Enhancing Molecular Generation with FragGPT-Guided Genetic Algorithms

Tiangai Yao, Zhangfan Yang, Junkai Ji, *Member, IEEE*

***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

De novo molecular design has emerged as a cornerstone of computational drug discovery, with various methodologies addressing the challenge of navigating vast chemical spaces to identify molecules with desired pharmacological properties. This section systematically reviews the evolutionary landscape of molecular optimization approaches, categorizing them by their core computational paradigms and analyzing their respective strengths and limitations.

A. Evolutionary Algorithms in Molecular Optimization

Evolutionary algorithms (EAs) have established themselves as robust optimization frameworks for molecular design, leveraging population-based search strategies to explore chemical space systematically. Traditional genetic algorithms encode molecules using string representations such as SMILES or graph-based structures, evolving populations through selection, crossover, and mutation operators guided by fitness functions derived from docking scores, bioactivity predictions, or multi-objective criteria.

Classical Genetic Approaches: Early implementations like Graph-GA [1] operate directly on molecular graph representations, employing specialized crossover and mutation operators designed for graph structures. These methods demonstrate computational efficiency through parameter-free designs but suffer from limited exploration diversity. Similarly, traditional GA frameworks often implement chemically-aware heuristics such as BRICS-guided edits and reaction-based mutations to maintain chemical validity and synthesizability.

T. Yao, Z. Yang, and J. Ji are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: 2400671005@mails.szu.edu.cn; yangzhangfan@szu.edu.cn; jijunkai@szu.edu.cn).

Manuscript received December 15, 2024; revised January 20, 2025.

1089-778X/25$31.00 © 2025 IEEE

Enhanced Evolutionary Strategies: More sophisticated approaches integrate neural network guidance to address exploration limitations. The GA+D framework [2] augments genetic optimization with discriminator networks, employing SELFIES molecular representation to guarantee chemical validity while using neural networks to guide the evolutionary process. However, these discriminator-based approaches often require extensive hyperparameter tuning and may introduce computational overhead that limits scalability.

Limitations of Pure Evolutionary Approaches: Despite their optimization prowess, vanilla genetic algorithms suffer from fundamental limitations: (1) *diversity collapse* - populations tend to converge prematurely around local optima, reducing scaffold novelty; (2) *mode-seeking behavior* - the selection pressure often drives populations toward similar chemical motifs, limiting exploration of diverse chemotypes; (3) *limited generative capacity* - traditional crossover and mutation operators struggle to introduce genuinely novel structural patterns beyond local modifications of existing molecules.

B. Deep Generative Models for Molecular Design

The advent of deep learning has revolutionized molecular generation, with various generative modeling paradigms demonstrating remarkable capabilities in learning chemical distributions and sampling novel molecular structures.

Autoregressive and Variational Models: Junction Tree Variational Autoencoders (JTVAE) [3] decompose molecules into chemically meaningful substructures, enabling structured generation while maintaining chemical validity. These approaches excel at capturing local chemical patterns but often struggle with long-range dependencies and complex multi-ring systems.

Reinforcement Learning Approaches: REINVENT [4] employs recurrent neural networks for SMILES-based generation through reinforcement learning paradigms, demonstrating the ability to optimize molecules toward specific objectives. However, these methods face challenges in balancing exploration and exploitation, often requiring careful reward engineering and suffering from training instability. More sophisticated approaches like MolDQN [5] formulate molecular design as sequential decision-making processes using Markov Decision Process principles, but they struggle with sample efficiency and may converge to suboptimal solutions.

Multi-Objective Generation: Advanced frameworks address multi-objective optimization challenges. RationaleRL [6] implements policy gradient optimization through message-passing neural networks for molecular graph generation, incorporating interpretable substructures. MARS [7] executes Markov Chain Monte Carlo sampling with temperature annealing schemes for multi-objective molecular discovery, utilizing graph neural networks for proposal generation.

3D-Aware Generation: Recent developments include 3D structure-aware approaches like Gen3D [8], which generates molecules considering three-dimensional constraints for structure-based drug design. While promising, these methods often require extensive computational resources and may struggle with complex binding pocket geometries.

