A Ground-Up Designed Controllable GPT for Molecular Optimization

Xuefeng Liu [1] **Songhao Jiang** [1] **Bo Li** [1] **Rick L. Stevens** [1 2]

Abstract

Large Language Models (LLMs) employ three popular training approaches: Masked Language Models (MLM), Causal Language Models (CLM), and Sequence-to-Sequence Models (seq2seq). However, each approach has its strengths and limitations, and faces challenges in addressing specific tasks that require controllable and bidirectional generation, such as drug optimization. To address this challenge, inspired by the biological processes of growth and evolution, which involve the expansion, shrinking, and mutation of sequences, we introduce CONTROLLABLEGPT. This initiative represents the first effort to combine the advantages of MLM, CLM, and seq2seq into a single unified, controllable GPT framework. It enables the precise management of specific locations and ranges within a sequence, allowing for expansion, reduction, or mutation over chosen or random lengths, while maintaining the integrity of any specified positions or subsequences. In this work, we designed CONTROLLABLEGPT for drug optimization from the ground up, which included proposing the Causally Masked Seq2seq (CMS) objective, developing the training corpus, introducing a novel pre-training approach, and devising a unique generation process. We demonstrate the effectiveness and controllability of CONTROLLABLEGPT by conducting experiments on drug optimization tasks for both viral and cancer benchmarks, surpassing competing baselines.

1. Introduction

The Generative Pre-trained Transformer (GPT) (Floridi and Chiriatti, 2020; Yenduri et al., 2023) has achieved significant success in applications such as ChatGPT (Ouyang et al., 2022; Wu et al., 2023) for chatbox, Copilot (Barke et al., 2023) for code generation and VideoGPT (Yan et al., 2021) for video creation. However, GPT operates unidirectionally and lacks controllability (Ethayarajh, 2019), which means it still faces the challenge in handling the real-world scenarios that require both controllable and bidirectional generation capabilities, as dictated by the current design of GPT.

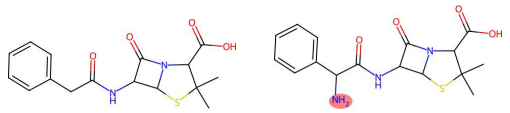
On the other hand, drug discovery (Berdigaliyev and Aljofan, 2020) has become increasingly important since the advent of COVID-19 (Muratov et al., 2021). The search for more effective drugs is becoming more urgent but remains underexplored. De Novo Drug Discovery incurs billions of dollars in costs and still confronts a high failure rate in its early stages (Tong et al., 2021). Drug Improvement addresses the limitations of De Novo drug discovery by building upon existing FDA-approved drugs. The DrugImprover framework (Liu et al., 2023a) leads the way in tackling drug optimization by using Tanimoto similarity (Landrum et al., 2013) to maintain beneficial properties while targeting multiple objectives with its novel Advantage-alignment Policy Optimization (APO) algorithm. It also provides a specialized dataset for optimizing drugs against COVID and cancer proteins. REINVENT 4 (He et al., 2021; 2022; Loeffler et al., 2024), in further, utilize the advanced generative capabilities of Transformers and large language models (LLMs) to address the drug optimization problem.

Although REINVENT 4 has further improved performance, demonstrated promising outcomes, and achieved state-of-the-art performance, its effectiveness is still limited due to the black-box nature of the LLMs' generation process. In this process, LLMs optimize for either maximum likelihood or specific objectives set by users during the decoding phase. More specifically, it cannot specify the generation specific locations or the length of tokens to be generated. Such limitations are further intensified in drug optimization with REINVENT 4: (1) If an important substructure exists within the original molecule, REINVENT 4 might miss it and fail to preserve it in the generated ones, even though retaining that substructure could be beneficial. (2) Real-world examples of drug improvement, such as the addition of an NH2 group to original drugs, demonstrate significant enhancements in effectiveness and reduced side effects. For instance, Ampicillin's modifications over Penicillin, illustrated in Figure 1, show increased activity range, stomach acid resis-

[1]Department of Computer Science, University of Chicago, Chicago, IL, USA [2]Argonne National Laboratory, Lemont, IL, USA. Correspondence to: Xuefeng Liu <xuefeng@uchicago.edu>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

tance, and better absorption while retaining the essential beta-lactam ring crucial for antibiotic activity. Ampicillin's main distinction from Penicillin is its side chain, altered to penetrate gram-negative bacteria's outer membranes more effectively. (3) If the fragment has associated side effects, drugs derived from it might inherit these issues, contradicting the goal of optimization. (4) Based on the given drug, we need to decide to what extent (e.g., how many atoms to add) to preserve or alter the original structure.



(a) Penicillin (b) Ampicillin

Figure 1: Penicillin in drug optimization. With adding a simple functional group NH2 (in red), Ampicillin has resolved the rash side effect bring about by Penicillin.

To address the challenges of the GPT model and drug optimization, this work introduces CONTROLLABLEGPT, a Bidirectional Causally Masked Seq2seq GPT tailored for controllable generation in drug optimization. Drawing inspiration from biological processes like growth and evolution, the training of this model leverages the SMILES representation (Weininger, 1988) to enable the expansion, contraction, and mutation of molecule sequences at any point while maintaining essential structures.

This model, trained with a novel causally masked seq2seq objective, merges causal, masked, and seq2seq language modeling, enabling comprehensive generative modeling with bidirectional context, facilitating precise modifications without disrupting the molecule's overall structure. The model offers controllable capabilities including: 1) Generating new molecules while preserving original beneficial functional groups or scaffolds. 2) Removing atoms or groups causing side effects. 3) Precisely adding new atoms at specified scales and positions, with size hints that guide the generation process. 4) Controlling mutations, allowing expansion or contraction based on sub-sequences to achieve specific molecular configurations.

In summary, our contributions are:

• We propose a Causally Masked Seq2seq (CMS) objective that merges causal, masked, and seq2seq language modeling. CMS enables precise management of specific locations and ranges within a sequence, facilitating expansion, reduction, or mutation over chosen or random lengths, all while preserving the integrity of any specified positions or subsequences.

• We developed CONTROLLABLEGPT completely from the ground up, training it on the CMS objective. This process included designing the training corpus, introducing a novel pre-training strategy, and devising a unique generation process.

• Through extensive experiments and ablation studies on real-world viral and cancer-related benchmarks, we demonstrate the effectiveness and controllability of CONTROLLABLEGPT in outperforming the competing baselines in improving upon existing molecules and drugs across targeted objectives, resulting in superior drug candidates.

2. Related Work

Causally Masked Language Modeling. Causal Language Modeling (CLM) is an autoregressive method employed in models such as GPT-4 (Achiam et al., 2023), predicting the next token using only prior token information. While effective in applications like text generation (Li et al., 2024) and dialogue systems (Hosseini-Asl et al., 2020), CLM's unidirectional approach is a limitation. Masked Language Modeling (MLM), used in models like BERT (Devlin et al., 2018), predicts hidden tokens using bidirectional context. Although it processes only about 15% of tokens during training, limiting some uses, MLM is still broadly applied in biology for representation learning (Chithrananda et al., 2020; Lin et al., 2023). Causally Masked objective improve MLM by providing a type of hybrid of causal and masked language models by enabling full generative modeling while also providing bidirectional context when generating the masked spans. However, the existing State-of-the-Art (SOTA) Causally Masked models (Aghajanyan et al., 2022) are still limited in enabling the controllable mutation function including conditional expansion and contraction, and the current design of masked tokens leads to misleading interpretations between the masked token and its context. In this work, we address both limitations by redesigning the masked tokens and incorporating a sequence-to-sequence model.

Sequence-to-Sequence Modeling. Seq2Seq, or Sequence-to-Sequence models (Dey and Salem, 2017; Graves and Graves, 2012; Xue et al., 2020; Wang et al., 2020; Ni et al., 2021), employ an encoder-decoder structure where the encoder interprets the input sequence, and the decoder constructs the output sequence. This method is frequently utilized in tasks such as machine translation (Chen et al., 2018; Tiwari et al., 2020; Wang et al., 2022), summarization (Prasad et al., 2020; Shi et al., 2021), and question-answering (Tang et al., 2018; Wu et al., 2020). Due to their ability to manage complex tasks that require transforming input into output, Seq2Seq models are highly versatile and suitable for a broad spectrum of NLP applications. Nevertheless, Seq2seq models exhibit limitations in coherence, context understanding, handling variable-length inputs, training efficiency, and capturing bidirectional context when compared to CLMs and MLMs.

