



This is a repository copy of *Expanding the RNA polymerase biocatalyst solution space for mRNA manufacture*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/214298/>

Version: Published Version

---

**Article:**

Curry, E. [orcid.org/0000-0002-5123-8866](https://orcid.org/0000-0002-5123-8866), Sedelnikova, S., Rafferty, J. [orcid.org/0000-0003-1183-7474](https://orcid.org/0000-0003-1183-7474) et al. (4 more authors) (2024) Expanding the RNA polymerase biocatalyst solution space for mRNA manufacture. *Biotechnology Journal*, 19 (6). 2400012. ISSN 1860-6768

<https://doi.org/10.1002/biot.202400012>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## RESEARCH ARTICLE

## Expanding the RNA polymerase biocatalyst solution space for mRNA manufacture

Edward Curry<sup>1</sup>  | Svetlana Sedelnikova<sup>2</sup> | John Rafferty<sup>2</sup> | Martyn Hulley<sup>3</sup> | Melinda Pohle<sup>1</sup> | George Muir<sup>1</sup> | Adam Brown<sup>1</sup><sup>1</sup>Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, UK<sup>2</sup>School of Biosciences, University of Sheffield, Sheffield, UK<sup>3</sup>Bioprocess Development, AstraZeneca, Cambridge, UK

## Correspondence

Adam J. Brown and Edward Curry, Department of Chemical and Biological Engineering, University of Sheffield, Mappin St., Sheffield S1 3JD, UK.

Email: [adam.brown@sheffield.ac.uk](mailto:adam.brown@sheffield.ac.uk); [eajcurry1@sheffield.ac.uk](mailto:eajcurry1@sheffield.ac.uk)

## Funding information

AstraZeneca; Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/T508664/1

## Abstract

All mRNA products are currently manufactured in in vitro transcription (IVT) reactions that utilize single-subunit RNA polymerase (RNAP) biocatalysts. Although it is known that discrete polymerases exhibit highly variable bioproduction phenotypes, including different relative processivity rates and impurity generation profiles, only a handful of enzymes are generally available for mRNA biosynthesis. This limited RNAP toolbox restricts strategies to design and troubleshoot new mRNA manufacturing processes, which is particularly undesirable given the continuing diversification of mRNA product lines toward larger and more complex molecules. Herein, we describe development of a high-throughput RNAP screening platform, comprising complementary in silico and in vitro testing modules, that enables functional characterization of large enzyme libraries. Utilizing this system, we identified eight novel sequence-diverse RNAPs, with associated active cognate promoters, and subsequently validated their performance as recombinant enzymes in IVT-based mRNA production processes. By increasing the number of available characterized functional RNAPs by more than 130% and providing a platform to rapidly identify further potentially useful enzymes, this work significantly expands the RNAP biocatalyst solution space for mRNA manufacture, thereby enhancing the capability for application-specific and molecule-specific optimization of both product yield and quality.

## KEYWORDS

cell-free screening, in vitro transcription, mRNA manufacture, RNA polymerase

## 1 | INTRODUCTION

The clinical success of SARS-Cov-2 vaccines established synthetic mRNA as an effective drug format, paving the way for hundreds of new mRNA-based vaccines and gene therapies to enter clinical trials.<sup>[1,2]</sup> This has resulted in a sharp increase in global demand for mRNA

production, and shifted mRNA manufacturing from a relatively niche process to one that underpins current and future strategies to treat monogenic disorders, cancer, and infectious diseases.<sup>[3–5]</sup> All such mRNA products are currently produced in standardized in vitro transcription (IVT) systems using single-subunit DNA-dependent phage RNA polymerase (RNAP) biocatalysts. While *Salmonella* phage SP6 and *Enterobacteria* phage T3 RNAPs can be utilized in certain contexts, the dominant biocatalyst choice for synthetic mRNA manufacture

**Abbreviations:** CGE, capillary gel electrophoresis; CLuc, *Cypridina* luciferase; dsRNA, double-stranded RNA; IVT, in vitro transcription; RNAP, RNA polymerase.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published 2024. This article is a U.S. Government work and is in the public domain in the USA. *Biotechnology Journal* published by Wiley-VCH GmbH.

is the *Enterobacteria* phage T7 RNAP. This enzyme has undergone extensive protein engineering to improve bioproduction performance, predominantly via strategies to reduce formation of immunogenic product-related impurities such as short-abortive transcripts<sup>[6,7]</sup> and double-stranded RNA (dsRNA) species.<sup>[8,9]</sup>