Limitations of Pure Generative Approaches: Despite impressive generative capabilities, deep learning models face significant challenges: (1) *objective steering difficulty* - directing generation toward multi-objective optima (e.g., simultaneous optimization of binding affinity, drug-likeness, and synthetic accessibility) remains non-trivial; (2) *sample inefficiency* - reinforcement learning-based approaches often require extensive sampling to achieve convergence; (3) *training instability* - policy gradient methods and adversarial training can be unstable, leading to mode collapse or poor convergence; (4) *limited exploitation* - purely generative approaches lack the systematic exploitation capabilities of evolutionary algorithms.

C. Hybrid and Multi-Strategy Approaches

Recognizing the complementary strengths of evolutionary and generative paradigms, recent research has explored hybrid methodologies that combine multiple optimization strategies.

Loose Coupling Strategies: Early hybrid approaches typically employ loose coupling mechanisms, such as initializing GA populations with generative models or periodically reseeding populations with neural proposals. While these methods demonstrate some improvement over individual approaches, they often underutilize model priors and may disrupt established GA dynamics.

Guided Exploration Methods: GEGL [9] represents a more sophisticated approach, guiding deep molecular optimization with genetic exploration. However, this method faces challenges in maintaining the balance between genetic algorithm rigour and neural network flexibility, often requiring extensive hyperparameter optimization.

Current Limitations of Hybrid Approaches: Existing hybrid methodologies suffer from several key limitations: (1) *integration depth* - most approaches implement shallow integration, failing to leverage the full potential of both paradigms; (2) *dynamic balance* - maintaining optimal balance between exploration (generative) and exploitation (evolutionary) components throughout optimization remains challenging; (3) *computational efficiency* - combining multiple complex algorithms often results in significant computational overhead; (4) *objective coordination* - ensuring that both generative and evolutionary components work toward consistent multi-objective goals is non-trivial.

D. Fragment-Based Molecular Design

Fragment-based approaches have gained prominence in both traditional medicinal chemistry and computational drug design, leveraging the principle that complex molecules can be decomposed into smaller, chemically meaningful building blocks.

Traditional Fragment Approaches: Classical fragment-based drug design relies on identifying small molecular fragments that bind to target proteins, subsequently linking or growing these fragments into larger, more potent compounds. This approach has proven successful in medicinal chemistry but often requires extensive manual intervention and domain expertise.

Computational Fragment Methods: Recent computational approaches have attempted to automate fragment-based design,

using techniques such as BRICS decomposition to break molecules into chemically sensible fragments. However, these methods often struggle with fragment recombination and may generate chemically implausible structures.

Our FragGPT-GA framework addresses the limitations of existing approaches through several key innovations: (1) *tight integration* - we implement deep integration between fragment-based GPT generation and genetic algorithm optimization, where the GPT operates as an intelligent diversity operator within the GA loop; (2) *dynamic masking* - our approach employs dynamic fragment masking that adapts exploration intensity throughout the optimization process; (3) *multi-objective coordination* - both generative and evolutionary components are coordinated toward consistent multi-objective goals through NSGA-II selection; (4) *chemical validity preservation* - the fragment-based GPT approach ensures high chemical validity while maintaining synthetic accessibility.

III. The Proposed FragGPT-GA Framework

A. Framework Overview and Design Philosophy

The FragGPT-GA framework represents a paradigm shift in molecular optimization by establishing a tight integration between evolutionary algorithms and generative AI models. Unlike existing hybrid approaches that employ loose coupling strategies, our framework embeds a fragment-based GPT model as an intelligent diversity operator directly within the genetic algorithm's iterative loop. This deep integration enables the system to maintain the systematic optimization capabilities of evolutionary algorithms while leveraging the generative power and learned chemical knowledge of large language models.

The core innovation lies in treating the GPT model not as an external generative component, but as an adaptive operator that intelligently expands the search space based on the current population's genetic material. This approach addresses the fundamental limitation of traditional GAs—diversity collapse—while maintaining the directed optimization pressure essential for convergence to high-quality solutions.