In this study, we introduce Causally Masked Seq2seq (CMS) modeling, conceptualizing the seq2seq model as a controllable conditional mutation component in biological sequences, and harnessing the strengths of seq2seq models, CLMs, and MLMs.

Controllable Generation. In the field of computer vision, the introduction of generative adversarial networks (Goodfellow et al., 2014) enhanced the quality of image generation. Subsequent research focused on methods to control the generative process and improve the estimation of generative distributions (Kingma, 2013; Chen et al., 2016; Arjovsky et al., 2017). In the realm of natural language processing, language models are often developed as conditional models tailored for specific text generation tasks (Brants et al., 2007; Sutskever, 2014; Rush, 2015). Typically, prompts created by models or those written by humans serve merely as a rough starting point for the generated text. This raises questions about how to achieve more explicit control over text generation. Recent advancements in transformer architecture (Vaswani, 2017; Radford et al., 2019) and diffusion models (Ho et al., 2020) have led to improved control in both text and image generation (Li et al., 2019; Keskar et al., 2019; Raffel et al., 2020; Li et al., 2022; Epstein et al., 2023; Zhang et al., 2023; Liang et al., 2024). However, these techniques are not specifically adapted for biological sequences, which may involve unique challenges in nature, such as expansion, reduction, or mutation at specific locations and ranges with desired properties. CONTROLLABLEGPT is specifically designed for biological sequences and addresses these challenges by introducing a novel CMS objective.

Large Language Models for Drug Optimization. Large language models have been employed in molecule generation, as evidenced by studies such as MolGPT (Bagal et al., 2021), C5T5 (Rothchild et al., 2021), and ChemGPT (Frey et al., 2023). More recently, ERP (Xuefeng et al., 2024) has utilized LLMs for drug discovery. In contrast, our work focuses on the drug optimization domain to improve upon existing drugs rather than designing from scratch. In the drug optimization domain, DrugImprover (Liu et al., 2023a) starts to effectively define the drug optimization problem by using reinforcement learning with a combination of multiple objectives. Moreover, it integrates Tanimoto similarity (Landrum et al.) as an additional term in the rewards function to ensure that the RL-fine-tuned model generates molecules similar to existing drugs. However, DrugImprover employs an LSTM as the generative model, which has limitations in scalability, capacity, and contextual understanding. Reinvent 4 (He et al., 2021; 2022; Loeffler et al., 2024) has made efforts in developing transformer-based generative models and achieving state-of-the-art performance in the Drug Optimization domain with an emphasis on pretraining with simply adopt REINFORCE (Williams, 1992) finetuning. Although pretraining aids in producing molecules that resemble those in the training dataset, it naturally limits the scope of exploration because of biases inherent in the training data. In addition, REINVENT 4 lacks controllability during generation, a crucial aspect for drug optimization. In this study, we tackle controllability issues and surpass the current state-of-the-art, REINVENT 4, in drug optimization benchmarks.

3. Preliminaries

LLM/CLM. Let $\mathbf{X} = [x_1, x_2, \cdots, x_n]$ be a sequence of tokens representing an input sentence (prompt), where each x_i is a token from a vocabulary $\mathcal{V}$. Let $\mathbf{Y} = [y_1, y_2, \cdots, y_T], y_i \in \mathcal{Y}$ be the output sequence of tokens with vocabulary $\mathcal{Y}$. $\mathcal{V}$ and $\mathcal{Y}$ are potentially different vocabularies. Note that $\mathbf{y}_{<t} = [y_1, \cdots, y_{t-1}], \mathbf{y}_T := Y$. T represents the length of sequence. Each training corpus begins with a start token [BOS], follows with a sequence of tokens $\mathbf{y}$ where each y_i belongs to $\mathcal{V}$, and concludes with a termination action [EOS]. Each molecule is depicted using a sequence of tokens $\mathbf{y}$ to assemble a SMILES string, applicable to both incomplete and complete molecular structures. Let us denote $\circ$ as string concatenation, and let $\mathcal{V}^*$ represent the Kleene closure of $\mathcal{V}$. The set of training corpus $\mathcal{C}$ is defined as: $\mathcal{C} := \{[\text{BOS}] \circ \mathbf{v} \circ [\text{EOS}] \mid \mathbf{v} \in \mathcal{V}^*\}$. The LLM generator policy π_θ, which is parameterized by a deep neural network (DNN) with learned weights θ, is defined as a product of probability distributions: $\pi_\theta(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{|\mathbf{y}|} \pi_\theta(y_t|\mathbf{x}, \mathbf{y}_{<t})$, where $\pi_\theta(y_t|\mathbf{x}, \mathbf{y}_{<t}) = P(y_t|\mathbf{y}_{<t}, X)$ is a distribution of next token y_t. The text generation decoding process is designed to select the most probable hypothesis from all possible candidates by addressing the following optimization problem: $\mathbf{y}^\star = \arg\max_{\mathbf{y} \in \mathcal{Y}_T} \log \pi_\theta(\mathbf{y}|\mathbf{x})$. CLM is a variant of language modeling where the objective can be formulated as $\max_\theta \sum_{i=1}^{n} \log P(x_i|\mathbf{X}_{<i}; \theta)$, where $P(x_i|\mathbf{X}_{<i}; \theta)$ is the conditional probability of observing token x_i given all the preceding tokens $\mathbf{X}_{<i}$.

MLM. In MLM, a subset (around 15%) of the tokens in $\mathbf{X}$ is randomly selected and replaced with a special token [MASK]. Let us denote this masked sequence as $\mathbf{M}$ and unmasked sequence as $\mathbf{S}$, $\mathbf{S} = \{x_i\}, x_i \in \mathbf{X}$ and $x_i \notin \mathbf{M}$. The objective of the MLM is to predict the original tokens of the masked positions based solely on the unmasked context $\mathbf{S}$, which can be represented as maximizing the likelihood: $\mathcal{L}_{MLM} = \prod_{i \in \mathbf{M}} P(x_i|\mathbf{S}; \theta)$, where $P(x_i|\mathbf{S}; \theta)$ represents the conditional probability of observing token x_i given the context provided by the unmasked tokens in $\mathbf{S}$. θ represents the parameters of the model. The parameters θ of the model are optimized to maximize the likelihood of the correct tokens at the masked positions. During training, the model learns to utilize the surrounding context to predict the

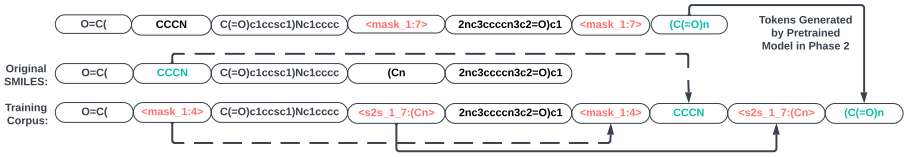


Figure 2: The visual representation of building the training corpus with both masked and seq2seq spans for seq2seq causal masked objective.

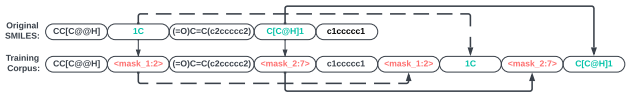


Figure 3: The visual representation of our causal masked objective on a molecule features two mask spans ($n = 2$), each with a specific size hint. The first span, <$mask_1 : 2$>, covers two tokens, and the second, <$mask_2 : 7$>, covers seven tokens.

masked tokens, which helps it develop a deep understanding of language structure and usage. MLM has proven effective for pre-training language models that are later fine-tuned for various downstream tasks.

Seq2Seq. Sequence-to-sequence (seq2seq) modeling is a framework in natural language processing designed to convert sequences from input sequence to output sequence. Seq2seq models typically consist of two main components: an encoder and a decoder, with model parameter θ_{enc} and θ_{dec} respectively. The encoder processes the input sequence $\mathbf{X}$ to a fixed-dimensional vector representation $\mathbf{c}$ to capture the semantic or contextual information. The decoder's objective is to generate the target sequence $\mathbf{Y}$ given the encoded representation $\mathbf{c}$. The objective in training seq2seq models is typically to maximize the log likelihood of the correct output sequence $\mathbf{Y}$ given the input sequence $\mathbf{X}$ across a dataset of paired sequences: $\max_{\theta_{enc},\theta_{dec}} \sum_{(\mathbf{X},\mathbf{Y})} \log P(\mathbf{Y}|\mathbf{X})$, where $\mathbf{P}$ is product of the conditional probabilities of each output token and $P(\mathbf{Y}|\mathbf{X}) = \prod_{j=1}^{n} P(y_j|\mathbf{Y}_{<j}, \mathbf{c}; \theta_{dec})$. Training involves adjusting both the encoder and decoder parameters to optimize this objective. Seq2seq models are powerful because they can handle variable-length input and output sequences and are capable of learning complex transformations between different types of sequence data.

