FADS and semi-rational design modified T7 RNA polymerase reduced dsRNA production, with lower terminal transferase and RDRP activities

Qiongwei Tang[1], Sisi Zhu[1], Nannan Hu[2], Sainan Yin[2], Yuhong Yang[2], Yigang Teng[1], Dongliang Song[1], Xiang Liu[1,2*]

1. Yeasen Biotechnology (Shanghai) Co., Ltd. Building 3, No. 537, Qingdai Road, Pudong New Area, Shanghai 201318, China.

2. Molefuture Biotechnology (Shanghai) Co., Ltd. Building 3, No. 537, Qingdai Road, Pudong New Area, Shanghai 201318, China.

* liuxiang@yeasen.com; liuxiang@molefuture.com

Abstract

T7 RNA polymerase (T7 RNAP) is the preferred tool for *in vitro* transcription (IVT), it synthesizes mRNA while accompanied by the generation of dsRNA by-products. This undesirable dsRNA triggers immune stress responses, compromises therapeutic efficacy, and raises safety concerns. To evolve T7 RNAP for reduced dsRNA, we pursued two complementary strategies. Firstly, the FADS (fluorescence-activated droplet sorting) based on molecular beacons was used to screen random libraries with diversity exceeding 10^5. Secondly, we constructed several single-site saturated libraries to facilitate the transition of T7 RNAP from the initiation to the elongation conformation. These libraries were screened using the traditional microplate-based dual-probe screening technique. Both approaches identified two dominant variants: Mut1 (V214A) and Mut7 (F162S/A247T) from FADS, Mut11 (K180E) and Mut14 (A70Q) from saturated libraries. Furthermore, the combinatorial mutant Mut17 (A70Q/F162S/K180E), generated via DNA shuffling, exhibited significantly reduced dsRNA production compared to the wild-type under various conditions, ranging from 0.18% to 1.80%, with a minimum value of 0.5 pg/μg. Cell experiments confirmed that variants generated capped-mRNA with similar quality and quantity to the wild-type, while significantly reducing immune stress response in cells. These results indicate the compatibility and broad potential applications of these mutations. We then observed a close correlation between the production of dsRNA and the activities of T7 RNAP in terminal transferase and RDRP. Particularly, the terminal transferase activity appears to play a critical role in dsRNA generation. These findings align with the mechanism of dsRNA formation during IVT and provide new screening criteria for further evolution of T7 RNAP.

Introduction

In recent years, mRNA-based drugs have garnered significant attention and attracted research focus. Particularly, mRNA vaccines, as a relatively new vaccine category, have demonstrated promise as an immunotherapy strategy[1,2,3]. Compared to traditional vaccines, mRNA offers several advantages. One crucial advantage is that the design and production of mRNA vaccines are independent of sequences. Different mRNA vaccines can utilize similar production and purification processes, enabling rapid and flexible development, and cost savings[4]. Additionally, mRNA vaccines can stimulate immune cells to secrete cytokines like tumor necrosis factor-α (TNF-α) and interferon-α (IFN-α),

activating enduring adaptive immune responses without additional adjuvants[5]. To ensure clinical application, multiple steps are involved in mRNA production, including mRNA generation, DNA digestion, purification, and downstream processing. RNA polymerase facilitates *in vitro* transcription (IVT) to produce mRNA from linear dsDNA (double-stranded DNA) templates, the products typically containing the target protein's open reading frame (ORF), UTR (untranslated region), 5'-cap, and 3'-poly(A) tail. Downstream purification processing involves precipitation and HPLC to remove impurities like redundant raw materials, DNA templates, proteins, and dsRNA (double-stranded RNA) by-products from transcription[6,7].

dsRNA, a pathogen-associated molecular pattern (PAMP), activates pattern recognition receptors in cellular compartments. Its recognition triggers the production of type I interferon, upregulating the expression and activity of protein kinase R (PKR; also known as EIF2AK2) and 2'-5'-oligoadenylate synthetase (OAS). Consequently, translation of intracellular proteins is suppressed, while degradation occurs for mRNA and ribosomal RNA[8,9,10]. To ensure the safety and efficacy of mRNA vaccines, it is necessary to remove dsRNA from the mRNA products. Currently, HPLC purification is an effective method for this purpose[8]. However, it increases time and costs, posing challenges for large-scale production and potentially leading to RNA degradation or reduced yield.

The bacteriophage T7 RNA polymerase (T7 RNAP) is the most commonly used enzyme for IVT. The 99 kDa protein consists of an N-terminal domain (residues 1-267) and a C-terminal domain (residues 268-883)[11-14]. T7 RNAP is known for its simplicity and ability to catalyze RNA synthesis with high yield and fidelity *in vitro*. However, during the transcription process, T7 RNAP can produce by-products such as abortive short RNAs, prematurely terminated RNAs, and 3'-extended RNAs, which creates favorable conditions for the formation of dsRNA[11,14-22].

The transcription of T7 RNAP involves three stages: initiation, elongation, and termination. During initiation, T7 RNAP recognizes specific promoters, opens the DNA double helix, and guides the template strand into the active site. The formation of this initiation complex marks the beginning of transcription. When the nascent mRNA reaches 2-6 nt, a conformational change occurs in the N-terminus of T7 RNAP. Multiple helices (D, E, F, G, I, J) in the promoter-binding region rotate, despite maintaining contact with the promoter. Simultaneously, helix C (residues 28-71) and subdomain H (residues 151-190) undergo a process of refolding. Helix C2 (residues 46-55) stacks onto the helix C1 (residues 28-41), resulting in the fusion of the original two short helices into a continuous long helix. Additionally, the loop (residues 173-186) within subdomain H transforms into an α-helix. These conformational rearrangements lead to the elimination of the promoter-binding site, expansion of the active pocket, and opening of the product RNA exit channel. Once the transcript reaches 10-12 nt, the conformational transition is complete, and the DNA promoter is released, marking the start of the elongation phase[11,13,14].

T7 RNAP continues to transcribe along the DNA template until it encounters a specific terminator or reaches the end of the linear DNA template, resulting in the generation of full-length RNA, known as run-off RNA[11,14-20]. However, premature termination during elongation can lead to incomplete RNAs, while non-template-dependent additions at the 3'-end of run-off RNA can generate 3'-extended products[21,22]. Additionally, the transition from initiation to elongation

conformation is a challenging process that often leads to the accumulation of short single-stranded RNA, known as abortive RNA[11,14-17]. With the assistance of RDRP (RNA-dependent RNAP) activity of T7 RNAP, these 3'-extended products and abortive RNAs contribute to the formation of loopback dsRNA or complementary pairing-dependent dsRNA, leading to immunostimulatory effects in host cells[21-25].

Fluorescence-activated droplet sorting (FADS) is an emerging microfluidic-based screening technology that has gained popularity in the directed evolution of enzymes and strain selection. It offers several advantages, including ultra-high throughput, high sensitivity, and broad applicability[26,27]. FADS operates on a mechanism similar to fluorescence-activated cell sorting (FACS), where the device detects fluorescence of each droplet instead of cell and activates the electrodes to divert positive droplets into collection channels[28]. A typical droplet in FADS contains a lysing reagent, fluorescent substrate, and a microbial cell that overexpresses a mutant of the target enzyme. After sorting, the DNA within positive droplets can be amplified through PCR and subsequently used for sequence analysis to determine the variation types and abundance of dominant variants[29]. While in certain cases, secreted proteins or special substrates can directly pass through the cell membrane to activated fluorescence of droplet, without the need for cell lysis. This method enables direct culture of cells within positive droplets, facilitating the amplification of target clones with desired mutations[29].

Compared to traditional screening techniques, FADS provides a remarkable improvement in throughput, exceeding 10^3-fold, and significantly reduces reagent consumption by more than 10^6-fold. This is achieved due to the ultra-high throughput of FADS, which can reach 10^6-10^8 droplets per day while allowing for a reduction in reaction volumes within droplets to the picolitre range[30]. FADS has proven effective in directed evolution of various enzymes by directly modifying substrates with fluorescent groups or coupling with other reactions, including proteases, polymerases, RNase, glycosidases, and dehydrogenases[31]. The emergence of similar technologies, such as AADS (Absorbance-activated droplet sorting), BADS (BDD electrode-activated droplet sorting), and MADS (Mass-activated droplet sorting), has greatly expanded the application of high-throughput screening in enzyme evolution[32]. Moreover, there is great anticipation regarding the combination of these high-throughput screening technologies with IVTT (*In vitro* transcription and translation). IVTT is considered a valuable approach for expressing enzymes with cytotoxicity while also eliminating protein expression variations caused by the host[33].

We propose that preventing the production of dsRNA during IVT is a more efficient and economical approach compared to post-transcriptional purification methods. As such, our study aimed to reduce dsRNA formation during transcription by evolution of T7 RNAP. Through a combination of directed evolution and semi-rational design, we identified several beneficial mutations. These mutations were then randomly combined, resulting in a T7 RNAP variant that more effectively reduces the generation of dsRNA.

Result

Designing the FADS procedure

To establish a screening system for T7 RNAP, we optimized a molecular beacon-based method to detect the products

of IVT. This real-time monitoring approach has shown a positive correlation between the fluorescence of the probe and RNA yield[34]. We hypothesized that positioning the molecular beacon at the 3'-end of long mRNA could potentially enable the distinction of RNA conformation and integrity. **Figure 1A (right panel)** illustrates our design, including various single-stranded DNA sequences representing target products (run-off RNA: BP-20), dsRNA with varying lengths (BP-0, BP-5, BP-10, BP-15: numbers indicate bases that can complementarily bind to the molecular beacon), and the incomplete mRNA.

After incubating equimolar amounts of single-stranded DNA with the molecular beacon at 37°C, we observed a significant increase in fluorescence intensity in the reaction tubes as the number of complementary bases increased. Notably, the DNA lacking a complementary region, simulating incomplete mRNA, exhibited the lowest fluorescent after incubation **(Figure 1A, left panel)**. These results indicate that this method can effectively discriminate the conformation and integrity of single-stranded nucleic acids, making it suitable for mutant screening. In IVT reactions, theoretically, mutants with higher fluorescence intensity are expected to produce fewer dsRNA or/and incomplete mRNA.

For screening T7 RNAP, FADS was chosen to meet the high-throughput requirements of protein evolution. We validated the aforementioned probe-based method by utilizing purified T7 RNAP and *E. coli* strains BL21 expressing T7 RNAP. The fluorescence of reaction tubes showed a significant increase with higher enzyme or cell concentrations **(Figure S1A and Figure S1B)**. Moreover, when different lengths of DNA were used as templates, the signal-to-noise ratio of the system remained above 2 **(Figure S1C)**, indicating that the probe-based method is applicable for FADS. To prevent premature fragmentation of cells, we used a PDMS (polydimethylsiloxane) chip with dual aqueous phases, where lysozyme was mixed with the cells on the chip and exerted its activity within the droplets, as shown in **Figure 1B**. The number of droplets was controlled to be roughly ten times the number of cells, aiming to ensure that each microdroplet contained at most one cell. Droplets were collected into a syringe and incubated at 37°C to facilitate cell lysis and RNA transcription. Subsequently, the droplets were introduced into a sorting chip, and only positive droplets exhibiting fluorescence above the threshold were collected for further investigation.

FADS-based screening of T7 RNAP

We generated several random mutant libraries with a diversity ranging from 10^5 to 10^6 **(Table 1)**, using error-prone PCR. T7 RNAP expression was induced in each library, and droplets were generated for further analysis. These droplets were incubated to facilitate lysozyme-mediated cell lysis and T7 RNAP-driven transcription. To determine the sorting threshold, droplets encapsulating *E. coli* expressing wild-type T7 RNAP were prepared using the same procedure and incubation conditions.