B. Architectural Components and Workflow

The FragGPT-GA framework operates through a sophisticated multi-stage iterative process, as depicted in Fig. 1. Each generation cycle encompasses five interconnected phases: (1) *fragment-based molecular decomposition*, (2) *intelligent GPT-driven diversity generation*, (3) *population-aware genetic operations*, (4) *multi-objective fitness evaluation*, and (5) *Pareto-optimal selection*. This design ensures that exploration and exploitation are balanced dynamically throughout the optimization process.

C. Fragment-Based Molecular Decomposition and Dynamic Masking

The first critical component of our framework involves sophisticated molecular fragmentation using BRICS (Breaking of Retrosynthetically Interesting Chemical Substructures) decomposition. For each molecule m_i in the current parent

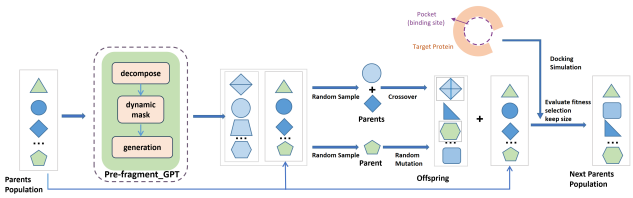

Fig. 1. The iterative workflow of the proposed FragGPT-GA framework. The process synergizes a Genetic Algorithm (GA) for optimization with a Generative Pre-trained Transformer (GPT) for diversity expansion, featuring dynamic fragment masking and multi-objective selection.

population P_g, we apply fragmentation to obtain a set of chemically meaningful fragments $F_i = \{f_1, f_2, \ldots, f_k\}$. This decomposition preserves chemical validity by breaking bonds at positions that correspond to common synthetic transformations.

Our dynamic masking strategy represents a significant advancement over static approaches. The number of fragments to mask adapts throughout the optimization process according to:

$$n_{mask}(g) = n_{initial} + \frac{g-1}{G_{max}-1} \cdot (n_{final} - n_{initial}) \quad (1)$$

where g represents the current generation, G_{max} is the maximum number of generations, $n_{initial}$ and n_{final} are the initial and final masking intensities, respectively. This adaptive strategy enables broad exploration in early generations (higher masking) while promoting targeted refinement in later stages (lower masking), thus maintaining an optimal exploration-exploitation balance throughout the evolutionary process.

D. GPT-Powered Intelligent Diversity Generation

The fragment-based GPT model serves as the core innovation of our framework, acting as an intelligent diversity operator that generates novel molecular candidates from masked fragment contexts. Unlike conventional genetic operators that perform local modifications, the GPT component can introduce entirely new structural motifs and scaffold families, effectively expanding the accessible chemical space.

GPT's Unique Advantages for Molecular Optimization:

1) **Population Distribution Expansion:** The GPT model possesses the remarkable ability to generate molecules that lie outside the current population's distribution. While traditional GA operators (crossover and mutation) are inherently constrained to explore regions proximate to existing population members, the GPT model can propose structurally novel compounds that introduce entirely new chemical scaffolds and functional groups into the population.
2) **Learned Chemical Knowledge Integration:** Having been pre-trained on extensive chemical databases, the GPT model encodes implicit knowledge about structure-activity relationships, pharmacophore patterns, and synthetic accessibility. This enables it to generate molecules that are not only novel but also chemically reasonable and potentially bioactive.

3) **Accelerated Optimization Convergence:** By introducing high-quality, diverse candidates at each generation, the GPT component accelerates the optimization process. Rather than relying solely on gradual evolutionary improvements, the algorithm can make significant leaps in chemical space toward promising regions, reducing the number of generations required for convergence.
4) **Scaffold Hopping Capability:** The GPT model excels at scaffold hopping—the ability to identify structurally distinct compounds with similar biological activities. This capability is particularly valuable in drug design, where exploring alternative scaffolds can lead to improved selectivity, reduced toxicity, or enhanced pharmacokinetic properties.