4. CONTROLLABLEGPT

In this section, we propose CONTROLLABLEGPT, a ground-up designed GPT model for molecular optimization. We first introduce the novel Causally Masked Seq2seq (CMS) Objective as the foundation of CONTROLLABLEGPT. Then, we discuss the design of GPT, including designing the training corpus, a pre-training strategy, and a generation process.

4.1. Causally Masked Seq2seq (CMS) Objective.

Masked, causal, and seq2seq language modeling each offer unique benefits and limitations. Masked models encode bi-directional contexts but only decode about 15% of the tokens during training. Causal models, being decoder-only, process every token but are restricted to left-to-right contexts. Seq2seq models are versatile yet often lack bidirectional context and precise generation control. To combine the strengths of MLM, CLM, and seq2seq models and draw inspiration from biological molecule evolution using SMILES representation—which allows for molecular expansion, shrinking, and mutation—we introduce the Causally Masked Seq2seq (CMS) Objective. The CMS objective enables per-token generation, incorporating optional bidirectional and seq2seq functionality for greater adaptability. It allows for precise control over specific positions and spans within sequences, supporting the expansion, contraction, or mutation of segments while maintaining the integrity of designated areas. The construction of the CMS objective involves the following steps:

Designing the corpus. Our methodology for developing the CMS objective to a SMILES (Weininger, 1988) string of length $\mathcal{L}$ begins with the most basic corpus suitable for the CLM objective as $\mathcal{C} = \{[BOS], x_1, \cdots, x_T, [EOS]\}$.

Blending the MLM objective. We then build the MLM objective on top of CLM. It involves a probability p to determine the total number of tokens to mask as $\lfloor \mathcal{L} \cdot p \rfloor$. Let us denote $N \in \mathbb{R}^+$ as the number of span of mask in the source document.

$$[BOS], x_1, \cdots, \underbrace{x_{idx_1}, \cdots, x_{idx_1+\lfloor \mathcal{L}\cdot p \rfloor}}_{<mask_n:\mathcal{L}\cdot p>}, \cdots, x_T, [EOS], \tag{1}$$

For $N = 1$, let us choose a random starting index $idx_1 \sim [0, \mathcal{L} - \lfloor \mathcal{L} \cdot p \rfloor - 1]$, and proceed to mask tokens in range $[idx_1, idx_1 + \lfloor \mathcal{L} \cdot p \rfloor]$. For $N = 2$, we divide $\lfloor \mathcal{L} \cdot p \rfloor$ into two segments, $m1$ and $m2$, ensuring $m1 + m2 = \lfloor \mathcal{L} \cdot p \rfloor$ and that each segment's length is uniformly selected from the range $[1, \lfloor \mathcal{L} \cdot p \rfloor]$. We then identify a starting point idx_1 within $[0, \mathcal{L} - \lfloor \mathcal{L} \cdot p \rfloor - 1]$ for the first mask span and a second starting point idx_2 from the range $[idx_1 + m1 + 1, \mathcal{L} - m2 - 1]$ for the second mask span,

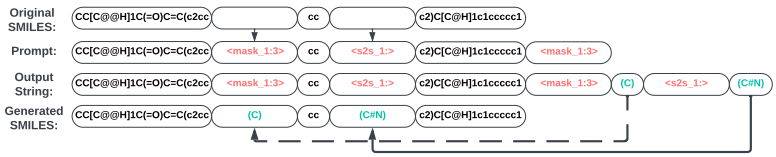


Figure 4: Expansion of an original molecule: Mask tokens (in red) are inserted into the SMILES string, prompting the generation of new segments (in green). These segments are then added to the molecule, showcasing the model's capability to expand molecular structures both creatively and precisely.

ensuring that the two masked segments are non-overlapping and sequentially ordered in the SMILES string. Following the same strategy for selection and masking of these segments, we could reach for any N. For the n_{th} span of mask, we replace the span by the token <$mask_n : \mathcal{L} \cdot p$>, where n and $\mathcal{L} \cdot p$ represents for the n_{th} masked segment with size hint length $\mathcal{L} \cdot p$, which specify the desired length of text to generate for replacing the mask conditioning on tokens length. Finally, we reposition the masked spans to the end of the SMILES string, maintaining their sequence order as illustrated in Fig. 3. In this work, we embed the size hint within the mask token as <$mask_i : n$> to avoid the ambiguity seen in prior works Aghajanyan et al. that use <$mask_i$>n. This format prevents misinterpretation by models, as numerical values in chemical structures can indicate ring closures or chain lengths.

Blending the seq2seq objective. Finally, we establish CMS objective by applying seq2seq objective on top of MLM and CLM. Initially, we train a GPT model using the MLM and CLM objectives, denoted as π_{CM}. Given a SMILES string, we randomly mask a seq2seq span starting at position s_1 and of length $\mathcal{L}$, ensuring it does not overlap with previously masked spans, while regard the remaining tokens as $\mathbf{Z}$. Our goal is to transform this s2s span $[x_{s_1}, \cdots, x_{s_1+\mathcal{L}}]$ into a target span with desired length $\mathcal{L}^t$. To create this training corpus, we utilize π_{CM} to generate $\mathcal{L}^t$ tokens $[m_1, \cdots, m_{\mathcal{L}^t}]$ with regarded to $\mathbf{Z}$. We then construct the training corpus by mapping the s2s span to the subsequence generated by π_{MLM}.

$$x_1, \cdots, x_{s_1-1}, \langle \text{mask_1} : \mathcal{L}^t \rangle, \tag{2}$$

$$\underbrace{x_{s_1+\mathcal{L}+1}, \cdots, x_T, \langle \text{mask_1} : \mathcal{L}^t \rangle}_{\text{Prompt based on the pretrained model in previous step}} \tag{3}$$

$$\rightarrow [m_1, \cdots, m_{\mathcal{L}^t}]$$

$$\underbrace{\begin{aligned} &x_1, \cdots, \langle \text{s2s_}i\text{_}\mathcal{L}^t : x_{s_1}, \cdots, x_{s_1+\mathcal{L}} \rangle, \\ &\cdots, x_T, \langle \text{s2s_}i\text{_}\mathcal{L}^t : x_{s_1}, \cdots, x_{s_1+\mathcal{L}} \rangle, [m_1, \cdots, m_{\mathcal{L}^t}] \end{aligned}}_{\text{Training corpus for seq2seq objective}} \tag{4}$$

where $\langle \text{s2s_}i\text{_}\mathcal{L}^t : x_{s_1}, \cdots, x_{s_1+\mathcal{L}} \rangle$ denotes the seq2seq objective conditioned on a specific subsequence $x_{s_1}, \cdots, x_{s_1+\mathcal{L}}$ and its bidirectional unmasked tokens. The index i indicates the i-th span, and $\mathcal{L}^t$ represents the target length of the generated subsequence. Unlike conventional sequence-to-sequence models, our work on seq2seq is also conditioned on and benefits from the bidirectional context surrounding the seq2seq span. This approach allows for the incorporation of task-specific length priors into prompts, resulting in outputs that are more precise and controlled.

4.2. The Design of the Controllable GPT.

Pretraining. In this work, we propose a novel three-phase training approach to train a GPT model under CMS objective.

In the initial phase. Our objective is to train a GPT specifically designed for understanding molecules. This training employs a CLM approach. CLM is an autoregressive technique where the model learns to predict the next token in a sequence based solely on the preceding tokens. This creates a unidirectional context model, which means it only considers past information and ignores any future context when making predictions. For this phase, the model is trained on a dataset comprised of texts about ligands. This dataset enables the model to accurately learn the representation of compounds, including their chemical structures and properties.

In the second phase. Building on the success of the LLM developed in Phase 1, which demonstrated high accuracy in generating molecular structures, we proceed to refine the model's training. This phase employs a causally masked objective with multiple mask tokens, each with a size hint, as illustrated in Figure 3. In this phase, the model, denoted as π_{CM}, benefits from both Causal Language Modeling (CLM) and Masked Language Modeling (MLM), which enhance CLM's performance by utilizing bidirectional context. π_{CM} is capable of generating molecules in a controlled manner, specifying both the target length and the position for expansion.