Although T7 RNAP typically generates high product yields and acceptable product quality profiles, it is highly unlikely that a single one-size-fits-all biocatalyst approach will be optimal for all mRNA manufacturing contexts. Indeed, other bioproduction processes rely on biocatalyst toolbox approaches, such as the wide range of evolved and engineered Chinese Hamster Ovary cell factories utilized for recombinant protein manufacture.<sup>[10]</sup> The current unavailability of such an RNAP toolbox for mRNA IVT platforms restricts (i) bioprocess design strategies, such as optimizing temperature set-point to achieve quality target product profiles, and (ii) molecule/application specific optimization of product yield and quality. The latter is particularly pertinent given that mRNA product lines are diversifying to include larger and more complex molecular formats that present new biomanufacturing challenges, such as circular RNA,<sup>[11–13]</sup> long self-amplifying RNA transcripts,<sup>[14,15]</sup> and linear mRNA incorporating novel cap structures and modified nucleotides.<sup>[16]</sup> Indeed, it should be anticipated that mRNA will follow the path of protein therapeutics, where product designers rapidly progressed from relatively simple molecules such as insulin to highly engineered formats (e.g., tri-specific antibodies) that require product-specific biocatalyst solutions.<sup>[17]</sup>

The available RNAP toolbox has recently been expanded by studies focused on identifying and characterizing individual enzymes with putative desirable bioproduction phenotypes. KP34 enhances 3' homogeneity of product molecules by  $\approx 40$ -fold compared to T7, a result of being unable to catalyze self-templated 3' RNA end extension.<sup>[18]</sup> Syn5 exhibits unusually high levels of processivity and can produce larger transcripts at higher quantities and quality than T7, owing to its natural function of producing extended polycistronic mRNAs.<sup>[19]</sup> VSW-3 reduces dsRNA impurities by  $\approx 100$ -fold compared to T7, due to lacking any detectable RNA-dependent RNAP activity.<sup>[20]</sup> While these hypothesis-driven approaches have successfully identified new biocatalysts with novel functionalities, only six characterized RNAPs are currently publicly available for mRNA manufacture (although we note that some additional unpublished enzymes may be utilized in industrial settings). Accordingly, mRNA manufacturing solution spaces are severely limited by the restricted number of RNAPs available for bioproduction. Moreover, currently utilized "standardized" biocatalysts such as T7 may have relatively poor performance characteristics relative to the hundreds of "untested" single-subunit phage RNAPs found in nature, including lower processivity<sup>[19]</sup> and higher generation rates of impurities such as dsRNA<sup>[21]</sup> and abortive transcripts.<sup>[6]</sup>

In this study, we address the paucity of biocatalysts available for IVT-based mRNA production. Using a combination of *in silico* and *in vitro* analyses we identify and functionally validate eight new sequence-diverse RNAPs, more than doubling the number of previously described enzymes for mRNA manufacture. In doing so, we describe development of a high-throughput screening system that can

be utilized to rapidly select and test future RNAP libraries, facilitating further expansion of the biocatalyst solution space. Provision of a substantially expanded RNAP toolkit enhances capabilities to design and troubleshoot new molecule-specific manufacturing processes, which will be particularly useful for optimizing yield and quality of complex next-generation mRNA products.

## 2 | MATERIALS AND METHODS

### 2.1 | RNAP library creation

A starting library of 351 predicted RNAP sequences was collated from Uniprot, comprising all sequences annotated as predicted phage DNA-directed RNAPs. RNAPs were clustered by grouping RNAPs sharing sequence identity  $>85\%$ , using the Clustal Omega online alignment tool.<sup>[22]</sup> RNAPs were further clustered using the MMSEQ2 online server, with a minimum sequence identity threshold of  $85\%$ , and coverage threshold of  $70\%$ .<sup>[23]</sup> A representative RNAP from each cluster was chosen by totaling the matrix identity score to determine which polymerase in each cluster was most divergent in sequence to all others. Promoters for remaining RNAPs were predicted by PHIRE (Phage *in silico* regulatory elements)<sup>[24]</sup> using parameters of string lengths – 20, window size – 30, and degeneracy – 4. Predicted promoter sequences were verified with the *PhagePromoter* tool,<sup>[25]</sup> with a probability threshold of 0.5.

### 2.2 | Plasmid construction

For coupled transcription-translation assay plasmids, RNAP sequences were synthesized and cloned into XhoI and XbaI restriction sites on the pTNT vector (Promega). To create the corresponding transcription templates, the NanoLuc gene (Promega) was cloned into XhoI and XbaI restriction sites in pTNT, before site directed mutagenesis to substitute the SP6 and T7 promoter with the promoter of interest. For RNAP overexpression plasmids, RNAP sequences were inserted between NdeI and XhoI sites on pET-29b (Novagen). Transcription templates were made by site directed mutagenesis to substitute the T7 promoter with the promoter of interest on pCMV-Cluc2 (New England Biolabs).