The GPT generation process proceeds as follows: for each masked fragment sequence M_i, the model autoregressively generates completion candidates using temperature-controlled sampling. The generation temperature T and top-k sampling parameters are carefully tuned to balance novelty and chemical validity.

E. Population-Aware Genetic Operations

Following GPT-based diversity generation, the framework performs sophisticated genetic operations on an expanded population pool $P_{expanded} = P_g \cup P_{GPT}$, where P_{GPT} represents the set of GPT-generated molecules. This expanded pool serves as the input for both crossover and mutation operations, enabling the genetic operators to work with a more diverse and potentially higher-quality substrate.

Enhanced Crossover Operations: Our crossover mechanism employs a fragment-aware approach that preserves chemically meaningful substructures while enabling recombination of beneficial motifs from different molecules. The process operates at the fragment level, ensuring that the resulting offspring maintain chemical validity and synthetic feasibility.

Chemical-Context Mutation: The mutation operator incorporates chemical knowledge to perform context-appropriate modifications. Rather than random SMILES string manipulations, mutations are guided by chemical rules and synthetic feasibility constraints, increasing the likelihood of generating viable compounds.

F. Multi-Objective Fitness Evaluation and Optimization

The framework employs a comprehensive multi-objective fitness function that simultaneously optimizes three critical pharmaceutical properties:

$$F(m) = [DS(m), QED(m), SA(m)] \tag{2}$$

where $DS(m)$ represents the docking score (to be minimized), $QED(m)$ quantifies drug-likeness (to be maximized), and $SA(m)$ measures synthetic accessibility (to be minimized). This multi-objective formulation ensures that optimization does not sacrifice drug-like properties or synthetic feasibility in pursuit of binding affinity alone.

The docking evaluation employs AutoDock Vina with standardized protocols, utilizing grid-based molecular docking to assess binding affinity against target proteins. The QED (Quantitative Estimate of Drug-likeness) incorporates multiple pharmaceutical properties including molecular weight, lipophilicity, and topological polar surface area. The SA score estimates synthetic accessibility based on fragment contributions and structural complexity.

Algorithm 1 FragGPT-GA Complete Framework

Input: Initial population P_0, max generations G_{max}, GPT model θ
Output: Pareto-optimal population P^*
Initialize population P_0 and evaluate $F(P_0)$
for $g = 1$ to G_{max} **do**
  // Fragment-based decomposition and masking
  $n_{mask} \leftarrow$ CalculateDynamicMask(g, G_{max})
  $M_g \leftarrow$ DecomposeAndMask(P_{g-1}, n_{mask})
  // GPT-powered diversity generation
  $P_{GPT} \leftarrow$ GPTGenerate(M_g, θ)
  // Population-aware genetic operations
  $P_{pool} \leftarrow P_{g-1} \cup P_{GPT}$
  $C_{cross} \leftarrow$ FragmentAwareCrossover(P_{pool})
  $C_{mut} \leftarrow$ ChemicalContextMutation(P_{pool})
  $C_g \leftarrow$ ChemicalFilter($C_{cross} \cup C_{mut}$)
  // Multi-objective evaluation
  Evaluate $F(C_g)$ = $[DS(C_g), QED(C_g), SA(C_g)]$
  // Pareto-optimal selection
  $P_g \leftarrow$ NSGA-II-Select($P_{g-1} \cup C_g$)
end for
return Non-dominated solutions from $P_{G_{max}}$

G. Pareto-Optimal Selection via NSGA-II

Selection operates under the NSGA-II (Non-dominated Sorting Genetic Algorithm II) framework, which maintains population diversity while driving convergence toward the Pareto-optimal front. The algorithm performs non-dominated sorting to identify solution fronts, followed by crowding distance calculation to preserve diversity within each front.

The selection process ensures that the framework maintains a balanced portfolio of solutions representing different trade-offs between the three objectives. This approach prevents the optimization from converging to a single-objective optimum and provides medicinal chemists with diverse options for further development.

