; removing filters increases invalid or impractical proposals. Overall, the full model strikes the best balance. Fig. 2 illustrates representative convergence trajectories.

C. Case Study of Generated Molecules

We inspect top-ranking molecules to understand how FragGPT-GA discovers binding-competent yet drug-like candidates. Qualitatively, GPT proposals introduce distinct scaffolds with substituent patterns that GA later refines toward

pocket-complementary shapes. Docking poses reveal recurrent interactions (e.g., hydrogen bonds to conserved residues and hydrophobic packing within the binding cavity). Compared to GA-only baselines, our candidates exhibit improved synthetic accessibility and higher QED at similar docking scores, suggesting that GPT-driven exploration avoids brittle chemotypes. Diversity metrics further indicate broader chemotype coverage without sacrificing validity. Representative molecules and binding mode depictions are provided in the supplementary figures.

VI. Conclusion

We introduced FragGPT-GA, a tightly coupled hybrid framework for de novo molecular design that fuses fragment-based GPT generation with GA optimization under multi-objective selection. The method sustains diversity, avoids premature convergence, and drives populations toward chemically plausible, high-affinity candidates. Future work will expand objective sets (e.g., ADMET proxies), investigate task-adaptive prompting for the GPT component, and explore transfer to additional protein families.

Appendix
Proof of the Zonklar Equations

Use `\appendix` if you have a single appendix.

Acknowledgment

The authors would like to thank...

References

[1] Mathematics Into Type. American Mathematical Society. [Online]. Available: https://www.ams.org/arc/styleguide/mit-2.pdf

[2] H. Sira-Ramirez, "On the sliding mode control of nonlinear systems." Syst. Control Lett., vol. 19, pp. 303-312, 1992.