### 2.3 | Cell free coupled transcription-translation assays

Coupled transcription-translation assays were carried out using the TNT SP6 Quick Coupled transcription/translation system (Promega). Reactions were assembled containing 8  $\mu\text{L}$  TnT Quick Master Mix, 1  $\mu\text{L}$  RNAP plasmid ( $40\text{ ng }\mu\text{L}^{-1}$ ), 1  $\mu\text{L}$  NanoLuc plasmid ( $80\text{ ng }\mu\text{L}^{-1}$ ), and 0.2  $\mu\text{L}$  1 mM methionine. The assay proceeded at  $30^\circ\text{C}$  for 1 h. Samples were then diluted 500-fold in nuclease free water, and added at a 1:1 ratio to pre-diluted NanoLuc luciferase assay substrate. Samples

were incubated in darkness for 5 min, before detection of luminescence by Molecular Devices ID5 plate reader, with an integration time of 10 s.

## 2.4 | Cell free coupled transcription-translation for quality assessment

To produce mRNA for quality assessment, coupled transcription-translation assays were carried out using the TNT SP6 Quick Coupled transcription/translation system (Promega). Reactions were assembled containing 64  $\mu\text{L}$  TnT Quick Master Mix, 1  $\mu\text{L}$  RNAP plasmid (400 ng  $\mu\text{L}^{-1}$ ), 1  $\mu\text{L}$  NanoLuc plasmid (400 ng  $\mu\text{L}^{-1}$ ), and 1.6  $\mu\text{L}$  1 mM methionine, and 12.4  $\mu\text{L}$  ddH<sub>2</sub>O. The assay proceeded at 30°C for 2 h. Polyadenylated mRNA was subsequently purified from the reaction by Dynabeads Oligo (dT)<sub>25</sub> (ThermoFisher) magnetic bead based affinity enrichment, following the manufacturers protocol. Briefly, 2 mg oligo (dT)<sub>25</sub> beads are equilibrated in binding buffer (20 mM Tris-HCl, pH 7.5, 1.0 M LiCl, 2 mM EDTA), before addition of mRNA containing samples. Beads were washed twice with washing buffer (10 mM Tris-HCl, pH 7.5, 0.15 M LiCl, 1 mM EDTA), before elution of bound poly(A) mRNA with ddH<sub>2</sub>O.

## 2.5 | Capillary gel electrophoresis (CGE)

CGE analysis of RNA was performed with a 5200 Fragment Analyzer System (Agilent), using the DNF-471 RNA Kit (15 nt) (Agilent). The capillary cassette used was FA 12-Capillary Array Short, 33 cm (Agilent). Samples were diluted to <100 ng  $\mu\text{L}^{-1}$  in nuclease free water. Before each separation, a pre-run voltage was applied (8 kV for 30 s), the capillaries were conditioned with the conditioning solution and the capillaries were dipped twice in the rinse buffer. Following this, the capillaries were filled with RNA separation gel (by pressure) and then the sample was introduced using a voltage injection (5 kV for 4 s). The separation was then conducted by applying a voltage of 8 kV for 45 min. Detection was carried out using laser induced fluorescence, by fluorescent dye tagging of the RNA.

## 2.6 | Transfection of NanoLuc mRNA into HEK293 cells

Suspension adapted Expi293f HEK cells (Thermo Fisher) were routinely cultured in Gibco FreeStyle 293 Expression Medium (Thermo Fisher). Cells were maintained in 30 mL volume in 125 mL Erlenmeyer flasks (Corning), at 37°C, 80% humidity, and 5% CO<sub>2</sub>, with agitation at 140 rpm. Cell density and viability was determined by the Vicell Blue automated cell counter system (Beckman Coulter). Cells for transfection were cultured in 24-shallow well plates (Corning), containing 500  $\mu\text{L}$  culture volume, with agitation at 240 rpm. For mRNA transfection, cells were seeded at a density of  $0.5 \times 10^6$  cells per mL in 24-shallow well plates, and incubated for 24 h. TransIT-mRNA trans-

fection reagent (Mirus) was used to transfect 3  $\mu\text{L}$  of mRNA per well as per manufacturer's instructions.