H. Algorithmic Implementation and Computational Efficiency

The complete FragGPT-GA algorithm is presented in Algorithm 1. The implementation features several efficiency optimizations: (1) *parallel evaluation* of molecular properties using multiprocessing, (2) *incremental learning* where the GPT model can optionally be fine-tuned on successful molecules from previous generations, and (3) *adaptive resource allocation* that dynamically adjusts computational resources based on population convergence metrics.

The modular architecture facilitates reproducibility and enables systematic ablation studies. Each component—fragment

decomposition, GPT generation, genetic operations, evaluation, and selection—operates as an independent module with well-defined interfaces, allowing for easy customization and extension of the framework.

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. 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.

D. 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.

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 docking score comparison across different methods. We compare against established baselines including MARS [7], MolDQN [5], GEGL [9], REINVENT [4], RationaleRL [6], JTVAE [3], Gen3D [8], GA+D [2], Graph-GA [1], AutoGrow 4.0 [10], RGA [11], and screening baseline [12]. As shown, FragGPT-GA achieves the best docking score with a TOP-1 value of -13.458 ± 0.442 kcal/mol, demonstrating superior binding affinity compared to all baseline methods including RGA (-12.869 ± 0.473 kcal/mol).

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

Method	Nov↑	Div↑	QED↑	SA↓
screening	0%	0.858±0.005	0.678±0.022	2.689±0.077
MARS	100%	**0.877±0.001**	0.709±0.008	2.450±0.034
MolDQN	100%	**0.877±0.009**	0.170±0.024	5.833±0.182
GEGL	100%	0.853±0.003	0.643±0.014	2.990±0.054
REINVENT	100%	0.857±0.011	0.445±0.058	2.596±0.116
RationaleRL	100%	0.717±0.025	0.315±0.023	2.919±0.126
JTVAE	98%	0.867±0.001	0.593±0.035	3.222±0.136
Gen3D	100%	0.870±0.006	0.701±0.016	3.450±0.120
GA+D	99%	0.834±0.035	0.405±0.024	5.024±0.164
Graph-GA	100%	0.811±0.037	0.456±0.067	3.503±0.367
Autogrow 4.0	100%	0.852±0.011	0.748±0.022	2.497±0.049
RGA	100%	0.857±0.020	0.742±0.036	2.473±0.048
FragGPT-GA	100%	0.845±0.024	**0.764±0.012**	**2.014±0.153**

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

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

$$S_{\text{single}}(m) = -\text{DockingScore}(m) \tag{3}$$

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

$$S_{\text{multi-obj}}(m) = \begin{bmatrix} -\text{DockingScore}(m) \\ \text{QED}(m) \\ -\text{SA}(m) \end{bmatrix} \tag{4}$$

TABLE III
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

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

$$S_{\text{comp}}(m) = \widehat{DS}(m) \times \text{QED}(m) \times \widehat{SA}(m) \in [0, 1] \quad (5)$$

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

$$\widehat{DS}(m) = -\frac{\text{clip}(\text{DockingScore}(m))}{20} \in [0, 1] \quad (6)$$

$$\widehat{SA}(m) = \frac{10 - \text{SA}(m)}{9} \in [0, 1] \quad (7)$$

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 III 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 ($S(\text{TOP-1}) = -13.314$, $S(\text{QED}) = 0.579$, $S(\text{SA}) = 2.645$).

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.

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

C. Comprehensive Baseline Comparison

To further validate the effectiveness of FragGPT-GA, we conducted extensive experiments comparing our method against AutoGrow4.0 and RGA across 10 diverse protein targets (1iep, 2rgp, 3eml, 3ny8, 3pbl, 4r6e, 4rlu, 4unn, 5mo4, 7l11). Figure 3 presents a comprehensive violin plot visualization showing the docking score distributions for all three methods on each individual protein target.