For the first round of mutant screening, the threshold was set at 1.5 times the maximum fluorescence of the wild-type, while for the second round, it ranged from 2 to 2.5 times the wild-type value. Considering the fluctuations in droplet fluorescence during each measurement, the screening threshold also varied and was typically determined by multiplying the value of the wild-type group obtained from the simultaneous test by a specific coefficient **(Table 1)**.

In the case of random library L3, 3539 and 122 positive droplets were collected after the first and second rounds of sorting, respectively, with fluorescence increases of 22.97% and 44.28% compared to the initial library **(Figure 2A)**. The gradual increase in fluorescence indicates an improved abundance of dominant variants in the library under equal cell count conditions. Similarly, library L5 yielded 6793 and 204 positive droplets, with fluorescence increases of 36.38% and 55.35%, respectively **(Figure 2B)**. The distribution of fluorescence is depicted in **Figure S2A and Figure S2B**. Following screening, a significant increase was observed in the proportion of droplets with high fluorescence, confirming the effectiveness of screening once again.

After PCR amplification, we conducted next-generation sequencing (NGS) on the DNA extracted from positive droplets. The single-point mutations with a frequency exceeding 5% are illustrated in **Figure 2C and Figure 2D**. Among them, V214A (21.93%) and V186I (15.55%) were the most frequent amino acid mutation types in their respective libraries. We also identified several high-frequency silent mutations, such as G664G in L3 and G626G in L5, indicating that these sites may serve as hotspots during error-prone PCR. Interestingly, Sanger sequencing revealed a double-mutant variant with F162S and A247T, which explains why the frequencies of F162S and A247T mutations in L5 are closely matched.

We isolated clones and purified the proteins containing the mentioned sites, obtaining T7 RNAP variants named Mut1 to Mut10 **(Table 1)**. Subsequently, we analyzed the dsRNA content in their transcription products. As shown in **Figure 2E and Figure 2F**, the dsRNA content of Mut1, Mut7, and Mut9 was significantly lower than that of the wild-type, confirming the validity of our screening. However, Mut9 exhibited a transcriptional activity only 0.1% of wild-type, resulting in extremely low RNA production. Coincidentally, Mut2 showed an approximate 8% improvement in RNA integrity, while its activity was substantially lower than wild-type **(data not shown)**. These two variants may require additional evolution steps to enhance their activity and RNA yield.

It is perplexing that the products of Mut3, Mut4, and Mut5 exhibited significantly higher dsRNA content **(Figure 2E)**. We speculate that this may be due to false-positive results of FADS, or these variants exhibited increased production of another source of dsRNA during transcription, namely the dsRNA dependence on complementary pairing[35]. Although we screened T7 RNAP with reduced dsRNA production using FADS successfully, the enrichment of low-yield variants like Mut9 and Mut2 highlights the need for stricter monitoring of yield during the screening.

Semi-rational design of T7 RNAP

Due to limited data, our structural and energetic analysis of Mut1 and Mut7 did not provide conclusive insights into the mechanism behind dsRNA reduction. Therefore, we decided to explore potential sites in spatially adjacent regions of F162, V214, or A247. On the other hand, we hypothesize that facilitating conformational transitions of T7 RNAP can significantly decrease the formation of abortive RNAs and ultimately reduce the dsRNA by-products[35]. An effective approach to achieve this goal is by reducing the free energy of the elongation conformation. This ensures a smoother transition of conformation and enhances the stability of the elongation complex, thus favoring the formation of run-off RNA. As described in the Materials and methods section, we conducted a semi-rational design exploration.

Through a comprehensive analysis of the protein structure and calculations using FoldX and PROSS **(Table S1)**, 43, 47, 70, 180, 189, 228, 237, 238, 243, 743, and 747 were selected as preliminary candidates **(Figure 3A and Figure 3B)**. Among these positions, 43, 47, 70, 180, and 189 are located in regions with significant conformational changes, while 228, 237, 238, 243, 743, and 747 are in the intercalating hairpin or specificity loop, adjacent to V214 and A247. Additionally, PROSS suggested the site 383 as a candidate **(Table S1)**, which is located in the non-template DNA strand-binding region. It believed to stabilize the non-template DNA and enhance the processivity of T7 RNAP, thus reducing incomplete RNA. Since mutations at 43 and 47 have already been proven to be beneficial for reducing dsRNA[36,37], we finally created ten single-site (A70, K180, D189, S228, V237, G238, T243, A383, I743, and L747) saturated libraries and employed the traditional microtiter plate method for screening. The selection of a saturated library aims to avoid potential functional impairments of site-directed mutagenesis caused by the limitations of simulation calculations.

To ensure a comprehensive evaluation of each variant, we improved the molecular beacon screening system. As shown in the upper right inset of **Figure 4A**, the 3'-end FAM probe is consistent with the function of beacon used in FADS, which is designed to screen mutants with reduced loopback dsRNA or incomplete RNA. The Cy5 probe positioned at the 5'-end of the RNA allows for the analysis of the yield of RNA[34], and perhaps provides an indication of the molar ratio between abortive RNA and run-off RNA. Sorting mutants based on the FAM to Cy5 ratio in each well ensures the enrichment of high-yield variants with significantly reduced dsRNA levels. We analyzed the fluorescence ratio changes over time for Mut1 and Mut10, observed that the ratios across various wells tended to converge and stabilize after 60 minutes. This indicates that the probes have annealed to the mRNA in a stable manner, and the fluorescence ratio at this point can represent the quality of the products. **(Figure S2 A and Figure S2 B)**.

After screening over 1000 mutants, we identified six mutants with fluorescence ratios higher than the wild-type **(Figure 4A)**: Mut11 (K180E), Mut12 (S228F), Mut13 (G238L), Mut14 (A70Q), Mut15 (A383H), and Mut16 (I743L). We then quantified the transcriptional products of these mutants. The total RNA yield of all six mutants did not significantly differ from the wild-type, indicating comparable overall transcription efficiency. Expectedly, the dsRNA content of Mut11 and Mut14 was significantly lower compared to the wild-type **(Figure 4B)**.

Similar to the FADS experiment, the screening results from the saturated library revealed certain variants, particularly Mut12 and Mut15, where the dsRNA content did not align with the fluorescence ratio ranking. This discrepancy implies the presence of additional factors during the screening process that were not considered. To capture a wider range of beneficial variants, we will lower the threshold during subsequent screening.

Generating Novel Mutants via DNA Shuffling

To further optimize the T7 RNAP, the four variants with low dsRNA content while meeting production requirements were selected. These mutants had specific mutations at five sites: V214A, F162S, A247T, K180E, and A70Q. We then constructed a DNA shuffling library consisting of approximately 500 clones, encompassing a theoretical total of 120 mutation types. Through screening with microtiter plates, we identified a variant that exhibited lower dsRNA

204 production compared to its parental T7 RNAP, named Mut17 (A70Q/F162S/K180E).

205 We subsequently analyzed the transcriptional products of Mut17 and its parental variants (Mut14, Mut11, and Mut7).
206 As shown in **Figure 5**, all four variants exhibited significant reductions in dsRNA production during transcription
207 compared to the wild-type. Mut11 exhibited approximately 7.28% of the wild-type's dsRNA content, while Mut14 and
208 Mut7 had dsRNA levels of 3.49% and 3.31%, respectively. Remarkably, Mut17, the second-generation mutant, showed
209 significantly lower dsRNA content of only 1.80% and as low as 0.007 ng/μg in the solvent system of screening.

210 ***Immunogenicity of IVT products from mutants***

211 Next, we evaluated the immunogenicity of the IVT products in murine RAW264.7 cells, and assessed the expression
212 of EGFP mRNA in HEK293 cells. The 1kb dsDNA served as the template for T7 RNAP and its variants, containing
213 the EGFP ORF. After introducing an equal amount of RNA, we detected the transcription and translation of IFN-β, a
214 key effector in the immune response of cells against dsRNA[38]. As shown in **Figure 6A and Figure 6B**, IFN-β mRNA
215 and protein were upregulated to varying degrees in RAW264.7 cells. The RNA synthesized by the wild-type T7 RNAP
216 exhibited the most robust immune response, whereas the RNA derived from the mutants displayed a muted response.
217 Specifically, the IFN-β mRNA of cells treated by Mut11 RNA was only 9.7% of the wild-type treated cells, and its
218 protein was 12.93 pg/mL. With the exception of Mut11, which did not fully align with *in vitro* dsRNA quantification,
219 the remaining mutants' dsRNA content matched the cellular response, highlighting their effectiveness in reducing
220 dsRNA **(Figure 5 and Figure 6)**. Moreover, the expression of EGFP showed no significant differences in quantity or
221 quality among mRNAs generated by different T7 RNAPs **(Figure 6C)**, meeting the requirements for industrial
222 applications.

223 Lastly, we conducted an evaluation of the transcriptional performance of Mut17, chosen as a representative variant,
224 under varying conditions. **Figure S4** illustrates the tests performed on Mut17 with and without cap analogs, using
225 diverse DNA templates. The results consistently demonstrated significantly lower dsRNA levels with Mut17, ranging
226 from 0.18% to 1.62% of the wild-type. IVD-B is consistent with the solvent system of screening, exhibiting a dsRNA
227 content of 1.62%, which closely matched the initial test result of 1.8% **(Figure 5)**. The incorporation of co-transcription
228 capping and the use of modified nucleotides further reduced dsRNA production, with a minimum as low as 0.5pg/μg.
229 However, the introduction of CleanCap® AU cap, especially when combined with N1-Me-Pseudo UTP, can affect the
230 overall yield of RNA. Therefore, subsequent optimization might be necessary in such cases.

231 ***RDRP and Terminal transferase activity of variants***

232 To investigate the impact of mutations on reducing dsRNA, we conducted *in silico* studies on all these mutations. Five
233 polymerases, including T7 RNAP wild-type and four variants, were simulated separately with ssRNA/ssRNA and
234 dsDNA/mRNA **(Figure 7)**. The former simulated dsRNA formation, while the latter replicated T7 RNAP transcription.
235 Since T7 RNAP does not require promoter recognition during RNA primer extension[21-25], we hypothesize that it
236 adopts an elongation conformation when interacting with ssRNA/ssRNA substrates. Therefore, the dsDNA/mRNA
237 substrate from 1MSW was modified to ssRNA/ssRNA as our structural model.

We conducted a comparative analysis of protein-substrate interactions in all simulated systems. Our findings revealed that, except for Mut14 (A70Q), the variants exhibited higher binding energy with the RNA template when T7 RNAP bound to the ssRNA/ssRNA complex **(Figure 7A and Figure 7C)**. This suggests a clear indication of reduced RDRP activity in the variants, as they show a decreased tendency to utilize RNA as a template. Similarly, this may have had an impact on the terminal transferase activity of T7 RNAP[37]. In addition, we calculated the energy during the extension process of T7 RNAP. As expected, the mutants exhibited lower binding energy with DNA template compared to the wild-type **(Figure 7B and Figure 7D)**. However, there was no significant trend observed in the binding energetics with mRNA or the non-template DNA strand. Both ssRNA/ssRNA and dsRNA/mRNA systems clearly indicates the differences in template selection between the T7 RNAP wild-type and mutants.

To investigate these, we designed experiments to test the relative RDRP activity of T7 RNAP. The variants were diluted to the same transcriptional activity (i.e., DNA-dependent RNAP activity, DDRP), and then incubated with a hairpin structural chimeric substrate. As shown in **Figure 8**, under different concentrations, all 4 variants exhibited less than 50% of the fluorescence value compared to the wild-type. This signifies a lower efficiency in the formation of double-stranded products by the variants, implying a decrease in RDRP activity, likely responsible for the decrease in dsRNA content. However, it remains unclear why Mut14 shows a lower binding energy than the wild-type, while also reducing the synthesis level of dsRNA **(Figure 5 and Figure 7C)**.