## 2.7 | RNAP expression and purification

pET-29b-RNAP plasmids were transformed into BL21 or NEB Shuffle *Escherichia coli* cells (New England Biolabs), and grown in 5 mL culture overnight at 37°C. Starter cultures were used to inoculate 500 mL LB broth, which was incubated at 37°C until an OD<sub>600</sub> of 0.4 was reached. At this point incubation temperature was lowered to 25°C. Protein expression was then induced by addition of 1 M IPTG (Sigma-Aldrich), before harvesting of cells after 8 h. Cell pellets were resuspended for purification in buffer A (50 mM Tris-HCl, 0.5 M NaCl, pH 8.0), and lysed by sonication. After removal of cell debris by centrifugation (73,000  $\times g$  for 15 min), cell free extract was applied to a 5 mL HisTrap HP column (Cytivia) at a flow rate of 5 mL min<sup>-1</sup>. The column was washed with 2 column volumes Buffer A + 40 mM imidazole (Sigma-Aldrich), before elution of protein in a gradient of imidazole from 0 to 300 mM over 10 column volumes. 5 mL of eluted protein was applied to a 1.6  $\times$  60 cm Superdex200 gel filtration column at 1.5 mL min<sup>-1</sup>, with 2 mL fractions collected after void volume. Fractions containing the RNAP of interest were concentrated to a final concentration of 1 mg mL<sup>-1</sup> and exchanged into RNAP storage buffer (50 mM Tris-HCl, 100 mM NaCl, 50% v/v glycerol, 10 mM DTT, 0.1 mM EDTA, 0.2% w/v Na<sub>3</sub> [All Sigma-Aldrich]), using a 50 kDa MWCO centrifugal filter (Sigma-Aldrich). RNAP preparations were stored at -20°C. Purity of final preparations was assessed by Tris-Glycine SDS-PAGE 4%-12% BT Novex gel with MES running buffer (Invitrogen).

## 2.8 | In vitro transcription

Plasmid templates for IVT were linearized with XbaI, and purified by ethanol precipitation. Transcription reactions using the Hiscribe IVT kit (New England Biolabs) were assembled to a final volume of 20  $\mu\text{L}$ . Reactions contained 2  $\mu\text{L}$  10X reaction buffer, 2  $\mu\text{L}$  of each NTP, 1  $\mu\text{g}$  of template DNA, and 2  $\mu\text{L}$  of T7 RNAP, or 2  $\mu\text{L}$  of novel RNAP. Transcription reactions were incubated for 2 h, before addition of 1  $\mu\text{L}$  DNase I, and further incubation for 20 min. Transcription reactions were purified using the Monarch RNA cleanup kit (New England Biolabs). mRNA concentration was quantified by nanodrop spectrophotometer, and product integrity assessed by agarose gel electrophoresis.

# 3 | RESULTS AND DISCUSSION

## 3.1 | Bioinformatic analysis of the potential RNA polymerase biocatalyst solution space

The biocatalyst solution space for mRNA production is currently limited to a handful of characterized RNAPs. To define the theoretical solution space, we extracted the sequence of all putative single

subunit RNAPs from Uniprot. At the time of conducting this analysis, 351 distinct single-stranded RNAPs had been predicted from publicly available genomics data. Accordingly, given that only six of these enzymes had been previously tested,  $\approx 98\%$  (i.e., 345 out of the potential 351 enzymes) of the potential solution space remained unexplored. We concluded that testing all 345 previously untested RNAPs would be inefficient, and, moreover, unnecessary, given that many of these enzymes will share similar performance characteristics. Indeed, we assumed that variation in bioproduction phenotype (e.g., enzyme processivity, impurity generation profiles) would be underpinned by significant differences in amino acid sequences. Accordingly, we sought to define distinct spots within the potential solution space by identifying RNAP clusters that shared minimal amino acid sequence homology (Figure 1A).

The Clustal Omega sequence alignment tool<sup>[22]</sup> was used to define RNAP clusters, whereby enzymes with  $>85\%$  global sequence identity were grouped into a single distinct family. This analysis identified 93 enzyme clusters, where the smallest and largest groups contained 1 and 32 RNAPs, respectively (Figure 1B). To interrogate local sequence similarities, these families were then analyzed using MMSEQ2, grouping RNAPs based on k-mer matching and the Smith–Waterman algorithm. Using sequence identity and coverage thresholds of 85% and 70%, respectively,<sup>[23]</sup> the number of discrete RNAP families was reduced from 93 to 44, where the majority of clusters (27) contained a single enzyme (Figure 1C).