The protein-specific analysis reveals that FragGPT-GA achieves superior or competitive performance across all 10 protein targets. Overall, FragGPT-GA demonstrates the most favorable docking scores with a mean of -10.31 kcal/mol, compared to AutoGrow4.0 (-10.27 kcal/mol) and RGA (-7.96 kcal/mol). Notably, FragGPT-GA exhibits remarkable consistency across different protein families, successfully identifying molecules with exceptional binding affinities (minimum -14.9 kcal/mol) while maintaining population diversity. The violin plots reveal that FragGPT-GA populations span broader score ranges, indicating effective exploration, while AutoGrow4.0 shows more conservative exploration patterns with tighter distributions. RGA, despite computational efficiency, consistently produces less favorable docking scores across all targets, confirming the advantage of our hybrid GPT-GA approach for molecular optimization.

D. 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.

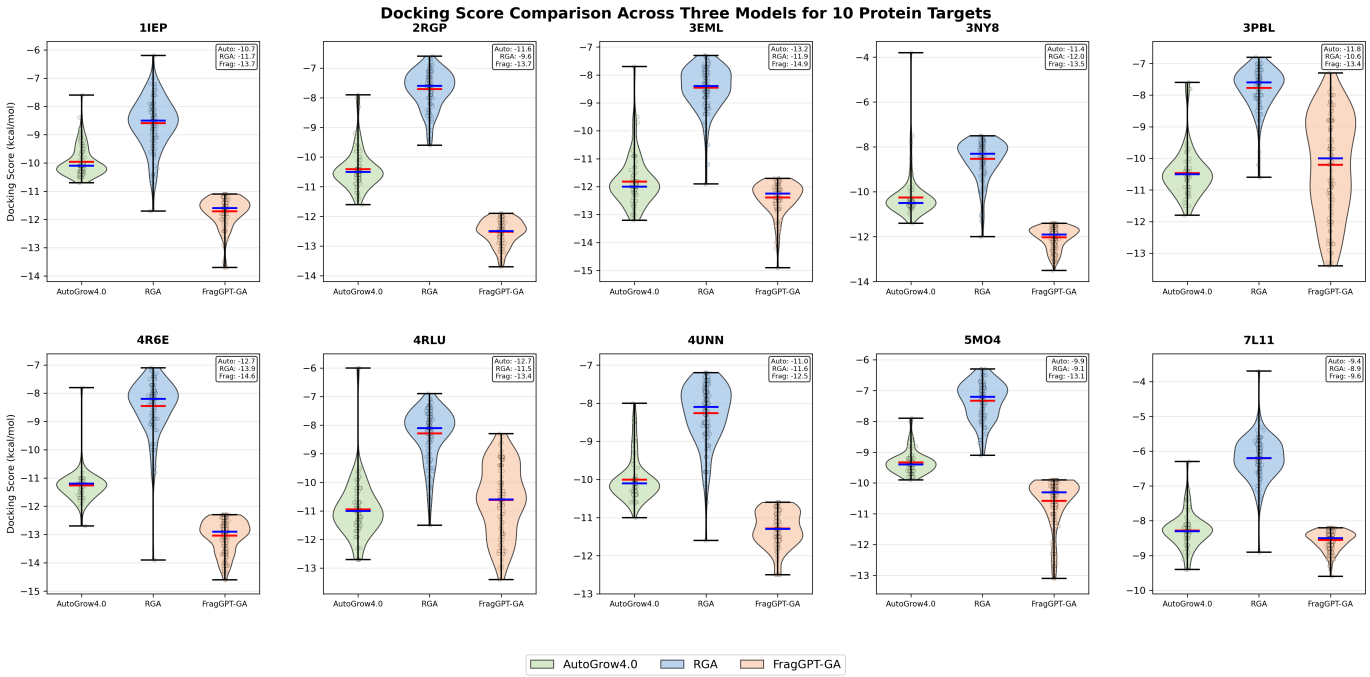

Fig. 3. Protein-specific docking score comparison across three models. Each subplot represents one protein target, with three violin plots showing the distribution of docking scores for AutoGrow4.0 (green), RGA (pink), and FragGPT-GA (blue). The violin plots display both population density and individual data points with statistical overlays (red lines: means, blue lines: medians). The top-right corner of each subplot shows the best docking score (TOP1) achieved by each model for that specific protein target. FragGPT-GA consistently demonstrates superior performance across most protein targets, achieving the best TOP1 scores while maintaining population diversity.