In the third phase. Ultimately, we achieve building the GPT under CMS objective by further refining the causally masked model,π_{CM}, through the integration of a sequence-

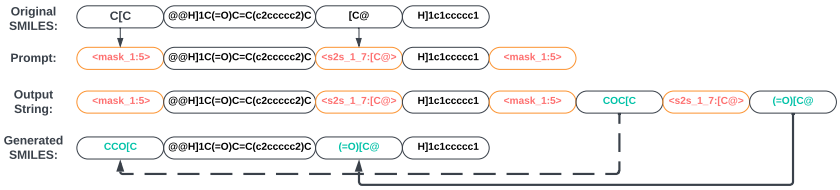


Figure 5: Modification of an original molecule. This figure illustrates the process of altering a molecule's structure. Key steps include replacing original segments with masked and sequence-to-sequence tokens (highlighted in red), generating new molecular segments (in green) by the model, and reintegrating these segments into the molecule.

to-sequence objective. We trained our model, denoted as π_{CMS}, using the training corpus outlined in Fig. 2 to refine the causally masked model π_{CM} developed in Phase 2. This advancement aims to enhance the model's controllable generation in terms of both contraction and mutation. It mimics the mutation behavior in biological sequences. Thus, our π_{CMS} achieves controllable generation in expansion, contraction, and mutation at specific positions or ranges, in either a random or specified length.

Loss function. Instead of altering the standard cross-entropy loss to consider the loss from predicting masked tokens negligible, we treat masked tokens like regular tokens, subject to the usual loss calculations. This method is used because our training data may contain multiple masked tokens, each with size hint information indicating the number of tokens to generate in place of the mask. Thus, it's crucial to accurately predict both the presence of these masked tokens and their corresponding size hints.

Generation Process. The prompt, output string, and generated SMILES for CONTROLLABLEGPT can be viewed in figure 5 and figure 4. More specifically, in the process of generating new molecular structures, CONTROLLABLEGPT employs a method that either modifies existing molecules or adds new elements to them without altering the original essential structure, showcasing the flexibility and precision of the model in generating novel molecular designs. This is illustrated through two examples:

Modifying the Original Molecule: Initially, two segments of the original molecule's SMILES string are identified and replaced with mask and seq2seq token respectively, which are placeholders indicating where and how long the new segments should be. These mask tokens are then processed by the model, which generates new segments in their place. The generated segments, highlighted in green, are repositioned to replace the original masked segments, effectively changing the molecule's structure and construct a new molecule. This process is depicted in Fig. 5, where the mask and seq2seq token are shown in red and the newly generated segments in green. The caption for Fig. 5 explains this process in detail, emphasizing the reintegration of generated tokens.

Adding to the Original Molecule Without Modification: In this scenario, instead of replacing parts of the SMILES string, one mask token and one seq2seq token are inserted at random positions within the string. These tokens serve as prompts for the model to generate new molecular segments that are then inserted into the specified positions, expanding the original molecule without altering its existing structure. This approach is visualized in Fig. 4, with the mask tokens again represented in green and the generated segments in red. The caption for Fig. 4 provides a clear explanation of this additive process.

5. Experiments

The language model. We employ the Byte Pair Encoding (BPE) method (Gage, 1994; Sennrich et al., 2015) to initially pre-train our tokenizer using raw SMILES strings, and GPT-2-like Transformers for causal language modeling. We use the standard 11M Drug-like Zinc dataset for training, excluding entries with empty scaffold SMILES. The dataset is divided into a 90/10 split for training and validation, respectively. (see Appendix A.1 for more details).

Dataset. We employ, from the most recent Cancer and COVID dataset of Liu et al. (2023a), 1 million compounds from the ZINC15 dataset docked to the 3CLPro (PDB ID: 7BQY) protein associated with SARS-CoV-2 and the RTCB (PDB ID: 4DWQ) human cancer protein.

Baselines. In this study, we use baseline models such as DrugImprover (Liu et al., 2023a), which leverages an LSTM-based generator fine-tuned with APO, Molsearch (Sun et al., 2022), a search-based strategy utilizing Monte Carlo Tree Search (MCTS) for molecule generation and optimization, MIMOSA (Fu et al., 2021), a graph-based molecular optimization method driven by sampling and DrugEx v3(Liu et al., 2023b), which utilizes transformer-based reinforcement learning for scaffold-driven drug optimization. Additionally, we incorporate the current state of art model, REINVENT 4, proposed by He et al. (2021; 2022); Loeffler et al. (2024), which trains a transformer to follow the Matched Molecular Pair (MMP) (Kenny and Sadowski, 2005; Tyrchan and Evertsson, 2017) guidelines. Specifically, given a set $\{(X, Y, Z)\}$, where X represents source molecule, Y the target molecule, and Z the property

Target	Algorithm	Avg Norm Reward ↑	Avg Top 10 % Norm Reward ↑	Docking ↓	Druglikeliness ↑	Synthesizability ↓	Solubility ↑	Similarity ↑
3CLPro	**Original**	0.532	0.689	-8.698	0.682	3.920	2.471	-
(PDBID:	**MMP** (Loeffler et al., 2024)	0.628 ± 0.001	0.718 ± 0.000	-8.259 ± 0.004	0.691 ± 0.001	2.682 ± 0.004	3.109 ± 0.020	0.862 ± 0.000
7BQY)	**Similarity (≥ 0.5)** (Loeffler et al., 2024)	0.615 ± 0.000	0.706 ± 0.001	-8.165 ± 0.024	0.697 ± 0.004	2.621 ± 0.006	3.180 ± 0.029	0.782 ± 0.001
	Similarity ([0.5, 0.7]) (Loeffler et al., 2024)	0.612 ± 0.001	0.701 ± 0.001	-8.187 ± 0.010	0.691 ± 0.001	2.611 ± 0.009	3.240 ± 0.014	0.756 ± 0.003
	Similarity (≥ 0.7) (Loeffler et al., 2024)	0.628 ± 0.001	0.718 ± 0.001	-8.214 ± 0.002	0.691 ± 0.002	2.717 ± 0.002	3.080 ± 0.016	0.881 ± 0.002
	Scaffold (Loeffler et al., 2024)	0.602 ± 0.001	0.703 ± 0.002	-8.116 ± 0.002	0.695 ± 0.001	2.728 ± 0.008	2.968 ± 0.038	0.776 ± 0.001
	Scaffold Generic (Loeffler et al., 2024)	0.617 ± 0.001	0.710 ± 0.002	-8.179 ± 0.012	0.701 ± 0.000	2.645 ± 0.008	3.090 ± 0.029	0.801 ± 0.000
	DrugImprover (Liu et al., 2023a)	0.432 ± 0.002	0.493 ± 0.005	-6.726 ± 0.007	0.506 ± 0.002	**1.306** ± 0.010	2.057 ± 0.011	0.531 ± 0.002
	Molsearch (Sun et al., 2022)	0.616 ± 0.001	0.726 ± 0.002	-8.855 ± 0.040	0.686 ± 0.001	3.105 ± 0.006	2.452 ± 0.008	**0.969** ± 0.001
	MIMOSA (Fu et al., 2021)	0.622 ± 0.001	0.734 ± 0.002	-8.800 ± 0.015	0.677 ± 0.004	3.105 ± 0.008	2.711 ± 0.010	0.959 ± 0.001
	DrugEx v3 (Liu et al., 2023b)	0.524 ± 0.001	0.613 ± 0.001	-8.089 ± 0.013	0.583 ± 0.002	3.095 ± 0.005	3.932 ± 0.008	0.495 ± 0.001
	CONTROLLABLEGPT (masks only)	0.668 ± 0.001	**0.743** ± 0.001	-9.083 ± 0.003	**0.718** ± 0.001	2.750 ± 0.001	3.630 ± 0.005	0.889 ± 0.001
	CONTROLLABLEGPT (mask + s2s)	**0.671** ± 0.001	**0.743** ± 0.001	**-9.150** ± 0.001	0.714 ± 0.001	2.763 ± 0.002	**3.672** ± 0.003	0.895 ± 0.001
RTCB	**Original**	0.536	0.698	-8.572	0.709	3.005	2.299	-
(PDBID:	**MMP** (Loeffler et al., 2024)	0.636 ± 0.000	0.731 ± 0.001	-8.465 ± 0.021	0.709 ± 0.001	2.599 ± 0.004	3.013 ± 0.013	0.845 ± 0.001
4DWQ)	**Similarity (≥ 0.5)** (Loeffler et al., 2024)	0.626 ± 0.000	0.723 ± 0.001	-8.511 ± 0.012	0.713 ± 0.002	2.543 ± 0.002	3.082 ± 0.031	0.760 ± 0.000
	Similarity ([0.5, 0.7]) (Loeffler et al., 2024)	0.622 ± 0.001	0.718 ± 0.000	-8.486 ± 0.021	0.713 ± 0.003	2.542 ± 0.005	3.101 ± 0.005	0.740 ± 0.001
	Similarity (≥ 0.7) (Loeffler et al., 2024)	0.639 ± 0.000	0.734 ± 0.001	-8.496 ± 0.009	0.718 ± 0.001	2.628 ± 0.001	2.868 ± 0.003	0.875 ± 0.002
	Scaffold (Loeffler et al., 2024)	0.609 ± 0.001	0.718 ± 0.000	-8.508 ± 0.026	0.711 ± 0.000	2.627 ± 0.002	2.803 ± 0.010	0.735 ± 0.002
	Scaffold Generic (Loeffler et al., 2024)	0.625 ± 0.001	0.722 ± 0.000	-8.544 ± 0.009	0.722 ± 0.002	2.551 ± 0.010	2.898 ± 0.005	0.768 ± 0.004
	DrugImprover (Liu et al., 2023a)	0.478 ± 0.001	0.618 ± 0.002	-8.701 ± 0.037	0.486 ± 0.002	**1.181** ± 0.010	2.026 ± 0.013	0.427 ± 0.001
	Molsearch (Sun et al., 2022)	0.625 ± 0.001	0.742 ± 0.001	-8.747 ± 0.009	0.719 ± 0.001	3.012 ± 0.004	2.273 ± 0.005	**0.950** ± 0.001
	MIMOSA (Fu et al., 2021)	0.631 ± 0.001	0.749 ± 0.001	-8.972 ± 0.011	0.706 ± 0.003	3.080 ± 0.007	2.561 ± 0.008	0.945 ± 0.001
	DrugEx v3(Liu et al., 2023b)	0.592 ± 0.001	0.668 ± 0.001	-8.762 ± 0.010	0.583 ± 0.002	2.488 ± 0.005	5.827 ± 0.010	0.393 ± 0.001
	CONTROLLABLEGPT (masks only)	0.675 ± 0.001	0.753 ± 0.001	-9.318 ± 0.002	**0.752** ± 0.001	2.674 ± 0.001	3.292 ± 0.002	0.883 ± 0.001
	CONTROLLABLEGPT (mask + s2s)	**0.678** ± 0.001	**0.755** ± 0.001	**-9.377** ± 0.003	0.751 ± 0.001	2.688 ± 0.001	**3.328** ± 0.005	0.890 ± 0.001