Utilization of novel RNAPs for mRNA manufacture requires concomitant identification of appropriate cognate promoter elements to drive product transcription. This is non-trivial as single subunit RNAPs typically display highly stringent promoter binding activity, where mutation of a single nucleotide can abolish transcriptional output.<sup>[26]</sup> Indeed, the high specificity binding activity of single subunit RNAPs to promoters is highlighted by recent work showing that any changes to the T7 promoter sequence results in substantial reductions in mRNA production efficiency.<sup>[27–29]</sup> Accordingly, identification of novel functional polymerases necessitates highly accurate promoter predictions. However, there are only limited publicly available tools to achieve this, where PHIRE<sup>[24]</sup> searches for conserved elements of defined length, and PhagePromoter utilizes machine learning models to classify specified phage sequences as “promoter” or “non-promoter”.<sup>[25]</sup> A representative polymerase was selected from each of the 44 families, and the associated phage genomes were investigated with both of these promoter prediction tools (Figure 1D). This analysis failed to accurately identify cognate promoters in 18/44 cases. Testing further RNAPs from these 18 families similarly failed to result in identification of useable elements, indicating that the amino acid sequence diversity within these clusters is associated with “unusual” cognate binding motifs that are significantly different to the promoter datasets that were used to train existing prediction tools. Accordingly,  $\approx 40\%$  of identified RNAP clusters could not be tested in vitro due to limitations in promoter prediction capabilities.

Cognate promoters were successfully predicted for the remaining 26 RNAPs, and optimal reaction temperatures were identified for each enzyme based on the growth temperature of corresponding phage

hosts (Table 1). Ten RNAPs had predicted temperature optima  $\leq 30^\circ\text{C}$ , which may be beneficial for mRNA product quality profiles given that IVT reactions performed at reduced temperatures are associated with decreased levels of product-related impurities.<sup>[20]</sup> Phylogenetic analysis of the 26 selected RNAPs confirmed that the final library comprised a panel of evolutionarily diverse enzymes, sharing no significant sequence similarity ( $\leq 75\%$  global sequence identity) with any the six previously characterized polymerases (Figure 1E). These polymerase-promoter pairs (Table 1) were taken forward for in vitro functional characterization, facilitating testing of  $\approx 60\%$  of the identified RNAP clusters.

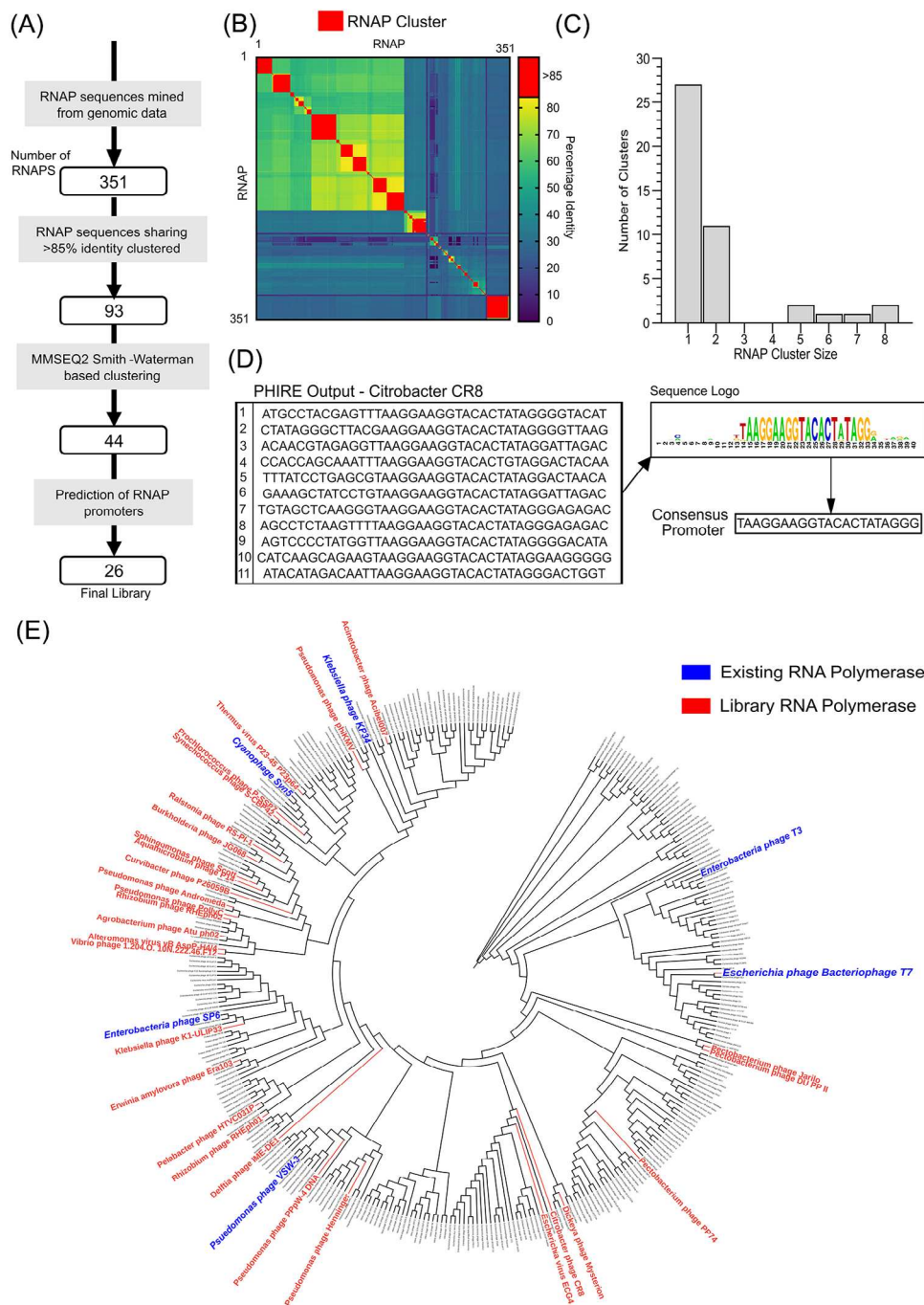
### 3.2 | Identification of novel active RNAP biocatalysts via high-throughput in vitro functional characterization