Appendix
Proof of the Zonklar Equations

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

Acknowledgment

The authors would like to thank...

References

[1] J. H. Jensen, "A graph-based genetic algorithm and generative model/monte carlo tree search for the exploration of chemical space," vol. 10, no. 12, pp. 3567–3572.

[2] A. Nigam, P. Friederich, M. Krenn, and A. Aspuru-Guzik, "Augmenting genetic algorithms with deep neural networks for exploring the chemical space."

[3] W. Jin, R. Barzilay, and T. Jaakkola, "Junction tree variational autoencoder for molecular graph generation," pp. 2323–2332.

[4] M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen, "Molecular de-novo design through deep reinforcement learning," vol. 9, no. 1, p. 48.

[5] Z. Zhou, S. Kearnes, L. Li, R. N. Zare, and P. Riley, "Optimization of molecules via deep reinforcement learning," vol. 9, no. 1, p. 10752.

[6] W. Jin, R. Barzilay, and T. Jaakkola, "Multi-objective molecule generation using interpretable substructures," pp. 4849–4859.

[7] Y. Xie, C. Shi, H. Zhou, Y. Yang, W. Zhang, Y. Yu, and L. Li, "Mars: Markov molecular sampling for multi-objective drug discovery."

[8] S. Luo, J. Guan, J. Ma, and J. Peng, "A 3d generative model for structure-based drug design," vol. 34, pp. 6229–6239.

[9] S. Ahn, J. Kim, H. Lee, and J. Shin, "Guiding deep molecular optimization with genetic exploration," vol. 33, pp. 12 008–12 021.

[10] J. O. Spiegel and J. D. Durrant, "AutoGrow4: An open-source genetic algorithm for de novo drug design and lead optimization," vol. 12, no. 1, p. 25. [Online]. Available: https://jcheminf.biomedcentral.com/articles/10.1186/s13321-020-00429-4

[11] T. Fu, W. Gao, C. W. Coley, and J. Sun, "Reinforced Genetic Algorithm for Structure-based Drug Design," pp. 6759–6778.

[12] T. Sterling and J. J. Irwin, "Zinc 15 – ligand discovery for everyone," vol. 55, no. 11, pp. 2324–2337.

[13] S. Lee, K. Kreis, S. P. Veccham, M. Liu, D. Reidenbach, S. Paliwal, A. Vahdat, and W. Nie, "Molecule Generation with Fragment Retrieval Augmentation."

[14] M. Abdelsayed, "AI-Driven Polypharmacology in Small-Molecule Drug Discovery," vol. 26, no. 14, p. 6996. [Online]. Available: https://www.mdpi.com/1422-0067/26/14/6996

[15] C. Ai, H. Yang, X. Liu, R. Dong, Y. Ding, and F. Guo, "MTMol-GPT: De novo multi-target molecular generation with transformer-based generative adversarial imitation learning," vol. 20, no. 6, p. e1012229. [Online]. Available: https://dx.plos.org/10.1371/journal.pcbi.1012229

[16] A. Cerveira, F. Kremer, D. d. A. Lourenço, and U. B. Corrêa. Evaluation Framework for AI-driven Molecular Design of Multi-target Drugs: Brain Diseases as a Case Study. [Online]. Available: http://arxiv.org/abs/2408.10482

[17] V. A. Cicirello, "Evolutionary Computation: Theories, Techniques, and Applications," vol. 14, no. 6, p. 2542. [Online]. Available: https://www.mdpi.com/2076-3417/14/6/2542

[18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," vol. 6, no. 2, pp. 182–197. [Online]. Available: http://ieeexplore.ieee.org/document/996017/

[19] Z. Fang, X. Wang, Y. Sun, and M. A. Adibhashimi, "Accelerating multi-objective optimization of concrete thin shell structures using graph-constrained GANs and NSGA-II," vol. 15, no. 1, p. 16090. [Online]. Available: https://www.nature.com/articles/s41598-025-00017-2