Subsequently, a 3'-end heterogeneity assay was employed to characterize the terminal transfer activity of T7 RNAP. We first transcribed RNA used mutants and wild-type T7 RNAP with DNA templates lacking a poly(A) sequence. Then, the *E. coli* poly(A) polymerase was utilized to add a poly(A) tail to the RNA. After reverse transcription, the nucleotide composition at the 3'-end was analyzed using NGS. **Figure 9** shows that all five T7 RNAPs produced a significant proportion of non-target RNA (n>0 or n<0). The most abundant among these were RNAs with an additional one or two nucleotides at the end (wild-type: n+2, others: n+1) **(Figure S5)**, confirming the strong terminal transferase activity of T7 RNAP.

When focusing on the proportion of RNA with n>0, which reflects the terminal transferase activity, we found that these RNAs accounted for 82.94% in the wild-type, while the mutants exhibited the following percentages: Mut11 (70.29%), Mut7 (67.88%), Mut14 (63.31%), and Mut17 (55.62%) **(Figure 9)**. Importantly, these percentages are highly consistent with the abundance of dsRNA in the products (**Figure 5**). These results indicate a significant reduction in terminal transferase activity of mutants, which, together with the decrease in RDRP activity, contributes to the decrease in dsRNA. Specifically, the terminal transferase activity may play a more crucial role, as evidenced by the lowest accumulation of n>0 RNAs observed in Mut17, which also exhibits the lowest dsRNA production **(Figure 5 and Figure 9)**. We evaluated the major RNAs (**Figure S6**: n-1 to n+3) in the IVT products as an indicator of RNA integrity. The mutants' major RNA showed a slightly lower percentage compared to the wild-type, attributed to an increase in n-10 and n-2 RNAs. However, as demonstrated in **Figure 6C**, minor losses of nucleotides likely have little effect on mRNA translation efficiency, potentially attributed to the UTR and poly(A) tail.

Considered that experimental validation establishes a significant correlation between RDRP and terminal transfer activities with the dsRNA content, screening strategies based on these activities are expected to provide new tools for further evolution of T7 RNAP.

Discussion

In our study, we utilized a combination of FADS-based directed evolution and computer-aided semi-rational design, and obtained several T7 RNAP variants with significantly reduced dsRNA. Among these variants, Mut17 (A70Q/F162S/K180E) exhibited the lowest dsRNA content in the products. When co-transcription capping was performed using CleanCap® AU cap and N1-Me-Pseudo UTP, the dsRNA levels reached 0.5 pg/μg. The testing of RDRP and terminal transferase activity revealed the mechanism behind the decreased dsRNA in the variants' products, as both activities closely associated with dsRNA formation demonstrated a significant decrease[35]. Specifically, the decrease in terminal transferase activity of variants appears to be the primary factor contributing to the reduction of dsRNA. This is supported by the strong positive correlation observed between this activity and dsRNA production among the different mutants **(Figure 5 and Figure 9)**. Based on the presented results, screening strategies focusing on the terminal transferase or RDRP activity would be a hopeful choice for further evolution of T7 RNAP.

Residue A70 belongs to the transition region between helix C and helix D, in close proximity to helix J and helix F. It is precisely positioned within the RNA exit channel constituted by Helix D, Helix J, and the specificity loop **(Figure 3A and 3B)**. The Mut14 mutation likely stabilizes the C/D helices, thus stabilizing the elongation conformation complex. Interestingly, Mut14 (A70Q) exhibits lower binding energy to RNA templates, potentially enhancing RDRP activity. However, in practice, both the RDRP and terminal transferase activities of Mut14 are reduced **(Figure 8 and Figure 9)**, which contributes to the decrease of dsRNA during IVT. This may imply the presence of more intricate factors influencing the RDRP activity of T7 RNAP. F162 and K180 reside on two helices of subdomain H, contributing to the stability of the transcription bubble during transcription **(Figure 3)**. While A247T is located in the transition region between the intercalating hairpin and J-helix. These three residues do not make direct contact with the template chain, yet simulations reveal stronger DNA template binding and weaker RNA template binding compared to the wild-type. This indicates that residues lacking direct interactions with the template can also alter the preference of T7 RNAP for the template. In an IVT system, there exists a competitive relationship between the transcriptional activity and the RDRP activity of the polymerase. When the polymerase prefers to bind to dsDNA/mRNA substrates, the activity of RDRP is weakened. Therefore, either enhancing the binding to DNA templates or reducing the binding to RNA templates is conducive to reducing the production of dsRNA during the *in vitro* transcription.

Before the design of the FADS procedure, we noted that even the wild-type T7 RNAP produced a minimal amount of dsRNA **(Figure 5)**, which significantly impedes FADS screening due to the difficulty in enriching candidates based on such minimal quantitative variations. However, in non-denaturing gel electrophoresis, we observed that dsRNA comprised about one-third to one-half of total RNA after IVT, identify by the acridine orange[39]. We concluded that FADS and this electrophoresis identify all RNA molecules containing complementary regions, which are present in a

considerable quantity. The reduction of such RNAs caused by the decrease in dsRNA is sufficient to distinguish mutants from the wild-type. While J2 antibodies recognize and bind to the complementary regions in RNA, and the more complementary bases there are, the more antibodies bind.

On the other hand, the molecular beacon method was initially designed for real-time monitoring of T7 RNAP transcription, where the fluorescence reflects RNA yield[34]. However, during our FADS screening, low-yield variants, such as Mut2 and Mut9, were unexpectedly enriched. We hypothesize that when using long templates, specifically 4k or 9k in our study, the fluorescence generated by beacons targeting the 3'-end reflects only the amount of run-off RNA, not the total RNA yield. Consequently, variants with lower yields but improved integrity or reduced dsRNA levels, such as Mut2 and Mut9, have the potential to be identified **(Figure 2)**. Fortunately, the dual-probe screening has successfully addressed this limitation, ensuring an acceptable yield of the generated mutants **(Figure 4)**. This method will also be applied to FADS by expanding the fluorescence channels of the equipment.

In the random library L6, we identified a pseudo-positive variant, Mut18, which accounted for a high proportion of 86% after two rounds of screening **(data not shown)**. Mut18 contained five mutations at positions K8N/Q232L/N233D/E242D/N488D. We observed that variant Mut18 displayed significantly reduced transcriptional (DDRP) activity, which can be attributed to mutations at positions 232/233/242 that affect the functionality of the "intercalating hairpin" during promoter recognition or template unwinding[40]. However, when incubated alone with the probe, it markedly triggers the fluorescence of FAM. This indicates that the molecular beacon can elicit fluorescence not only through its binding to the target mRNA but also via direct interaction with Mut18. We suspect that Mut18 exhibits a higher affinity for hairpin structural probes compared to the T7 promoter, leading to their unwinding and subsequent fluorescence activation. This further prompted us to develop alternative screening strategies based on the direct activity of T7 RNAP.

Table 1. Parameters for the screening of L3 and L5 by FADS.

Libraries	Lib3			Lib5		
Diversity	5.7×10^5			2.6×10^6		
RFU	**Round 1**	**Round 2**	**Mutant**	**Round 1**	**Round 2**	**Mutant**
WT [a]	< 1.2	< 2.5	**Mut1:** V214A	< 2.7	< 2.3	**Mut6:** V186I
Threshold	2.0	5.2	**Mut2:** N233D	4.1	5.7	**Mut7:** F162S/A247T
Ratio	> 1.67	> 2.08	**Mut3:** A465T **Mut4:** A615T	> 1.52	> 2.48	**Mut8:** F416L **Mut9:** L106P
Count	3539	122	**Mut5:** M369T	6793	204	**Mut10:** D87G

a. The maximum fluorescence of the wild-type serves as the threshold benchmark.

Materials and methods

Oligonucleotides and reagents

Synthetic oligonucleotides were obtained from Sangon Biotechnology. NTP solutions, carbenicillin, pyrophosphatase, and Murine RNase inhibitor were acquired from Yeasen Biotechnology. Anhydrotetracycline, inorganic salt, and

333 glycerol were purchase from Sigma-Aldrich.

334 **Beacon testing applied to artificial dsRNA**

335 A total of 30 pM of artificial dsRNA **(Figure 1A)** was mixed with 500 nM molecular beacon and 2 μL of 10 ×
336 Transcription Buffer (CAT#: 10670ES, Yeasen Biotechnology). The volume was then adjusted to 20 μL with ddH_2O.
337 The mixture was incubated at 95°C for 5 minutes and placed on ice for 10 minutes. Fluorescence was measured using
338 a SpectraMax iD3 Microplate Detection System (Molecular Devices) with excitation/emission wavelengths of
339 480/525nm (Cutoff 520nm).

340 **FADS**

341 To construct a random library, error-prone PCR was employed as previously described[41]. The Mn^{2+} concentration
342 was 0.1 mM, and *E. coli* DH10B-Plus (CAT#: DE1072, Weidi Biotechnology) was chosen as the host. After overnight
343 culture, the library was inoculated into fresh LB medium at a 2% ratio and incubated at 37°C for 2 hours. To induce
344 the expression of T7 RNAP, 200 ng/mL anhydrotetracycline was added to the culture, followed by further incubation
345 at 16°C for 16 hours. After harvesting, the pellet corresponding to 1OD was washed with 1 × Transcription Buffer and
346 resuspended. For liquid phase A, a mixture was prepared by combining 5 μL of the bacterial suspension with 95 μL of
347 1 × Transcription Buffer and 20 mM NTPs. Liquid phase B was prepared by dissolving 1 μg DNA template, 1 μM
348 beacon, and 0.2 mg/mL lysozyme in 100 μL of 1 × Transcription Buffer. Using the DREM cell droplet generation and
349 sorting system (TmaxTree Biotechnology), liquid phases A and B were mixed in a 1:1 ratio on a PDMS chip,
350 subsequently generating droplets with an approximate diameter of 25 μm. The same procedure was employed to
351 prepare droplets containing *E. coli* expressing wild-type T7 RNAP. Both sets of droplets were incubated at 37°C for
352 2.5 hours to ensure complete cell lysis and in-droplet transcription reactions. The droplets containing the library were
353 sorted and positive droplets were collected. The sorting threshold was set as 1.5-2.5 times the maximum fluorescence
354 intensity observed in the wild-type droplets **(Table 1)**. A double volume of emulsion breaker (Drop-breaker, FluidicLab)
355 was added to facilitate phase separation. The upper aqueous phase was then extracted for PCR amplification, and the
356 products were used for sequencing and construction of a new library.

357 **Purification of T7 RNAP and variants**

358 The plasmids were separately transformed into *E. coli* BL21 (CAT#: EC1001, Weidi Biotechnology), and cells were
359 collected after protein expression. T7 RNAP and the variants were purified using Ni-NTA chromatography in 20 mM
360 Tris-HCl pH 7.5, 100 mM NaCl, 10% Glycerol, and 10 mM Imidazole. The purified samples were then dialyzed
361 overnight against 100 mM Tris-HCl pH 7.9, 100 mM NaCl, 2 mM EDTA, 10 mM DTT, 0.1% Triton X-100, and 50%
362 Glycerol. In the microplate screening, the same purification procedure is used, but with a reduced volume of around 4
363 mL.