Previous studies focused on identifying new RNAP biocatalysts for mRNA manufacture have relied on recombinant production of individual “test” enzymes in *E. coli* cell-hosts, prior to characterization in IVT reactions.<sup>[18,19,30]</sup> This time-consuming method is undesirable for characterization of a large RNAP library, particularly given that manufacture of complex proteins at appropriate yield and quality can require significant process optimization.<sup>[31,32]</sup> Moreover, variation in recombinant protein stability and purity may prevent accurate quantification of relative enzyme activities across the library. Accordingly, to functionally characterize our 26 novel RNAPs in parallel, we developed a high-throughput testing platform that does not require production and purification of each polymerase. This was achieved by adapting a cell-free coupled transcription-translation system that has previously been employed to rapidly assess activity of variant T7 polymerases.<sup>[33,34]</sup> As shown in Figure 2A, this platform utilizes a mastermix containing rabbit reticulocyte lysate and recombinant SP6 RNAP to facilitate in vitro production of a “test” RNAP, which then in turn drives expression of a Nano-luciferase reporter-gene under the control of its cognate promoter.

Protein coding sequences and predicted cognate promoter elements for each of the 26 test enzymes were chemically synthesized and inserted into the appropriate RNAP screening platform vectors (Figure 2A). Resulting plasmid-pairs were individually mixed with the SP6 RNAP-rabbit reticulocyte lysate mastermix, and luciferase production was measured after incubating the reaction for 1 h at  $30^\circ\text{C}$  (recommended assay reaction temperature). As shown in Figure 2B, 8/26 enzymes were functionally active, driving luciferase expression levels that ranged from 10% to 161% of that achieved using the control T7 RNAP. Accordingly,  $\approx 70\%$  of tested enzymes were non-functional, highlighting the difficulty associated with identifying novel RNAP biocatalysts.

There was no significant correlation between predicted enzyme temperature optima and observed activity at  $30^\circ\text{C}$ . However, to further assess the impact of reaction parameters on polymerase performance, we tested enzyme activities at increased ( $37^\circ\text{C}$ ) and decreased ( $20^\circ\text{C}$ ) temperatures. While the same eight RNAPs were functional at





**FIGURE 1** Bioinformatics-driven design of an RNA polymerase (RNAP) “test” library (A). Putative RNAPs were clustered using pairwise global sequence identity analysis (B), and subsequently grouped into 44 distinct families according to local sequence similarities (C). Representative polymerases from families for which accurate cognate promoter prediction was possible (D) were phylogenetically analyzed to validate evolutionary diversity, represented in the circular cladogram showing all 351 analyzed RNAPs (E).

20°C, only three of these enzymes displayed activity at 37°C. Moreover, apart from R6 which drove highest luciferase expression levels under all conditions tested, the relative performance of polymerases varied with temperature. Although our objective was to identify effective polymerase-promoter pairs, rather than to precisely elucidate their relative performance characteristics, these data indicate that enzymes active over a narrow range of temperatures may be incorrectly categorized as non-functional. However, we concluded that this

was unlikely when testing across three separate temperature set-points, and that enzyme inactivity in our screening platform was more likely due to either (i) inaccurate annotation/sequencing of putative RNAP coding sequences or (ii) incorrect promoter prediction.