Table 1: **Main results.** A comparison of eight baselines including Original, six baselines from REINVENT 4 {MMP, Similarity (≥ 0.5), Similarity $\in [0.5, 0.7)$, Similarity ≥ 0.7, Scaffold, Scaffold Generic}, DrugImprover, Molsearch, MIMOSA, DrugEx v3 and CONTROLLABLEGPT on multiple objectives based on 3CLPro and RTCB datasets. The top two results are highlighted as **1st** and 2nd. Results are reported for 5 experimental runs.

change between X and Y, the model learns a mapping from $(X, Z) \in \mathcal{X} \times \mathcal{Z} \implies Y \in \mathcal{Y}$ during training. REINVENT 4 defined six different kinds of property change Z, including MMP for user-specified changes, different similarity thresholds, and scaffold-based alterations, where molecules share the same scaffold or generic scaffold. All baselines are fine-tuned using the cancer and COVID dataset with their respective fine-tuning methods.

Critics and evaluation metric. We evaluate seven key attributes for pharmaceutical drug discovery: 1) *Average normalized reward* is the average of the normalized values of the docking score, drug-likeness, synthesizability, solubility, and similarity across all valid molecules. This is regarded as the most crucial metric.; 2) *Average top 10% normalized reward* is the average of the normalized reward of the top 10% of molecules based on their average normalized reward; 3) *Docking score* (generated, for efficient calculation, with a surrogate docking model: see Appendix A.6) evaluates the potential of a drug to inhibit the target site. 4) *Druglikeness* assesses the probability of a molecule being a suitable drug candidate; 5) *Synthesizability* measures the synthesizability of a molecule, assigning a score of 1 for easy synthesis and a score of 10 for difficult synthesis (Ertl and Schuffenhauer, 2009); and 6) *Similarity* evaluates the similarity between original and generated SMILES using Tanimoto similarity. 7) *Solubility* evaluates a molecule's potential to dissolve in water, often measured by the water-octanol partition coefficient (LogP).

5.1. Main results

Table 1 illustrates the performance comparison between CONTROLLABLEGPT and the competing baseline methods. The results indicate that CONTROLLABLEGPT outperforms the competing baselines across all the metrics except for synthesizability. Notably, CONTROLLABLEGPT achieves the highest Tanimoto similarity score, surpassing both the current state-of-the-art, REINVENT 4, and its six variants. This implies that molecules optimized by CONTROLLABLEGPT not only exhibit structures more similar to the original drug compared to existing methods but also demonstrate improved properties across various metrics.

Additionally, when compared to the original baseline, the drugs generated by CONTROLLABLEGPT significantly enhance the original drug across the desired aspects. These results underscore the superiority and effectiveness of CONTROLLABLEGPT in controllable optimization of original drugs, preserving beneficial structures while optimizing diverse properties. In addition, CONTROLLABLEGPT with both masked and seq2seq tokens outperforms the masked token only. This demonstrate that the GPT model developed with our CMS objective surpasses causally masked modeling because of the added capabilities of CMS, such as the controllable mutation function that enables conditional expansion and contraction.

5.2. Ablation studies

Adding to the original molecule without modification. Table 2 (Task 1) visualizes the addition to the original molecule while preserving the com-

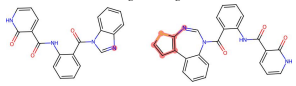
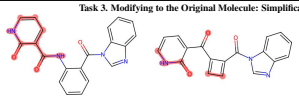
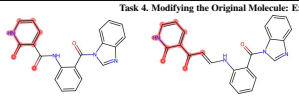
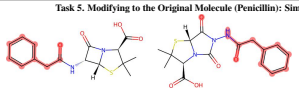
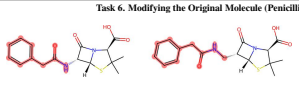
Description	Task 1. Adding to the Original Molecule Without Modification	Task 2. Modifying the Original Molecule
Molecule		
Original SMILES	O=C(Nc1ccccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O	-
Prompt	O=C(Nc1ccccc1C(=O)n1cnc2<mask_1:7>ccccc21)c1ccc[nH]c1=O<mask_1:7>	O=C(<mask_1:3>ccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O<mask_1:3>
Masked → Generated [token, length]	[None, 0] → [scc2c2, 7]	[Nc1cc, 5] → [c1s, 3]
Generated SMILES	O=C(Nc1ccccc1C(=O)n1cnc2scc2c2ccccc21)c1ccc[nH]c1=O	O=C(c1scc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O
Description	Task 3. Modifying to the Original Molecule: Simplification	Task 4. Modifying the Original Molecule: Expansion
Molecule		
Original SMILES	O=C(Nc1ccccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O	-
Prompt	O=C(<s2s_1_2:Nc1cc>ccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O<s2s_1_2:Nc1cc>	O=C(<s2s_1_10:Nc1cc>ccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O<s2s_1_10:Nc1cc>
Masked → Generated [token, length]	[Nc1cc, 5] → [c1, 2]	[Nc1cc, 5] → [/C=C/Nc1cc, 10]
Generated SMILES	O=C(c1ccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O	O=C(/C=C/Nc1ccccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O
Description	Task 5. Modifying to the Original Molecule (Penicillin): Simplification	Task 6. Modifying the Original Molecule (Penicillin): Expansion
Molecule		
Original SMILES	CC1([C@@H](N2[C@H](S1)[C@@H](C2=O)NC(=O)CC3=CC=CC=C3)C(=O)O)C	-
Prompt	C1([C@@H](N2[C@H](S1)<mask_1:6>(C2=O)NC(=O)CC3=CC=CC=C3)C(=O)O)C<mask_1:6>	CC1([C@@H](N2[C@H](S1)[C@@H]<s2s_10:(C2=O)>NC(=O)CC3=CC=CC=C3)C(=O)O)C<s2s_10:(C2=O)>
Masked → Generated [token, length]	[[C@@H], 6] → [(C2=O), 6]	[(C2=O), 6] → [(C2=O)[CH], 10]
Generated SMILES	CC1([C@@H](N2[C@H](S1)C(=O)N(C2=O)NC(=O)CC3=CC=CC=C3)C(=O)O)C	CC1([C@@H](N2[C@H](S1)[C@@H](C2=O)[CH]NC(=O)CC3=CC=CC=C3)C(=O)O)C
Toxicity Score	(Original) 2.54 → (Generated) 2.35	(Original) 2.54 → (Generated) 2.11