364 ***In vitro* transcription assays**

365 Gradient dilutions of variants, wild-type, and control (CAT#: 10626ES, Yeasen Biotechnology) T7 RNAPs were used
366 to assess relative activity based on their *in vitro* transcription RNA yield. The IVT of T7 RNAP was performed

367 following the procedure of the T7 High Yield RNA Synthesis Kit (CAT#: 10623ES, Yeasen Biotechnology). Briefly, a
368 mixture of T7 RNAP, 1 μg DNA template, 10 mM NTPs, and 2 μL of 10× Transcription Buffer was prepared, and the
369 volume was adjusted to 20 μL with DEPC-treated water. The reaction mixture was incubated at 37°C for 2 hours, and
370 the RNA was precipitated using LiCl. The pellet was washed twice with pre-chilled 70% ethanol and subsequently
371 dissolved in DEPC-treated water. The resuspended RNA was quantified using a NanoDrop™ One (ThermoFisher
372 Scientific). 250 U of T7 RNAP were added typically for IVT, aimed at comparing the performance of mutants. For co-
373 transcription capping experiments, an additional 10 mM of CleanCap® Reagent AU (CAT#: N-7114, TriLink
374 Biotechnology) or CleanCap® Reagent AG (3'OMe) (CAT#: N-7413, TriLink Biotechnology) analog was added.

375 **dsRNA immunoblots**

376 A total of 2.4 μg of RNA from each variant were spotted onto Nytran SuPerCharge (SPC) blotting membranes (Cat#:
377 10416216, Cytiva) and dried at 80°C for 40 minutes. The membranes were subsequently blocked in TBS-T buffer (20
378 mM Tris, pH 7.4, 150 mM NaCl, 0.1% (v/v) Tween-20) containing 5% (w/v) non-fat dried milk. For the detection of
379 dsRNA, the membranes were incubated with Mouse monoclonal antibody J2 (Cat#: ab288755, Abcam) at a dilution
380 of 1/1000 for 1 hour at room temperature. After washing 4-6 times, the membranes were immersed in TBS-T buffer
381 containing Peroxidase-conjugated Donkey Anti-Mouse IgG (H+L) at a dilution of 1/2000 and incubated for 1 hour at
382 room temperature. Following another round of washing (4-6 times), Super ECL Detection Reagent (CAT#: 36208ES,
383 Yeasen Biotechnology) was used for visualization, and the images were recorded using the ChemiDoc MP Imaging
384 System (Bio-Rad).

385 **Semi-rational design**

386 The structure of initiation complex from a crystal structure (PDB: 1QLN), and elongation complex from a crystal
387 structure (PDB: 1MSW). We removed waters and repaired the missing gaps (residues 1-5, 56-71 in the 1QLN and
388 residues 1, 233-240, 364-374 in the 1MSW) used Modeller[42,43]. The free energy differences between the folded and
389 unfolded structures of both the initiation and elongation conformations were predicted using a global PositionScan
390 with FoldX[44], and these differences were compared to those of the wild-type (ΔΔG). Additionally, PROSS was used
391 to predict mutations that are beneficial to the stability of the elongation conformation[45]. Structural analysis was
392 performed using the visualization software PyMol[46].

393 **dsRNA ELISA**

394 To quantify the dsRNA in the IVT products, a Double-stranded RNA (dsRNA) ELISA kit (CAT#: 36717ES, Yeasen
395 Biotechnology) was used for testing the RNA samples. Briefly, standard substance and RNA samples were diluted to
396 appropriate concentrations. 100 μL of each dilution was added to the reaction wells and incubated at room temperature
397 for 1 hour. After discarding the supernatant, 100 μL of 1× Detection Antibody was added to each well and incubated
398 at room temperature for another 1 hour. The samples were then discarded and the wells were washed. 100 μL of 1×
399 Streptavidin-HRP was added to each well and incubated at room temperature for 30 minutes. TMB Substrate and Stop
400 Solution were subsequently added, mixed, and the absorbance values were measured using the SpectraMax iD3
401 Microplate Detection System (Molecular Devices), under the condition of Ex = 450nm / Em = 650nm. A standard

curve was generated to calculate the dsRNA content of each sample.

Molecular dynamics simulation

The three-dimensional structure of the wild-type T7 RNAP was obtained by repairing the missing amino acids in the crystal structure (PDB: 1MSW) using the Modeller software. The mutations of T7 RNAP was obtained by mutating the wild-type structure using BIOVIA-Discovery Studio Visualizer (BIOVIA Corp). The dsDNA/mRNA substrate was derived from the ternary complex substrate of the crystal structure (PDB: 1MSW). The structure of the ssRNA/ssRNA was modified from the dsDNA/mRNA substrate with BIOVIA-Discovery Studio Visualizer: fist remove the non-template DNA, and then modified the DNA template to RNA strand, while retaining the 10 nucleotides at the 3'-end. The final ssRNA/ssRNA had a template strand of 5'-UUCGCCGUGU-3' and a primer strand of 5'-GACACGGCGA-3'. All molecular dynamics simulations were performed using the GROMACS-2022.1 software package[47-50], using the AMBER99SB force field with PARMBSC0 nucleic acid parameters and the Tip3p water model[51-53]. The steepest descent algorithm was used for energy minimization, and the Particle-Mesh Ewald method was used to treat long-range electrostatic interactions[54]. The ion concentration to neutralize the system were maintained 0.15M NaCl. The cutoff distance of van der Waals force (vdW) and short-range electrostatic interaction was set to 10 Å. The system temperature was maintained at 310K, with the neighbor list updated every 10 steps, a time step length of 2fs, and a MD simulation of 50ns.The interactions between protein and substrate were calculated from simulated trajectories using the gmx energy module in Gromacs. The interactions between protein and substrate energy includes Van der Waals forces and electrostatic interactions. The ele and vdW interaction energies were calculated at a cut-off distance of 40 Å.

RDRP activity

T7 RNAP and variants were diluted to the following concentrations: 125 U/μL, 62.5 U/μL, 31.25 U/μL, respectively, based on their activities (DDRP activity). For each dilution, 1 μL of the enzyme was added to the previously mentioned TVT system. A single-stranded chimeric nucleotide chain (5'-AAAAAAAAAAAAAAAAAAAAdCdCdCdCdCUUUUUUUUUU-3') at a concentration of 1.25 μM was used as the transcription template, which forms a stable hairpin structure after incubation at 65°C for 2 minutes followed by 4°C for 2 minutes. During a 10-minute incubation at 37°C, the T7 RNAP extended the 3'-end of the chimeric nucleotide chain, resulting in the formation of a complete double-stranded product. The reaction was then quenched by incubating at 65°C for 5 minutes. According to the protocol of Picogreen dsDNA Quantitation Reagent (Cat#: 12641ES, Yeasen Biotechnology), a double-stranded dye was added to each reaction well. The fluorescence was measured using the SpectraMax iD3 Microplate Detection System (Molecular Devices) with excitation at 450nm and emission at 650nm to determine the relative RDRP activity for each variant.

Heterogeneity analysis of the RNA 3'-end

To add a poly(A) tail to the 3'-end of RNA, *E. coli* poly(A) Polymerase (CAT#: 14801ES, Yeasen Biotechnology) and 1mM ATP was added to the samples and incubated at 37°C for 30 minutes. Tailed products were then subjected to reverse transcription using M-MLV (H-) Reverse Transcriptase (Cat#: 11300ES, Yeasen Biotechnology) with a poly(T)

primer, followed by PCR amplification of cDNA for sequencing.

Transfection of cells

HEK293 and murine RAW264.7 cells were plated in Culture Flask (Cat#: 84022ES, Yeasen Biotechnology) and incubated for 36 h. After treatment with trypsin-EDTA, the cells plated in 24-well plates at a density of 0.5×10^5 cells per well. Following the Lipofectamine 3000 Reagent (ThermoFisher Scientific) protocol, cells were grown to approximately 90% confluence before transfection. For transfection, 2 μg of capped RNA transcribed from different variants was dissolved in 50 μL of Opti-MEM Medium along with 4 μL of P3000 reagent. The solution was thoroughly mixed with diluted Lipofectamine 3000 Reagent at a 1:1 ratio and incubated at room temperature for 15 minutes. Subsequently, 50 μL volumes of different lipofectamine/RNA solutions were added to the wells of the three kinds of cells. The cells were then incubated at 37°C for 24 hours. After incubation, the cells were washed with PBS and released by treatment with 0.5% Trypsin-EDTA. The HEK293 cells were used for EGFP fluorescence analysis, while murine RAW264.7 cells were utilized for quantification of IFN-β mRNA and protein.

Reporter cell line immunogenicity assays

The immunogenicity of RNA from each variant was tested in murine RAW264.7 cells, as previously described[36]. The qPCR primers for mouse β-actin and IFN-β are as follows: Mactin-F: 5'-AACAGTCCGCCTAGAAGCAC-3', Mactin-R: 5'-CGTTGACATCCGTAAAGACC-3'; MIFN-F: 5'-GCCTTTGCCATCCAAGAGATGC-3', MIFN-R: 5'-ACACTGTCTGCTGGTGGAGTTC-3'. The quantification of IFN-β protein was performed using the IFN-β/IFNB ELISA Kit (Cat#: CSB-E04945m, Huamei Biotechnology), following the manufacturer's protocol.

References

[1] Huang T, Peng L, Han Y, Wang D, He X, Wang J, Ou C. Lipid nanoparticle-based mRNA vaccines in cancers: Current advances and future prospects. Front Immunol. 2022 Aug 26;13:922301. doi: 10.3389/fimmu.2022.922301. PMID: 36090974; PMCID: PMC9458914.

[2] Verbeke R, Lentacker I, De Smedt SC, Dewitte H. The dawn of mRNA vaccines: The COVID-19 case. J Control Release. 2021 May 10;333:511-520. doi: 10.1016/j.jconrel.2021.03.043. Epub 2021 Mar 30. PMID: 33798667; PMCID: PMC8008785.

[3] Petsch B, Schnee M, Vogel AB, Lange E, Hoffmann B, Voss D, Schlake T, Thess A, Kallen KJ, Stitz L, Kramps T. Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection. Nat Biotechnol. 2012 Dec;30(12):1210-6. doi: 10.1038/nbt.2436. Epub 2012 Nov 25. PMID: 23159882.

[4] Schlake T, Thess A, Fotin-Mleczek M, Kallen KJ. Developing mRNA-vaccine technologies. RNA Biol. 2012 Nov;9(11):1319-30. doi: 10.4161/rna.22269. Epub 2012 Oct 12. PMID: 23064118; PMCID: PMC3597572.

[5] Van Lint S, Renmans D, Broos K, Dewitte H, Lentacker I, Heirman C, Breckpot K, Thielemans K. The ReNAissanCe of mRNA-based cancer therapy. Expert Rev Vaccines. 2015 Feb;14(2):235-51. doi: 10.1586/14760584.2015.957685. Epub 2014 Sep 29. PMID: 25263094.

[6] Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine. 2021 Apr 15;39(16):2190-2200. doi: 10.1016/j.vaccine.2021.03.038. Epub 2021 Mar 24. PMID: 33771389; PMCID: PMC7987532.

[7] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines - a new era in vaccinology. Nat Rev Drug Discov. 2018 Apr;17(4):261-279. doi: 10.1038/nrd.2017.243. Epub 2018 Jan 12. PMID: 29326426; PMCID: PMC5906799.

[8] Karikó K, Muramatsu H, Ludwig J, Weissman D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. Nucleic Acids Res. 2011 Nov;39(21):e142. doi: 10.1093/nar/gkr695. Epub 2011 Sep 2. PMID: 21890902; PMCID: PMC3241667.

[9] de Haro C, Méndez R, Santoyo J. The eIF-2alpha kinases and the control of protein synthesis. FASEB J. 1996 Oct;10(12):1378-87. doi: 10.1096/fasebj.10.12.8903508. PMID: 8903508.