Although enzyme activity in the cell-free screening system may not be directly predictive of performance in IVT-based mRNA manufacturing processes, it is notable that polymerase R6 drove higher levels of luciferase expression than T7 in all conditions tested (increase ranging

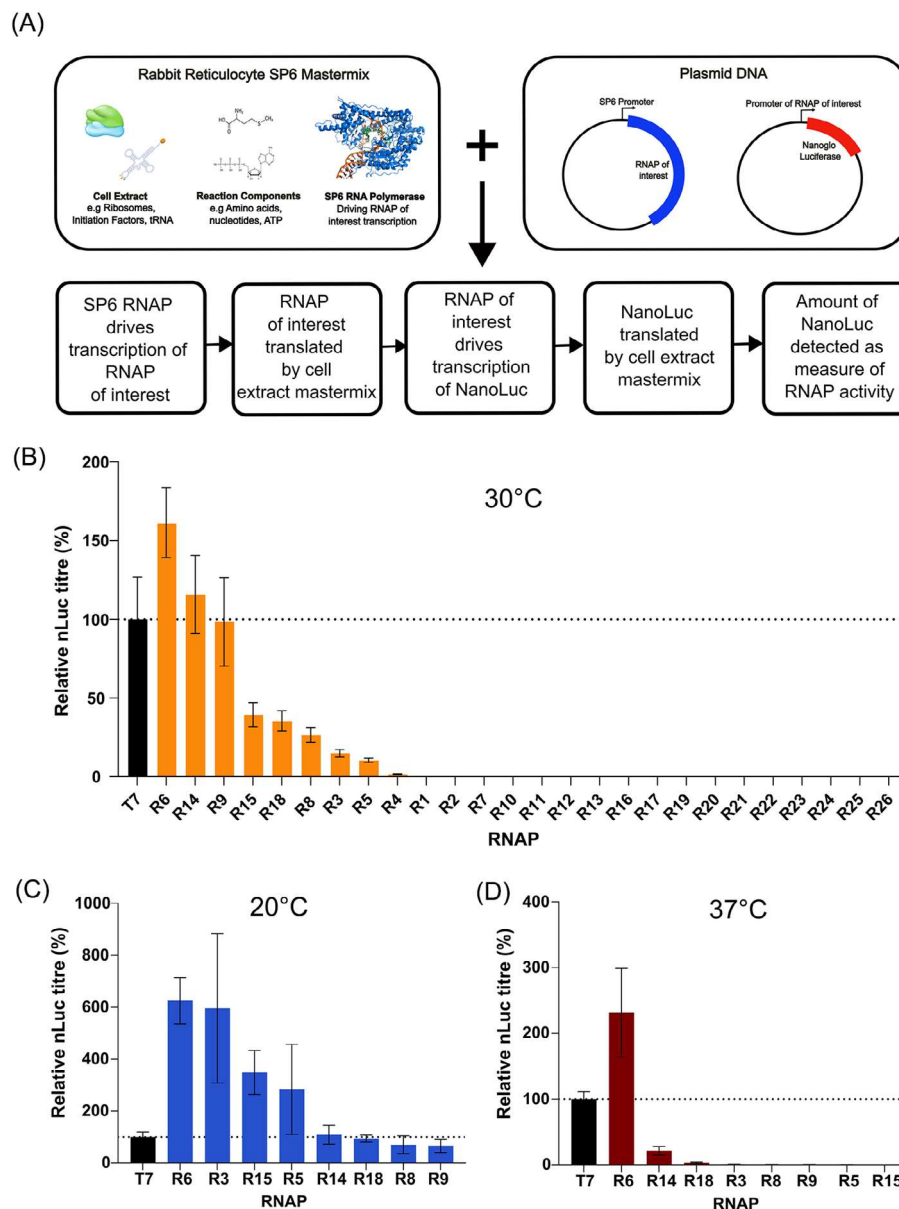
**TABLE 1** Bioinformatically identified single-subunit RNA polymerases selected for in vitro functional characterization.