Table 2: **Ablation studies. Task 1&2:** Examples using masking and size hints for controllable generation. **Task 3&4:** Examples using Seq2Seq and size hints for controllable generation. **Task 5&6:** Examples using CONTROLLABLEGPT to reduce the toxicity of Penicillin while preserving its fundamental structure.

plete original structure. In this experiment, a given original molecule with the SMILES representation O=C(Nc1ccccc1C(=O)n1cnc2ccccc21)c1ccc[nH]c1=O serves as the basis. Our objective is to extend the ring in the molecule. We designed the prompt by adding a mask token $< mask_1 : 7 >$ to the specific position adjacent to the ring in the SMILES. Finally, we obtained the generated molecule with the desired features (additional ring in red) while maintaining the completeness of the original molecule structure. This study demonstrates the ability of CONTROLLABLEGPT to extend at specific positions with a specific length.

Modifying the Original Molecule. In this experiment, our goal is to alter a portion of the original molecule by modifying bonds and atoms connecting the two rings. For this purpose, we construct the prompt by substituting the original structure $Nc1cc$ with a masked token $< mask_1 : 3 >$. Table 2 (Task 2) illustrates the modification of the original molecule by removing the ring and introducing a few atoms, while retaining the majority of the structure. This demonstrate the ability of CONTROLLABLEGPT by modifying partial of molecule and random generated in specific length.

Conditional Modifying to the Original Molecule: Contraction and Expansion. This experiment aims to showcase conditional modifications to the original molecule. Unlike Task 1&2, where the focus is on modifications and expansion in a random manner, here we concentrate on generating subsequences conditioned on a partial molecule. We undertake two tasks: expanding and shrinking partial molecules based on a given subsequence. For the simplification task, we successfully reduce a length 5 subsequence, $Nc1cc$, to a length 2 token using the prompt token $< s2s_1_2 : Nc1cc >$. Conversely, for the expansion task, we extend the subsequence to a length of 10 tokens using the prompt token $< s2s_1_10 : Nc1cc >$. Both tasks yield the desired molecules, as depicted in Table 2 (Task 3&4). This demonstrates that CONTROLLABLEGPT is capable of generating molecules controllably for contraction and expansion, conditioned on specific segments of the molecule, to target specific lengths of subsequences.

Penicillin Toxicity Reduction. In this study, we utilize the ToxSmi Model (Born et al., 2023), which was trained on the Tox21 (tox) dataset, encompassing 12 different types of environmental toxicities. The toxicities reported in Table 2 (Tasks 5&6) represent the sum of 12 toxicity scores. The original molecule, Penicillin, has a predicted toxicity score of 2.54. Our proposed controllable methods demonstrate a significant reduction in the toxicity scores of the generated molecule, while preserving the core scaffold structure for preseving desired beneficial properties.

6. Conclusion

In this study, we introduce the novel Causally Masked Seq2Seq (CMS) objective and CONTROLLABLEGPT, which allows precise control over specific sequence areas for expansion, reduction, or mutation while preserving key regions and biologial structure. CONTROLLABLEGPT demonstrated superiority over eight competing baselines in Covid and Cancer drug optimization benchmarks, maintaining high Tanimoto similarity and enhancing drug properties. It also demonstrated its controllability through specific examples in ablation studies. This method highlights CONTROLLABLEGPT's capability for precise generation in drug optimization tasks, despite its limitations. For future directions, we encourage applying CONTROLLABLEGPT in fields beyond our current research scope.

ACKNOWLEDGEMENTS

This work is supported by the RadBio-AI project (DE-AC02-06CH11357), U.S. Department of Energy Office of Science, Office of Biological and Environment Research, the Improve project under contract (75N91019F00134, 75N91019D00024, 89233218CNA000001, DE-AC02-06-CH11357, DE-AC52-07NA27344, DE-AC05-00OR22725), the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

References

Ncats toxicology in the 21st century (tox21). https://ncats.nih.gov/tox21. 8

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. 2

Armen Aghajanyan, Dmytro Okhonko, Mike Lewis, Mandar Joshi, Hu Xu, Gargi Ghosh, and Luke Zettlemoyer. Htlm: Hyper-text pre-training and prompting of language models. *arXiv preprint arXiv:2107.06955*, 2021. 5

Armen Aghajanyan, Bernie Huang, Candace Ross, Vladimir Karpukhin, Hu Xu, Naman Goyal, Dmytro Okhonko, Mandar Joshi, Gargi Ghosh, Mike Lewis, et al. Cm3: A causal masked multimodal model of the internet. *arXiv preprint arXiv:2201.07520*, 2022. 2

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017. 3

Viraj Bagal, Rishal Aggarwal, PK Vinod, and U Deva Priyakumar. MolGPT: Molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling*, 62(9):2064–2076, 2021. 3

Shraddha Barke, Michael B James, and Nadia Polikarpova. Grounded copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages*, 7(OOPSLA1):85–111, 2023. 1

Nurken Berdigaliyev and Mohamad Aljofan. An overview of drug discovery and development. *Future medicinal chemistry*, 12(10):939–947, 2020. 1

Jannis Born, Greta Markert, Nikita Janakarajan, Talia B. Kimber, Andrea Volkamer, María Rodríguez Martínez, and Matteo Manica. Chemical representation learning for toxicity prediction. *Digital Discovery*, pages –, 2023. doi: 10.1039/D2DD00099G. URL http://dx.doi.org/10.1039/D2DD00099G. 8

Thorsten Brants, Ashok Popat, Peng Xu, Franz Josef Och, and Jeffrey Dean. Large language models in machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 858–867, 2007. 3

Mia Xu Chen, Orhan Firat, Ankur Bapna, Melvin Johnson, Wolfgang Macherey, George Foster, Llion Jones, Niki Parmar, Mike Schuster, Zhifeng Chen, et al. The best of both worlds: Combining recent advances in neural machine translation. *arXiv preprint arXiv:1804.09849*, 2018. 2

Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in neural information processing systems*, 29, 2016. 3

Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta: large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*, 2020. 2

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 2

Rahul Dey and Fathi M Salem. Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pages 1597–1600. IEEE, 2017. 2

Dave Epstein, Allan Jabri, Ben Poole, Alexei Efros, and Aleksander Holynski. Diffusion self-guidance for controllable image generation. *Advances in Neural Information Processing Systems*, 36:16222–16239, 2023. 3

Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1:1–11, 2009. 7

Kawin Ethayarajh. How contextual are contextualized word representations? comparing the geometry of bert, elmo, and gpt-2 embeddings. *arXiv preprint arXiv:1909.00512*, 2019. 1

Luciano Floridi and Massimo Chiriatti. Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30:681–694, 2020. 1

Nathan C Frey, Ryan Soklaski, Simon Axelrod, Siddharth Samsi, Rafael Gomez-Bombarelli, Connor W Coley, and Vijay Gadepally. Neural scaling of deep chemical models. *Nature Machine Intelligence*, 5(11):1297–1305, 2023. 3

Tianfan Fu, Cao Xiao, Xinhao Li, Lucas M Glass, and Jimeng Sun. Mimosa: Multi-constraint molecule sampling for molecule optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 125–133, 2021. 6, 7

Philip Gage. A new algorithm for data compression. *The C Users Journal*, 12(2):23–38, 1994. 6

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 3

Alex Graves and Alex Graves. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45, 2012. 2

Jiazhen He, Huifang You, Emil Sandström, Eva Nittinger, Esben Jannik Bjerrum, Christian Tyrchan, Werngard Czechtizky, and Ola Engkvist. Molecular optimization by capturing chemist's intuition using deep neural networks. *Journal of cheminformatics*, 13(1):1–17, 2021. 1, 3, 6, 13

Jiazhen He, Eva Nittinger, Christian Tyrchan, Werngard Czechtizky, Atanas Patronov, Esben Jannik Bjerrum, and Ola Engkvist. Transformer-based molecular optimization beyond matched molecular pairs. *Journal of cheminformatics*, 14(1):18, 2022. 1, 3, 6, 13