[10] Liang SL, Quirk D, Zhou A. RNase L: its biological roles and regulation. IUBMB Life. 2006 Sep;58(9):508-14. doi: 10.1080/15216540600838232. PMID: 17002978.

[11] Yin YW, Steitz TA. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. Science. 2002 Nov 15;298(5597):1387-95. doi: 10.1126/science.1077464. Epub 2002 Sep 19. PMID: 12242451.

[12] Cheetham GM, Jeruzalmi D, Steitz TA. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature. 1999 May 6;399(6731):80-3. doi: 10.1038/19999. Erratum in: Nature 1999 Jul 1;400(6739):89. PMID: 10331394.

[13] Cheetham GM, Steitz TA. Structure of a transcribing T7 RNA polymerase initiation complex. Science. 1999 Dec 17;286(5448):2305-9. doi: 10.1126/science.286.5448.2305. PMID: 10600732.

[14] Durniak KJ, Bailey S, Steitz TA. The structure of a transcribing T7 RNA polymerase in transition from initiation

to elongation. Science. 2008 Oct 24;322(5901):553-7. doi: 10.1126/science.1163433. PMID: 18948533; PMCID: PMC2892258.

[15] Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res. 1987 Nov 11;15(21):8783-98. doi: 10.1093/nar/15.21.8783. PMID: 3684574; PMCID: PMC306405.

[16] Martin CT, Muller DK, Coleman JE. Processivity in early stages of transcription by T7 RNA polymerase. Biochemistry. 1988 May 31;27(11):3966-74. doi: 10.1021/bi00411a012. PMID: 3415967.

[17] Ling ML, Risman SS, Klement JF, McGraw N, McAllister WT. Abortive initiation by bacteriophage T3 and T7 RNA polymerases under conditions of limiting substrate. Nucleic Acids Res. 1989 Feb 25;17(4):1605-18. doi: 10.1093/nar/17.4.1605. Erratum in: Nucleic Acids Res 1989 Jun 12;17(11):4430. PMID: 2646596; PMCID: PMC331825.

[18] Song H, Kang C. Sequence-specific termination by T7 RNA polymerase requires formation of paused conformation prior to the point of RNA release. Genes Cells. 2001 Apr;6(4):291-301. doi: 10.1046/j.1365-2443.2001.00420.x. PMID: 11318872.

[19] Macdonald LE, Durbin RK, Dunn JJ, McAllister WT. Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. J Mol Biol. 1994 Apr 29;238(2):145-58. doi: 10.1006/jmbi.1994.1277. PMID: 8158645.

[20] Lyakhov DL, He B, Zhang X, Studier FW, Dunn JJ, McAllister WT. Pausing and termination by bacteriophage T7 RNA polymerase. J Mol Biol. 1998 Jul 10;280(2):201-13. doi: 10.1006/jmbi.1998.1854. PMID: 9654445.

[21] Gholamalipour Y, Karunanayake Mudiyanselage A, Martin CT. 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character-RNA-Seq analyses. Nucleic Acids Res. 2018 Oct 12;46(18):9253-9263. doi: 10.1093/nar/gky796. PMID: 30219859; PMCID: PMC6182178.

[22] Triana-Alonso FJ, Dabrowski M, Wadzack J, Nierhaus KH. Self-coded 3'-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase. J Biol Chem. 1995 Mar 17;270(11):6298-307. doi: 10.1074/jbc.270.11.6298. PMID: 7534310.

[23] Cazenave C, Uhlenbeck OC. RNA template-directed RNA synthesis by T7 RNA polymerase. Proc Natl Acad Sci U S A. 1994 Jul 19;91(15):6972-6. doi: 10.1073/pnas.91.15.6972. PMID: 7518923; PMCID: PMC44320.

[24] Arnaud-Barbe N, Cheynet-Sauvion V, Oriol G, Mandrand B, Mallet F. Transcription of RNA templates by T7 RNA polymerase. Nucleic Acids Res. 1998 Aug 1;26(15):3550-4. doi: 10.1093/nar/26.15.3550. PMID: 9671817; PMCID: PMC147742.

[25] Nacheva GA, Berzal-Herranz A. Preventing nondesired RNA-primed RNA extension catalyzed by T7 RNA polymerase. Eur J Biochem. 2003 Apr;270(7):1458-65. doi: 10.1046/j.1432-1033.2003.03510.x. PMID: 12654001.

[26] Markel U, Essani KD, Besirlioglu V, Schiffels J, Streit WR, Schwaneberg U. Advances in ultrahigh-throughput screening for directed enzyme evolution. Chem Soc Rev. 2020 Jan 2;49(1):233-262. doi: 10.1039/c8cs00981c. PMID: 31815263.

[27] Vallejo D, Nikoomanzar A, Paegel BM, Chaput JC. Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. ACS Synth Biol. 2019 Jun 21;8(6):1430-1440. doi: 10.1021/acssynbio.9b00103. Epub 2019 May

23. PMID: 31120731; PMCID: PMC6942126.

[28] Longwell CK, Labanieh L, Cochran JR. High-throughput screening technologies for enzyme engineering. Curr Opin Biotechnol. 2017 Dec;48:196-202. doi: 10.1016/j.copbio.2017.05.012. Epub 2017 Jun 15. PMID: 28624724.

[29] Neun S, Kaminski TS, Hollfelder F. Single-cell activity screening in microfluidic droplets. Methods Enzymol. 2019;628:95-112. doi: 10.1016/bs.mie.2019.07.009. Epub 2019 Aug 8. PMID: 31668237.

[30] Gantz M, Neun S, Medcalf EJ, van Vliet LD, Hollfelder F. Ultrahigh-Throughput Enzyme Engineering and Discovery in In Vitro Compartments. Chem Rev. 2023 May 10;123(9):5571-5611. doi: 10.1021/acs.chemrev.2c00910. Epub 2023 May 1. PMID: 37126602; PMCID: PMC10176489.

[31] Bunzel HA, Garrabou X, Pott M, Hilvert D. Speeding up enzyme discovery and engineering with ultrahigh-throughput methods. Curr Opin Struct Biol. 2018 Feb;48:149-156. doi: 10.1016/j.sbi.2017.12.010. Epub 2018 Feb 3. PMID: 29413955.

[32] Gantz M, Aleku GA, Hollfelder F. Ultrahigh-throughput screening in microfluidic droplets: a faster route to new enzymes. Trends Biochem Sci. 2022 May;47(5):451-452. doi: 10.1016/j.tibs.2021.11.001. Epub 2021 Nov 27. PMID: 34848125.

[33] Holstein JM, Gylstorff C, Hollfelder F. Cell-free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. ACS Synth Biol. 2021 Feb 19;10(2):252-257. doi: 10.1021/acssynbio.0c00538. Epub 2021 Jan 27. PMID: 33502841; PMCID: PMC7901014.

[34] Marras SA, Gold B, Kramer FR, Smith I, Tyagi S. Real-time measurement of in vitro transcription. Nucleic Acids Res. 2004 May 20;32(9):e72. doi: 10.1093/nar/gnh068. PMID: 15155820; PMCID: PMC419623.

[35] Wu MZ, Asahara H, Tzertzinis G, Roy B. Synthesis of low immunogenicity RNA with high-temperature in vitro transcription. RNA. 2020 Mar;26(3):345-360. doi: 10.1261/rna.073858.119. Epub 2020 Jan 3. PMID: 31900329; PMCID: PMC7025508.

[36] Dousis A, Ravichandran K, Hobert EM, Moore MJ, Rabideau AE. An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. Nat Biotechnol. 2023 Apr;41(4):560-568. doi: 10.1038/s41587-022-01525-6. Epub 2022 Nov 10. PMID: 36357718; PMCID: PMC10110463.

[37] Wu H, Wei T, Yu B, Cheng R, Huang F, Lu X, Yan Y, Wang X, Liu C, Zhu B. A single mutation attenuates both the transcription termination and RNA-dependent RNA polymerase activity of T7 RNA polymerase. RNA Biol. 2021 Oct 15;18(sup1):451-466. doi: 10.1080/15476286.2021.1954808. Epub 2021 Jul 27. PMID: 34314299; PMCID: PMC8677023.

[38] Mu X, Hur S. Immunogenicity of In Vitro-Transcribed RNA. Acc Chem Res. 2021 Nov 2;54(21):4012-4023. doi: 10.1021/acs.accounts.1c00521. Epub 2021 Oct 22. PMID: 34677064; PMCID: PMC9127547.

[39] Mu X, Greenwald E, Ahmad S, Hur S. An origin of the immunogenicity of in vitro transcribed RNA. Nucleic Acids Res. 2018 Jun 1;46(10):5239-5249. doi: 10.1093/nar/gky177. PMID: 29534222; PMCID: PMC6007322.

[40] Borkotoky S, Murali A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. Int J Biol Macromol. 2018 Oct 15;118(Pt A):49-56. doi: 10.1016/j.ijbiomac.2018.05.198. Epub 2018 May 27. PMID: 29847782.

[41] Cirino PC, Mayer KM, Umeno D. Generating mutant libraries using error-prone PCR. Methods Mol Biol. 2003;231:3-9. doi: 10.1385/1-59259-395-X:3. PMID: 12824595.

[42] Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Protein Sci. 2000 Sep;9(9):1753-73. doi: 10.1110/ps.9.9.1753. PMID: 11045621; PMCID: PMC2144714.

[43] Webb B, Sali A. Protein Structure Modeling with MODELLER. Methods Mol Biol. 2017;1654:39-54. doi: 10.1007/978-1-4939-7231-9_4. PMID: 28986782.

[44] Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol. 2002 Jul 5;320(2):369-87. doi: 10.1016/S0022-2836(02)00442-4. PMID: 12079393.

[45] Weinstein JJ, Goldenzweig A, Hoch S, Fleishman SJ. PROSS 2: a new server for the design of stable and highly expressed protein variants. Bioinformatics. 2021 Apr 9;37(1):123-125. doi: 10.1093/bioinformatics/btaa1071. PMID: 33367682; PMCID: PMC7611707.

[46] Delano, Warren L. “The PyMOL Molecular Graphics System.” (2002).

[47] Páll S, Zhmurov A, Bauer P, Abraham M, Lundborg M, Gray A, Hess B, Lindahl E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. J Chem Phys. 2020 Oct 7;153(13):134110. doi: 10.1063/5.0018516. PMID: 33032406.

[48] Abraham, Mark James, Murtola, Teemu, Schulz, Roland, Páll, Szilárd, Smith, Jeremy C., Hess, Berk, and Lindahl, Erik. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. United States: N. p., 2015. Web. doi:10.1016/j.softx.2015.06.001.

[49] Páll, S., Abraham, M.J., Kutzner, C., Hess, B., Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In: Markidis, S., Laure, E. (eds) Solving Software Challenges for Exascale. EASC 2014. Lecture Notes in Computer Science(), vol 8759. Springer, Cham. https://doi.org/10.1007/978-3-319-15976-8_1

[50] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, Erik Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics, Volume 29, Issue 7, April 2013, Pages 845–854, https://doi.org/10.1093/bioinformatics/btt055.

[51] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010 Jun;78(8):1950-8. doi: 10.1002/prot.22711. PMID: 20408171; PMCID: PMC2970904.

[52] DePaul AJ, Thompson EJ, Patel SS, Haldeman K, Sorin EJ. Equilibrium conformational dynamics in an RNA tetraloop from massively parallel molecular dynamics. Nucleic Acids Res. 2010 Aug;38(14):4856-67. doi: 10.1093/nar/gkq134. Epub 2010 Mar 11. PMID: 20223768; PMCID: PMC2919701.