[20] B. Huang, R. Cheng, Z. Li, Y. Jin, and K. C. Tan. EvoX: A Distributed GPU-accelerated Framework for Scalable Evolutionary Computation. [Online]. Available: http://arxiv.org/abs/2301.12457

[21] A. Kadan, K. Ryczko, E. Lloyd, A. Roitberg, and T. Yamazaki, "Guided multi-objective generative AI to enhance structure-based drug design," vol. 16, no. 29, pp. 13 196–13 210. [Online]. Available: https://xlink.rsc.org/?DOI=D5SC01778E

[22] T. Khater, S. A. Alkhatib, A. AlShehhi, C. Pitsalidis, A. M. Pappa, S. T. Ngo, V. Chan, and V. K. Truong, "Generative artificial intelligence based models optimization towards molecule design enhancement," vol. 17, no. 1, p. 116. [Online]. Available: https://jcheminf.biomedcentral.com/articles/10.1186/s13321-025-01059-4

[23] J. Meyers and N. Brown. FragGT: Fragment-based evolutionary molecule generation with gene types. [Online]. Available: https://chemrxiv.org/engage/chemrxiv/article-details/65cf28f266c1381729c70206

[24] Qingfu Zhang and Hui Li, "MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition," vol. 11, no. 6, pp. 712–731. [Online]. Available: http://ieeexplore.ieee.org/document/4358754/

[25] Y. Tian, T. Zhang, J. Xiao, X. Zhang, and Y. Jin, "A Coevolutionary Framework for Constrained Multiobjective Optimization Problems," vol. 25, no. 1, pp. 102–116. [Online]. Available: https://ieeexplore.ieee.org/document/9122020/

[26] T. V. T. Tran and T. S. Hy, "Protein Design by Directed Evolution Guided by Large Language Models."

[27] Z. Wang, K. Zhang, Z. Zhao, Y. Wen, A. Pandey, H. Liu, and K. Ding. A Survey of Large Language Models for Text-Guided Molecular Discovery: From Molecule Generation to Optimization. [Online]. Available: http://arxiv.org/abs/2505.16094

[28] C. Wang, J. Zhao, L. Jiao, L. Li, F. Liu, and S. Yang. When Large Language Models Meet Evolutionary Algorithms: Potential Enhancements and Challenges. [Online]. Available: http://arxiv.org/abs/2401.10510

[29] X. Wu, S.-h. Wu, J. Wu, L. Feng, and K. C. Tan. Evolutionary Computation in the Era of Large Language Model: Survey and Roadmap. [Online]. Available: http://arxiv.org/abs/2401.10034

[30] X. Xia, Y. Zhang, X. Zeng, X. Zhang, C. Zheng, and Y. Su, "CMOMO: A deep multi-objective optimization framework for constrained molecular multi-property optimization," vol. 26, no. 4, p. bbaf335. [Online]. Available: https://academic.oup.com/bib/article/doi/10.1093/bib/bbaf335/8195924

[31] X. Xia, Y. Liu, C. Zheng, X. Zhang, Q. Wu, X. Gao, X. Zeng, and Y. Su, "Evolutionary Multiobjective Molecule Optimization in an Implicit Chemical Space," vol. 64, no. 13, pp. 5161–5174. [Online]. Available: https://pubs.acs.org/doi/10.1021/acs.jcim.4c00031

[32] J. Yoo, W. Jang, and W.-H. Shin, "From part to whole: AI-driven progress in fragment-based drug discovery," vol. 91, p. 102995. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0959440X25000132

[33] T. Zhang, X. Guo, Y. Li, N. Li, R. Zheng, W. Dong, and W. Ding, "A Dual-Stage and Dual-Population Algorithm Based on Chemical Reaction Optimization for Constrained Multi-Objective Optimization," vol. 13, no. 8, p. 2484. [Online]. Available: https://www.mdpi.com/2227-9717/13/8/2484