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3

Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. A simple language model for task-oriented dialogue. *Advances in Neural Information Processing Systems*, 33:20179–20191, 2020. 2

Peter W Kenny and Jens Sadowski. Structure modification in chemical databases. *Chemoinformatics in drug discovery*, pages 271–285, 2005. 6

Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*, 2019. 3

Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 3

Greg Landrum et al. RDkit: Open-source cheminformatics software. https://www.rdkit.org. Accessed Oct 2023. 3

Greg Landrum et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8(31.10):5281, 2013. 1

Bowen Li, Xiaojuan Qi, Thomas Lukasiewicz, and Philip Torr. Controllable text-to-image generation. *Advances in neural information processing systems*, 32, 2019. 3

Junyi Li, Tianyi Tang, Wayne Xin Zhao, Jian-Yun Nie, and Ji-Rong Wen. Pre-trained language models for text generation: A survey. *ACM Computing Surveys*, 56(9):1–39, 2024. 2

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. Diffusion-lm improves controllable text generation. *Advances in Neural Information Processing Systems*, 35:4328–4343, 2022. 3

Xun Liang, Hanyu Wang, Yezhaohui Wang, Shichao Song, Jiawei Yang, Simin Niu, Jie Hu, Dan Liu, Shunyu Yao, Feiyu Xiong, et al. Controllable text generation for large language models: A survey. *arXiv preprint arXiv:2408.12599*, 2024. 3

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. 2

Xuefeng Liu, Songhao Jiang, Archit Vasan, Alexander Brace, Ozan Gokdemir, Thomas Brettin, and Fangfang Xia. Drugimprover: Utilizing reinforcement learning for multi-objective alignment in drug optimization. In *NeurIPS 2023 Workshop on New Frontiers of AI for Drug Discovery and Development*, 2023a. 1, 3, 6, 7, 13

Xuhan Liu, Kai Ye, Herman WT van Vlijmen, Adriaan P IJzerman, and Gerard JP van Westen. Drugex v3: scaffold-constrained drug design with graph transformer-based reinforcement learning. *Journal of Cheminformatics*, 15 (1):24, 2023b. 6, 7

Hannes H Loeffler, Jiazhen He, Alessandro Tibo, Jon Paul Janet, Alexey Voronov, Lewis H Mervin, and Ola Engkvist. Reinvent 4: Modern ai–driven generative molecule design. *Journal of Cheminformatics*, 16(1): 20, 2024. 1, 3, 6, 7

Eugene N Muratov, Rommie Amaro, Carolina H Andrade, Nathan Brown, Sean Ekins, Denis Fourches, Olexandr Isayev, Dima Kozakov, José L Medina-Franco, Kenneth M Merz, et al. A critical overview of computational approaches employed for covid-19 drug discovery. *Chemical Society Reviews*, 50(16):9121–9151, 2021. 1

Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith B Hall, Daniel Cer, and Yinfei Yang. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. *arXiv preprint arXiv:2108.08877*, 2021. 2

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022. 1

Chandrika Prasad, Jagdish S Kallimani, Divakar Harekal, and Nicy Sharma. Automatic text summarization model using seq2seq technique. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pages 599–604. IEEE, 2020. 2

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019. 3

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67, 2020. 3

Daniel Rothchild, Alex Tamkin, Julie Yu, Ujval Misra, and Joseph Gonzalez. C5T5: Controllable generation of organic molecules with transformers. *arXiv preprint arXiv:2108.10307*, 2021. 3

AM Rush. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*, 2015. 3

Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*, 2015. 6

Tian Shi, Yaser Keneshloo, Naren Ramakrishnan, and Chandan K Reddy. Neural abstractive text summarization with sequence-to-sequence models. *ACM Transactions on Data Science*, 2(1):1–37, 2021. 2

Mengying Sun, Jing Xing, Han Meng, Huijun Wang, Bin Chen, and Jiayu Zhou. Molsearch: search-based multi-objective molecular generation and property optimization. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 4724–4732, 2022. 6, 7

I Sutskever. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*, 2014. 3

Duyu Tang, Nan Duan, Zhao Yan, Zhirui Zhang, Yibo Sun, Shujie Liu, Yuanhua Lv, and Ming Zhou. Learning to collaborate for question answering and asking. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1564–1574, 2018. 2

Gaurav Tiwari, Arushi Sharma, Aman Sahotra, and Rajiv Kapoor. English-hindi neural machine translation-lstm seq2seq and convs2s. In *2020 International Conference on Communication and Signal Processing (ICCSP)*, pages 871–875. IEEE, 2020. 2

Xiaochu Tong, Xiaohong Liu, Xiaoqin Tan, Xutong Li, Jiaxin Jiang, Zhaoping Xiong, Tingyang Xu, Hualiang Jiang, Nan Qiao, and Mingyue Zheng. Generative models for de novo drug design. *Journal of Medicinal Chemistry*, 64(19):14011–14027, 2021. 1

Christian Tyrchan and Emma Evertsson. Matched molecular pair analysis in short: algorithms, applications and limitations. *Computational and structural biotechnology journal*, 15:86–90, 2017. 6

Archit Vasan, Rick Stevens, Arvind Ramanathan, and Vishwanath Venkatram. Benchmarking language-based docking models. 2023. 14

A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 3

Changhan Wang, Yun Tang, Xutai Ma, Anne Wu, Sravya Popuri, Dmytro Okhonko, and Juan Pino. Fairseq s2t: Fast speech-to-text modeling with fairseq. *arXiv preprint arXiv:2010.05171*, 2020. 2

Wenxuan Wang, Wenxiang Jiao, Yongchang Hao, Xing Wang, Shuming Shi, Zhaopeng Tu, and Michael Lyu. Understanding and improving sequence-to-sequence pretraining for neural machine translation. *arXiv preprint arXiv:2203.08442*, 2022. 2

David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988. 2, 4

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992. 3

Linjuan Wu, Peiyun Wu, and Xiaowang Zhang. A seq2seq-based approach to question answering over knowledge bases. In *Semantic Technology: 9th Joint International Conference, JIST 2019, Hangzhou, China, November 25–27, 2019, Revised Selected Papers 9*, pages 170–181. Springer, 2020. 2

Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang. A brief overview of chatgpt: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10 (5):1122–1136, 2023. 1

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mt5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*, 2020. 2

Liu Xuefeng, Tien Chih-Chan, Ding Peng, Jiang Songhao, and Stevens Rick. Entropy-reinforced planning with large language models for de novo drug discovery. *ICML 2024*, 2024. 3, 13, 15

Wilson Yan, Yunzhi Zhang, Pieter Abbeel, and Aravind Srinivas. Videogpt: Video generation using vq-vae and transformers. *arXiv preprint arXiv:2104.10157*, 2021. 1

Gokul Yenduri, Gautam Srivastava, Praveen Kumar Reddy Maddikunta, Rutvij H Jhaveri, Weizheng Wang, Athanasios V Vasilakos, Thippa Reddy Gadekallu, et al. Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *arXiv preprint arXiv:2305.10435*, 2023. 1

Hanqing Zhang, Haolin Song, Shaoyu Li, Ming Zhou, and Dawei Song. A survey of controllable text generation using transformer-based pre-trained language models. *ACM Computing Surveys*, 56(3):1–37, 2023. 3

A. Appendix

A.1. Pre-training Details

We used the ZINC dataset, filtering for Standard, In-Stock, and Drug-Like molecules, resulting in approximately 11 million molecules.

In the second phase of pre-training, we first trained for 10 epochs using a single mask. Subsequently, we trained for another 40 epochs with an equal probability of using either one or two masks. For each epoch, the masks were regenerated to create a more comprehensive masked dataset.

In the third phase of pre-training, we applied different mask configurations with specific probabilities: [one mask (0.1), two masks (0.1), one mask and one seq2seq (0.4), two masks and one seq2seq (0.4)] and train 20 epochs.Similar to the second phase, the masks were regenerated for each epoch to enhance the comprehensiveness of the masked dataset.

A.2. Baselines fine-tuning datasets

As outlined in Section 5, all baseline models are fine-tuned using the Cancer and COVID dataset, following their respective fine-tuning methodologies. For this process, we utilize one million compounds from the ZINC15 dataset, docked to the 3CLPro protein (PDB ID: 7BQY), which is linked to SARS-CoV-2, and the RTCB protein (PDB ID: 4DWQ), associated with human cancer. These datasets, sourced from the latest Cancer and COVID dataset by Liu et al. (2023a), are consistently applied across all baselines.