[53] Sorin EJ, Pande VS. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophys J. 2005 Apr;88(4):2472-93. doi: 10.1529/biophysj.104.051938. Epub 2005 Jan 21. PMID: 15665128; PMCID: PMC1305346.

[54] Essmann, Ulrich et al. “A smooth particle mesh Ewald method.” Journal of Chemical Physics 103 (1995): 8577-8593.

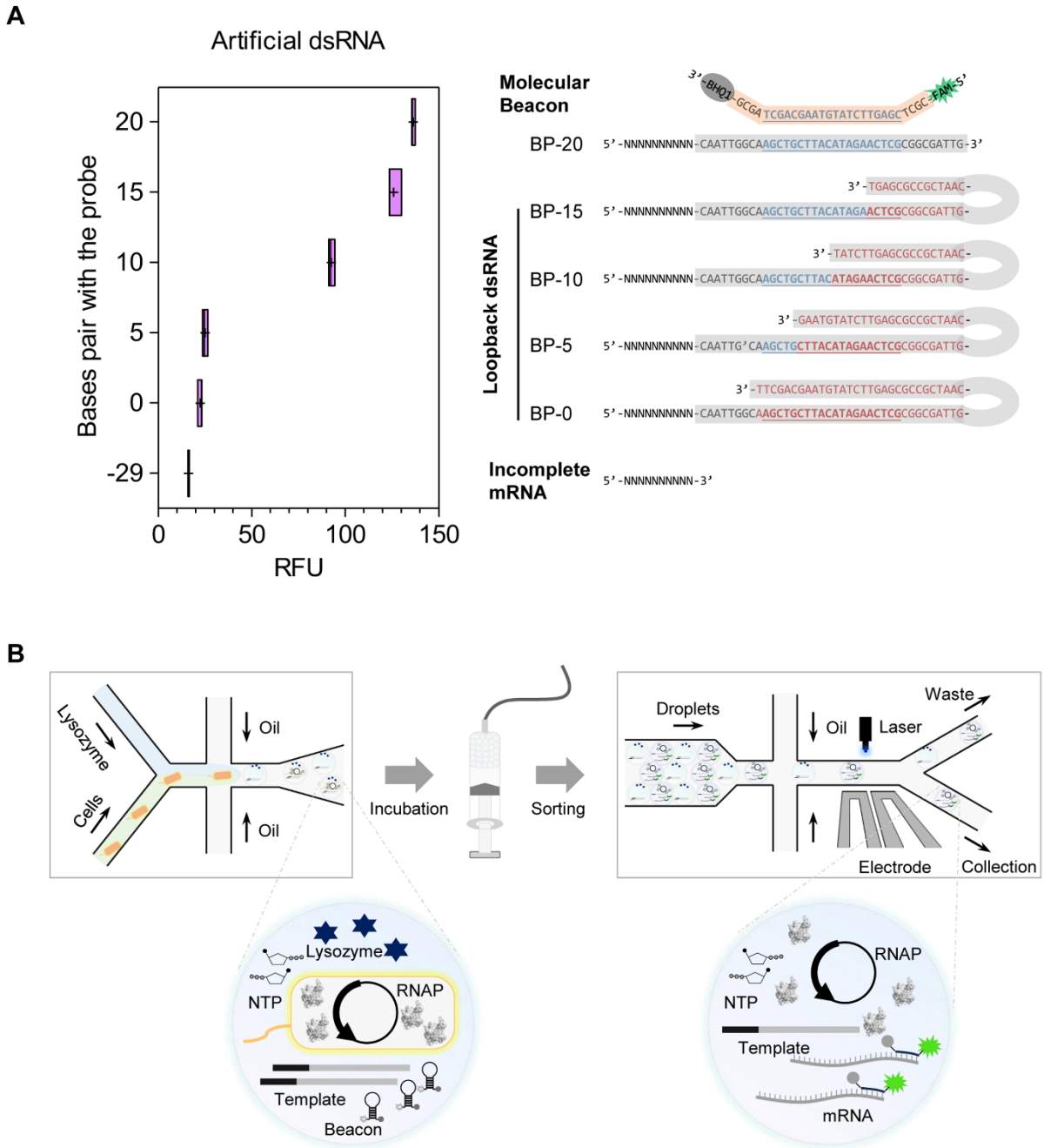


Figure 1. Overview of the FADS:

A. Relative fluorescence units (RFU) of the molecular beacon binding with ssDNA. Equimolar single-stranded DNA was denatured, annealed, and incubated with molecular beacons at 37°C for 10 minutes. Then the fluorescence was measured (n=12). BP-20 represents the run-off RNA, which serves as the target product. BP-0, BP-5, BP-10, and BP-15 are designed to simulate dsRNA, where the numbers indicate the complementary bases on ssDNA that bind to the molecular beacon. "+" represents the mean value.

B. Generation and sorting of droplets. The droplets contain all the necessary components for IVT, along with lysozyme and beacon. A dual aqueous phases chip channel was used to prevent premature cell lysis. Droplets were incubated in a syringe to facilitate lysozyme-mediated cell lysis and T7 RNAP-driven transcription after formation. Finally, the droplets were reinjected into a sorting chip for fluorescence detection and screening.

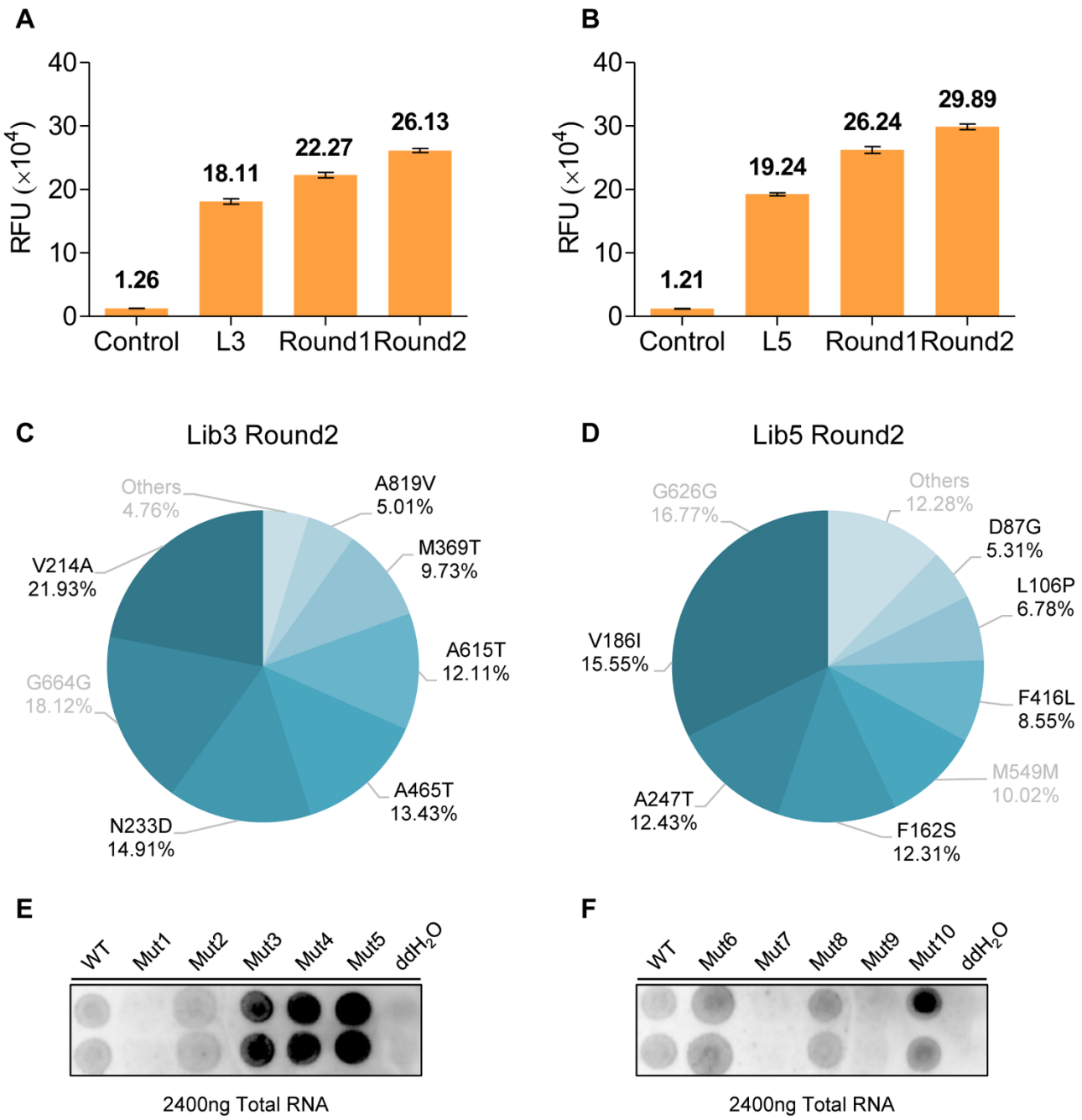


Figure 2. Random library screening and variants evaluation:

A (L3) and B (L5). RFU of IVT products from equivalent bacterial pellets. DNA from positive droplets used for next-round library (Round1 and Round2). The *E. coli* libraries cultured for protein expression and 0.05OD cells were collected. After cell disruption, the fluorescence generated by the IVT was detected (n=3).

C (L3) and D (L5). Distribution and proportions of mutations. Following two rounds of screening, DNA within the positive droplets was amplified via PCR and subsequently analyzed for sequencing using NGS. Silent mutations and mutations with frequencies below 5% are indicated in gray font.

E and F. dsRNA immunoblots of variants. The IVT products of variants were diluted and spotted onto nylon membranes, and the J2 antibody was used to identify dsRNA. Mutant information: Mut1 (V214A), Mut2 (N233D), Mut3 (A465T), Mut4 (A615T), Mut5 (M369T), Mut6 (V186I), Mut7 (F162S/A247T), Mut8 (F416L), Mut9 (L106P), and Mut10 (D87G). 5μL ddH_2O was used as the negative control.

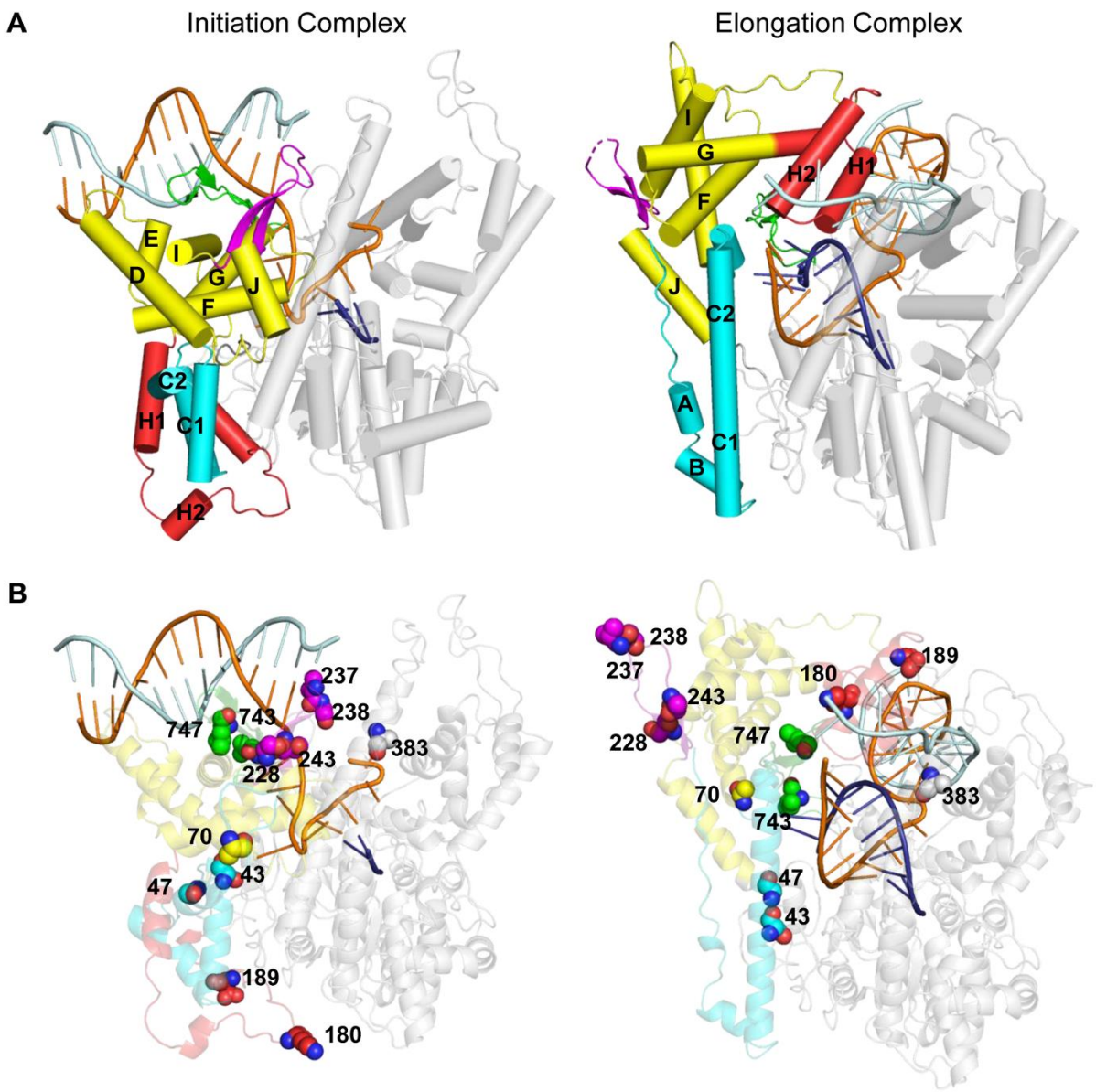


Figure 3. Structure-guided semi-rational design for decreased dsRNA:

A. Comparison of the structures of the T7 RNAP Initiation (left, PDB: 1QLN) and Elongation (right, PDB: 1MSW) complexes. The molecules have been similarly oriented by superposition of their palm domains. The C-terminal domain is shown as gray transparent cartoon to allow views of the DNA and RNA. The N-terminal domain is colored as follows: promoter binding domain is yellow, the C-helix is cyan, the subdomain H domain is red, the specificity loop is green, the intercalating hairpin opens the double DNA is magenta. The template DNA is shown in orange, non-template DNA in pale cyan and RNA in deep blue.

B. Locations of the recommended positions. Recommended positions are shown as spheres. The missing gaps in both the initiation and elongation complexes were repaired using Modeller. Proteins and substrates are colored as A. The protein is rendered as a transparent cartoon to highlight the recommended positions. The figure was generated using the PyMOL software.

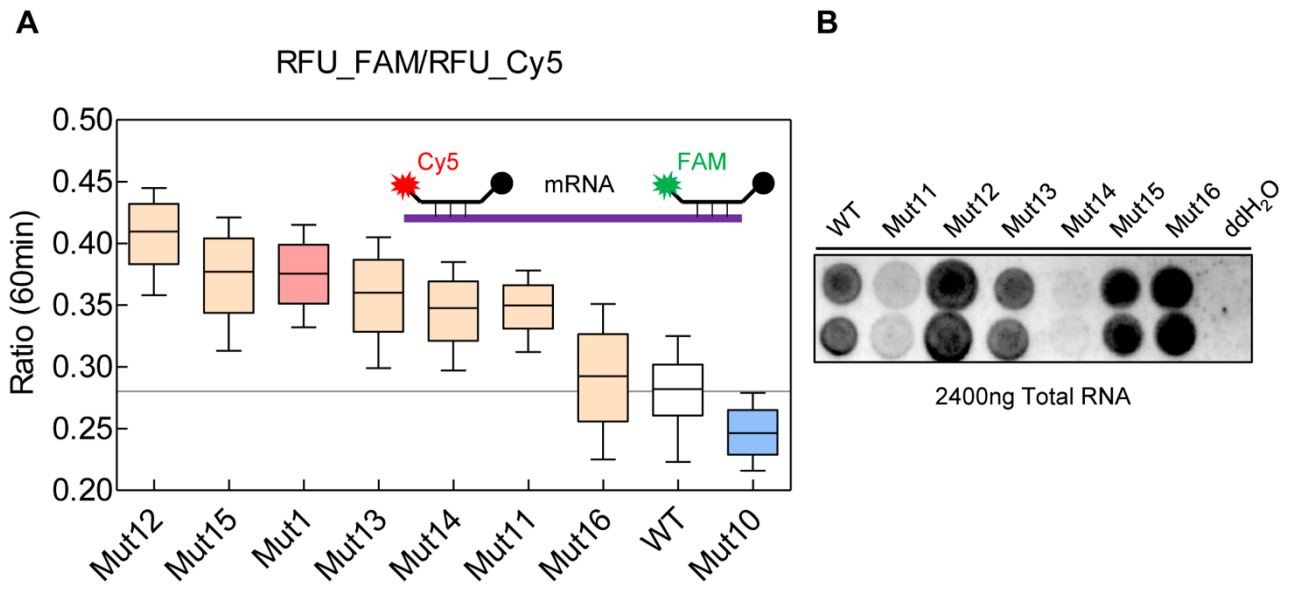


Figure 4. Microtiter plate screening and variant evaluation:

A. Dual probe detection of RNA products. The purified T7 RNAP mutants were used for IVT, and the fluorescence of Cy5 and FAM was measured after 60 minutes. After a round of primary screening, each candidate was re-inoculated into a microplate for secondary screening (n>6). The superior mutants were identified by ranking them based on the FAM/Cy5 ratio. The fluorescence ratio of Mut1 and Mut10 serves as positive and negative controls, respectively. The black line marks the wild-type ratio of 0.2802, and any values above it is considered candidate mutants. The illustrations in the top-right corner indicate the probe positions and fluorescence types, with the 3'-end FAM probe designed to screen for mutants with reduced loopback dsRNA or incomplete RNA, and the Cy5 probe allowing for analysis of RNA yield and the molar ratio between abortive RNA and run-off RNA.

B. dsRNA immunoblots of variants. The IVT products of variants were diluted and spotted onto nylon membranes, and the J2 antibody was used to identify dsRNA. Mutant information: Mut11 (K180E), Mut12 (S228F), Mut13 (G238L), Mut14 (A70Q), Mut15 (A383H), and Mut16 (I743L). 5μL ddH_2O was used as the negative control.

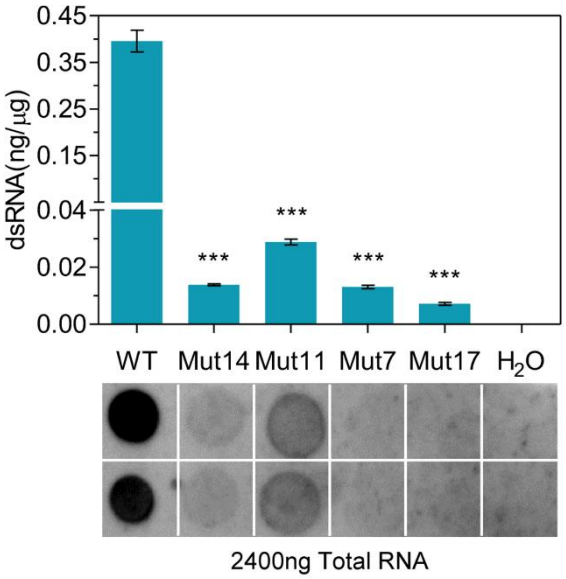


Figure 5. dsRNA content of Mut17 and its parental mutations:

Upper: Quantification of dsRNA by ELISA, comparing wild-type and mutant variants (Mut14, Mut11, Mut7, and Mut17), based on standard dsRNA (composed of natural deoxyribonucleotides). N=3. *** indicates $p < 0.0001$. Lower: Immunoblots of dsRNA variants, with 5μL ddH_2O as negative control. Mutant details: Mut14 (A70Q), Mut11 (K180E), Mut7 (F162S/A247T), Mut17 (A70Q/F162S/K180E).

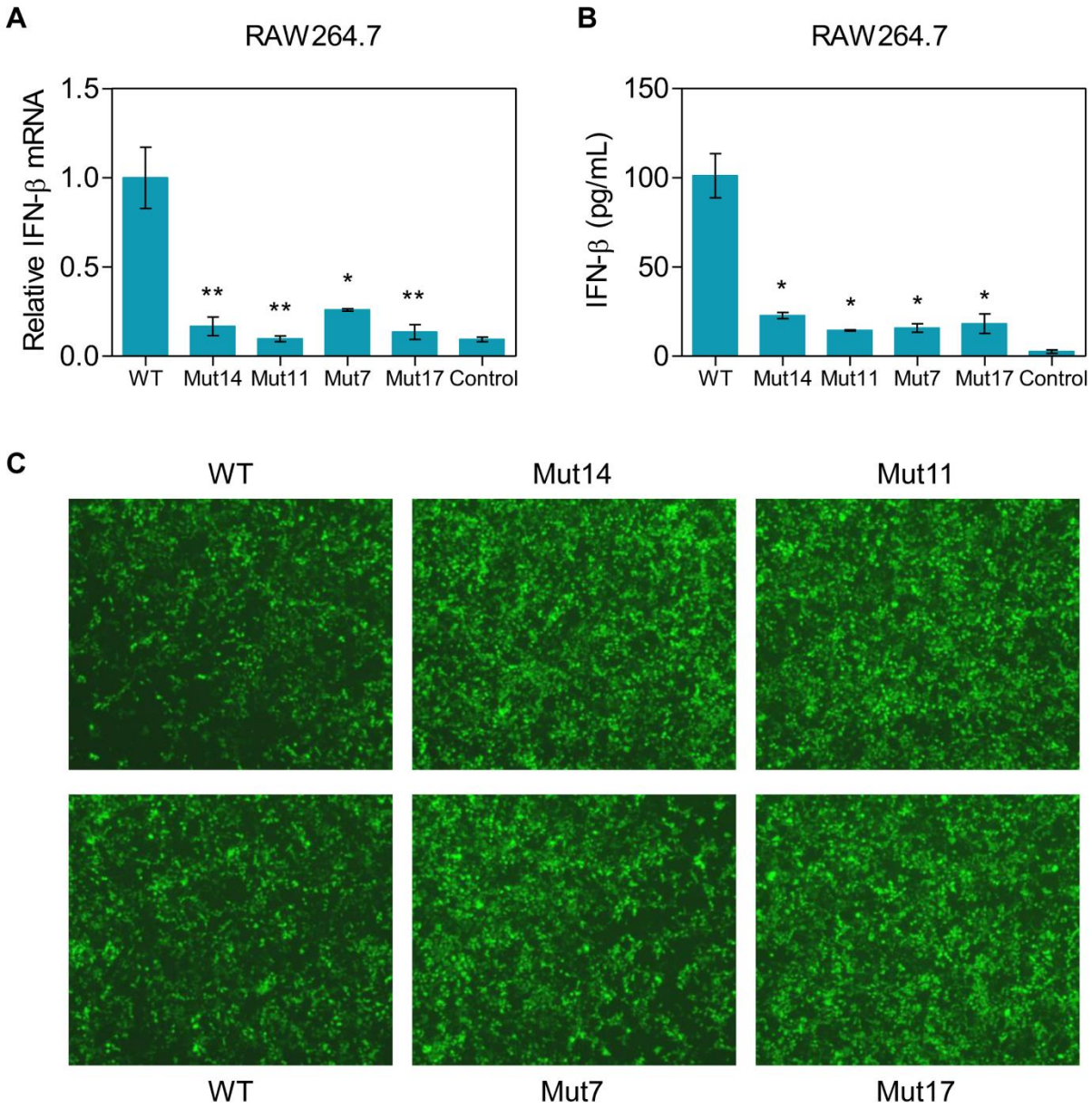


Figure 6. IFN-β response and EGFP expression in mammalian cells:

A. mRNA of IFN-β in murine RAW264.7 cells. Cells were seeded in 24-well plates at a density of 0.5×10^5 cells/well and transfected with 2 μg of each T7 RNAP transcript in triplicate. Following 24 h of incubation, total RNA was extracted, reverse transcribed, and analyzed for IFN-β mRNA, normalized to β-actin expression. Lipofectamine transfection served as a control. Significance levels: ** $p < 0.005$, * $p < 0.05$.