Additionally, these datasets are employed for molecular generation in our proposed methods, with further details on the generation process provided in Section A.3.

A.3. Generation

For each mask and seq2seq, we utilize three random variables: the start index, the number of tokens to be masked, and the number of tokens to be generated. During generation, we apply two settings: [one mask + one seq2seq, and two masks], resulting in a total of six random variables for each setting.

During the generation phase, we randomly sample these six variables 10,000 times, using them as prompts for generation, regardless of whether the generated SMILES are valid or not. In addition, for a given prompt molecule, we adopt TOPPK (Xuefeng et al., 2024) for generation strategy.

After generation, for each prompt molecule/SMILES, we select the top 10 generated molecules/SMILES based on their average normalized reward. The mean of these top 10 molecules/SMILES is then used to obtain the final result for the prompt molecule/SMILES. For a fair evaluation, we adopt it for baselines as well.

A.4. Baseline REINVENT 4

Following are detailed description of six different kinds of property change Z included in REINVENT 4 He et al. (2022; 2021)

- **MMP:** There are user-defined desirable property changes between molecules X and Y.
- **Similarity** ≥ 0.5**:** The Tanimoto similarity between molecules X and Y is greater than 0.5.
- **Similarity** $\in [0.5, 0.7)$**:** The Tanimoto similarity between the pair (X, Y) ranges from 0.5 to 0.7.
- **Similarity** ≥ 0.7**:** The Tanimoto similarity between molecules X and Y is greater than 0.7.
- **Scaffold:** Molecules X and Y share the same scaffold.
- **Scaffold generic:** Molecules X and Y share the same generic scaffold.

A.5. BPE Tokenization

Byte Pair Encoding (BPE) is a tokenization algorithm initially designed for data compression and later adapted for use in NLP, particularly in the preprocessing of text for deep learning models. The core idea behind BPE is to iteratively merge the most frequent pair of consecutive bytes (or characters in the context of text) into a single, new byte (or token), thereby reducing the size of the data to be processed. This method has been particularly influential in the development of language models and machine translation systems. The BPE method follows these main steps:

1. **Initial vocabulary preparation:** The text is divided into a sequence of characters or symbols, and a special end-of-word symbol (like <\w> or another unique marker) is added to each word to distinguish between the same character sequence occurring within a word and at the end of a word.

2. **Frequency Count:** The algorithm counts the frequency of each pair of adjacent characters (or symbols) in the text.

3. **Iterative Merging:**

   - Identify the most frequent pair of adjacent characters.
   - Merge this pair into a new single symbol (this does not mean changing the text itself but rather how the algorithm interprets the text).
   - Update the frequency count of all pairs, considering the newly created symbol.
   - Repeat this process for a predetermined number of iterations or until a desired vocabulary size is reached.

4. **Tokenization:** Once the merging process is complete, the original text can be tokenized (i.e., divided into a sequence of tokens) using the final set of symbols, including the merged ones. This results in a text representation where frequent words or subwords are encoded as single tokens, and less common words are broken down into smaller tokens.

A significant benefit of BPE lies in its capacity to manage rare and out-of-vocabulary words effectively. Since BPE operates at the character level, it can segment words that were not encountered during training, thus reducing the negative effects of unfamiliar words on the model's performance. In contexts where tokens of various lengths are randomly masked and relocated to the end of the sequence, as proposed in section 4.1, there's a high likelihood of generating a considerable number of unfamiliar tokens. BPE's approach is particularly beneficial here, as it ensures that the model can still process and understand these novel token sequences by breaking them down into familiar subunits, thereby maintaining robustness and reducing the potential degradation in performance due to unexpected or rare words.

A.6. Surrogate model

The surrogate model (Vasan et al., 2023) is a simplified version of a BERT-like transformer, widely employed in natural language processing. In this model, tokenized SMILES strings are inputted and then positionally embedded. The outputs are subsequently fed into a series of five transformer blocks, each comprising a multi-head attention layer (with 21 heads), a dropout layer, layer normalization with residual connection, and a feedforward network. The feedforward network consists of two dense layers followed by dropout and layer normalization with residual connection. Following the stack of transformer blocks, a final feedforward network is employed to produce the predicted docking score. The validation r^2 values are 0.842 for 3CLPro dataset and 0.73 for the RTCB dataset.

A.7. Performance scales with mask length

We conducted an analytical experiment to examine how performance scales with mask length. The results show that the smaller the difference between the length of the generated span and the masked span, the higher the validity will be. In our settings, the validity reaches its highest at 90% when the length of the generated span is between 5 and 10.

A.8. Computing infrastructure and wall-time comparison

We trained our docking surrogate models using 4 nodes of a supercomputer, each node equipped with CPUs (64 cores) and 4 A100 GPUs. The training time for each model was approximately 3 hours. We performed additional pretraining on a cluster consisting of CPU nodes (approximately 280 cores) and GPU nodes (approximately 110 Nvidia GPUs, ranging from Titan X to A6000, primarily configured in 4- and 8-GPU setups).

Pretraining utilizes 8 A100 GPUs, while one single generation uses a single Tesla T4 GPU. Based on the computing infrastructure, pretraining details as described in Appendix A.1 and generation details as described in Appendix A.3, we obtained the wall-time comparison in Table 3 as follows.

A.9. Hyperparameters and architectures

Table 4 provides a list of hyperparameter settings we used for our experiments.

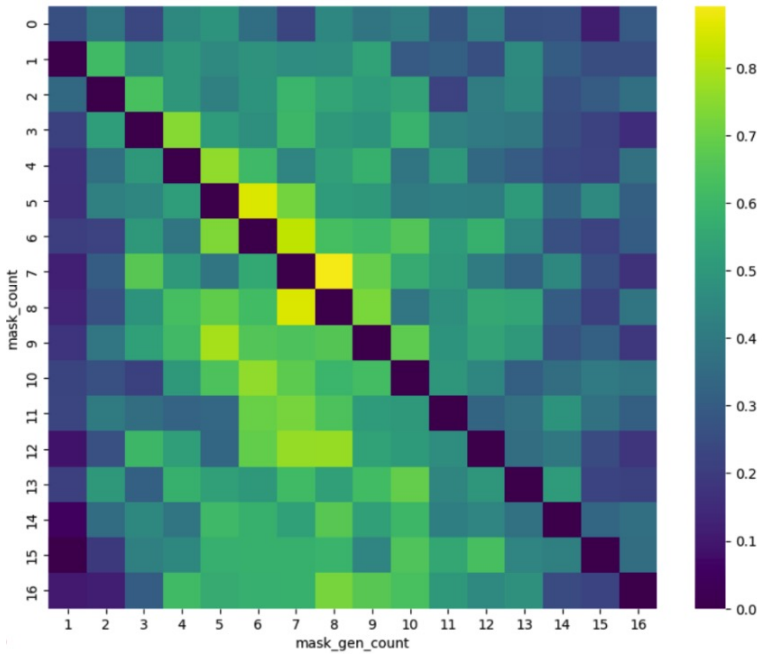


Figure 6: The x-axis of this heatmap is the number of tokens that are generated, and the y-axis of this heatmap is the number of tokens that are masked in prompt. Among 10k molecules generated, each cell indicates the average validity of the molecule generated for a specific combination of [generated tokens, masked prompt tokens]. Lighter colors indicate higher validity, while darker colors represent lower validity, as shown by the color bar on the right side of the heatmap.

	Total Run Time
Initial Phase Pretraining	18h
Second Phase Pretraining	48h
Third Phase Pretraining	20h
Generation 10k times for one molecule	15mins

Table 3: Wall-time comparison between different methods.

For experimentation, 1280 molecules from each of the RTCB and 3CLPro datasets, with docking scores ranging from -14 to -6, are selected. This range is based on (Xuefeng et al., 2024).

In addition, when calculating the average normalized reward for the original molecule, where similarity is not considered, we select the weights for docking, drug-likeness, synthesizability, and solubility as $[0.25] \times 4$.

Parameter	Value
Pretraining	
Learning rate	$5 \times e^{-5}$
Batch size	24
Optimizer	Adam
# of Epochs for Training Initial Phase	10
# of Epochs for Training Second Phase	50
# of Epochs for Training Third Phase	20
Model # of Params	$124M$
Generation	
# of Molecules Optimized	1280
TopK	$[10, 15, 20]$
TopP	$[0.85, 0.9, 0.95]$

Table 4: Hyperparameters.