B. Quantification of IFN-β in murine RAW264.7 cells. Cellular extracts were analyzed for IFN-β protein levels via ELISA using Mouse IFN-β as a standard. Lipofectamine transfection served as a control. Significance level: * $p < 0.05$.

C. Image of EGFP-expressing HEK293 cells. HEK293 cells were transfected with EGFP mRNA using an identical protocol. After 24 h of incubation, cells were washed, diluted, and visualized under a fluorescence microscope.

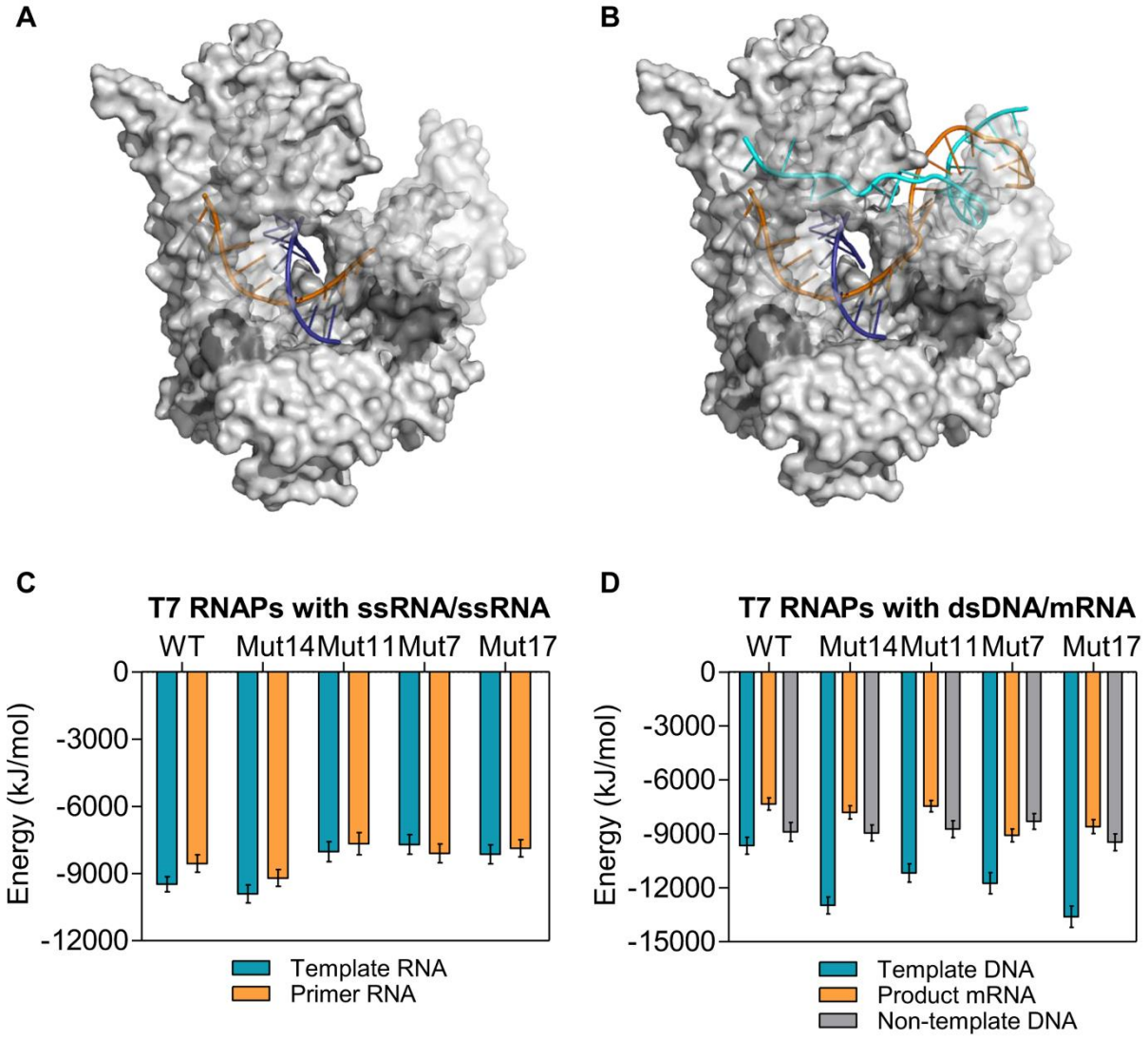
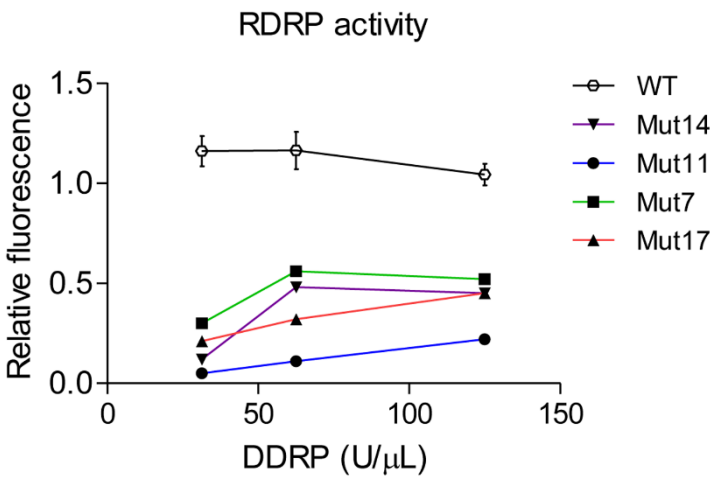


Figure 7. Molecular Dynamics Simulation of the interaction between T7 RNAPs and ssRNA/ssRNA (A and C) or dsDNA/mRNA (B and D):

A and B. Schematic diagram of the binding of T7 RNAP with ssRNA/ssRNA (A) or dsDNA/mRNA (B). T7 RNAP is shown as a gray surface with residues 331-406 and 569-688 set to transparent to allow views of DNA and RNA. The template DNA or RNA are shown in orange, non-template DNA in pale cyan, mRNA or primer RNA in deep blue. The figure was generated using the PyMOL software.

C and D. The interaction energetics between T7 RNAPs and ssRNA/ssRNA (C) or dsDNA/mRNA (D). All-atom MD simulations were performed using the GROMACS software package. The interaction energetics include electrostatic and Van der Waals contributions averaged from simulated trajectories.



678

679 **Figure 8. Relative RDRP activity:**

680 The hairpin oligonucleotide serves as a substrate to distinguish RDRP from DDRP activity of T7 RNAP. It forms a
681 hairpin with a dC loop and RNA stem. Short poly(U) is complementarily paired with long poly(A) sequences to mimic
682 a primer-template complex. T7 RNAP was diluted to 125 U/μL, 62.5 U/μL, and 31.25 U/μL for RDRP activity testing
683 (n=3).

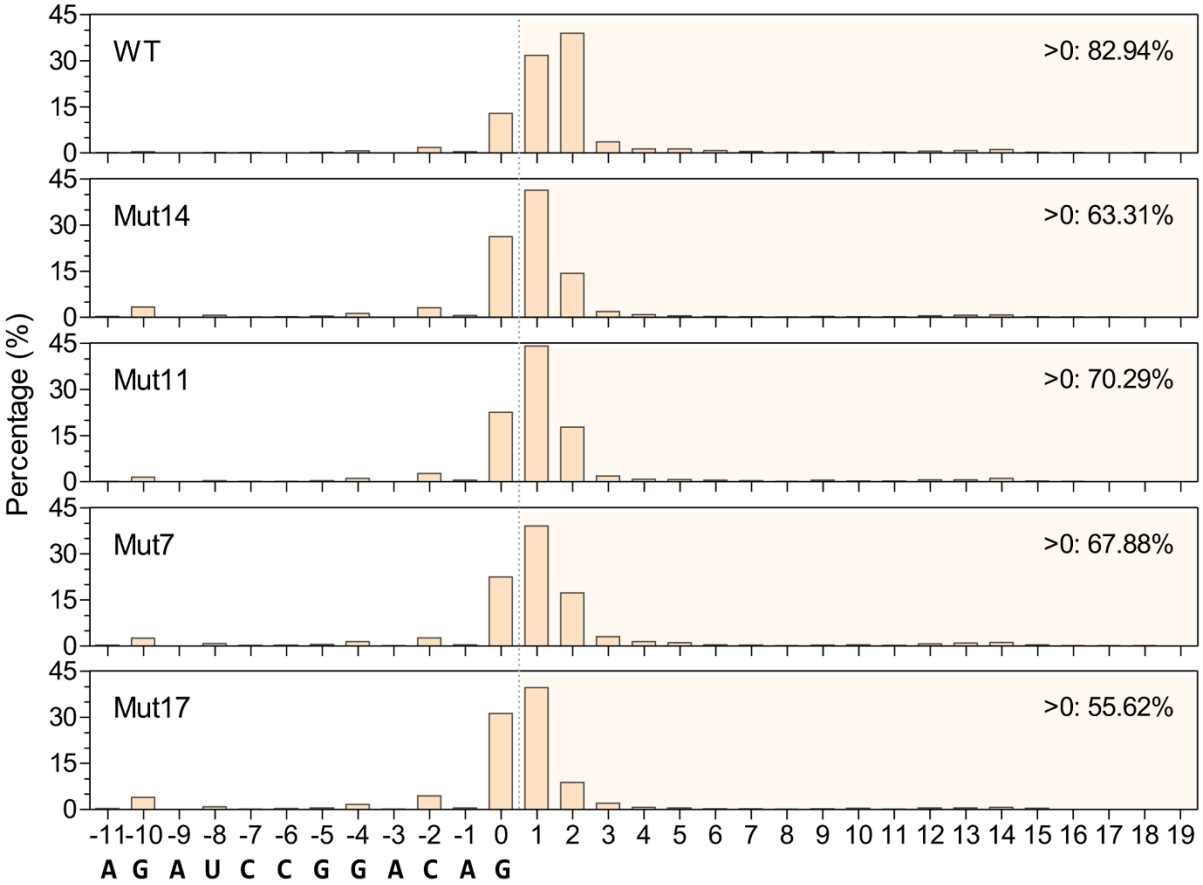


Figure 9. Heterogeneity in the 3'-end of RNA products:

Reverse transcription was performed on poly(A)-tailed transcripts, followed by NGS analysis to assess the terminal base composition of each RNA. The bottom bases represent the 3'-end sequence of target RNA. Each graph includes numbers on the right side, indicating the percentage of RNA molecules that are longer than the DNA template. At the terminal of RNA, the addition of bases is random, and the numbers indicate the quantity of bases without specifying their types.