Phage name	Assigned number	Temperature optimum [°C]	Uniprot protein ID	Promoter prediction
<i>Prochlorococcus</i> phage P-SSP7	R1	14–26	Q58N45	AAAATTCTTCAAGTTTACAA
<i>Synechococcus</i> phage S-CBP42	R2	20	A0A096VKW2	CACTTCCACTCAACCAACCG
<i>Erwinia amylovora</i> phage Era103	R3	28	A2I7X6	AATAACCACCCAGTATAGAAGGAA
<i>Agrobacterium</i> phage Atu ph02	R4	28	A0A223VZI2	TTATCCTTCGTATAAGGAATA
<i>Dickeya</i> phage Myserion	R5	28	A0A385IGY0	CTTAAATCATCACTATTAG
<i>Pectobacterium</i> phage PP74	R6	30	A0A1J0MEG1	TAATACGACTCACTATTGGGAA
<i>Aquamicrobium</i> phage P14	R7	30	A0A1L5C074	TTTCGGTACGCTCTAGCA
<i>Pectobacterium</i> phage DU PP II	R8	30	A0A2D2W5U8	TTATTAACGACTCACTACTAGGAA
<i>Pectobacterium</i> phage Jarilo	R9	30	A0A2S1GSW7	TAATAACGACTCACTATTAGAAG
<i>Sphingomonas</i> phage Scott	R10	30	A0A346FDD2	TCGGGTGTGCGATTTCCTTAC
<i>Ralstonia</i> phage RS-PI-1	R11	35	A0A1S6L1D6	GTCTGAAGTCGTCGAGCAGC
<i>Burkholderia</i> phage JG068	R12	37	U3PFP4	TCAGTAGACTATCTAG
<i>Acinetobacter</i> Acibel007	R13	37	A0A075DXW8	CTGTACTCACAGCTCAATTT
<i>Delftia</i> phage IME-DE1	R14	37	A0A0F7INH1	GTTAGCCACACCATTGAAGACCC
<i>Pseudomonas</i> phage Henninger	R15	37	A0A2K9VHD7	TTAAACCCCTCACTATGGCTACA
<i>Pseudomonas</i> phage PollyC	R16	37	A0A2K9VHU7	CTCACTCACGACCCAAATTC
<i>Pseudomonas</i> phage phiKMV	R17	37	Q7Y2D9	CGACCCCTCCCTACTCCGGCCTTAAAT
<i>Citrobacter</i> phage CR8	R18	37	W6PP41	TAAGGAAGGTACACTATAGGG
<i>Thermus</i> virus P23-45 P23p64	R19	65	A7XX94	TTATTCCTTTA
<i>Pelagibacter</i> phage HTVC031P	R20	16–23	A0A4Y1NTX3	AACTAATGCTCAATTTAGAGATA
<i>Rhizobium</i> phage RHEph01	R21	25–30	L7TQW5	ATTACCCCTCCCTTAAGCAAAG
<i>Rhizobium</i> phage RHEph02	R22	25–30	L7TJC5	TTAATCCTCACTATTAGGATAA
<i>Curvibacter</i> phage P26059B	R23	25–30	A0A384UH57	GCAACATTACAGGTACTGAA
<i>Pseudomonas</i> phage PPpW-4	R24	25–30	V5YUU1	TAAAAACCCCTCACTGAAACAGGG
<i>Vibrio</i> phage 10N	R25	35–37	A0A2I7RNL7	ACTTACCTTCACTATAGCAGCA
<i>Alteromonas</i> virus vB AspP-H44	R26	Not found	A0A220YL66	TGGTGACTACAGAGCAGCAG

between 160% and 620%), including a 220% increase at T7s optimum reaction temperature (37°C). Five further RNAPs (R3, R5, R9, R14, R15) facilitated luciferase titers greater than or equal to that achieved with T7 in at least 1 reaction condition. Accordingly, these enzymes may exhibit higher processivity/catalytic activity than T7 and could therefore have potential use in enhancing mRNA production yields. Moreover, their use may permit simplified downstream processing operations via reduced formation of product-related impurities, particularly as many of these RNAPs exhibit relatively high activities at low temperatures.<sup>[30]</sup>

Beyond raw catalytic power, that is, the relative rate at which each an enzyme produces mRNA from a given mass of DNA template (Figure 2), there are several other properties that would ideally be characterized by an in vitro RNAP screening platform. A complete comparative assessment would include measurement of a wide range of impurities, however current methodologies to quantify critical attributes such as dsRNA levels and polyA tail homogeneity are incompatible with a high-throughput small-scale testing module. While a comprehensive quantitative analysis is therefore intractable, we reasoned that the utility of the platform would be significantly

enhanced by developing a standardized secondary product-quality test to further evaluate RNAPs that pass the initial “functionality assessment”. As shown in Figure 3A, using the eight identified functional enzymes, we performed scaled-up (80 µL) cell-free coupled transcription-translation assays, before Oligo-(dT) based magnetic bead affinity enrichment of resulting luciferase mRNA molecules. Purified product samples were split into two downstream quality assessment tests to evaluate (i) generation of product-related impurities (i.e., proportion of full-length mRNA) by CGE analysis and (ii) sequence fidelity/functionality by transfection into HEK 293 cells followed by quantification of Luciferase protein levels. The resulting data profiles were then compared to the T7 control, where enzymes that produced a single sharp peak in the electropherogram, and measurable active luciferase in the cell-based assay were designated as passing the quality evaluation step (Figure 3B). These data confirmed that all eight enzymes produced functional homogenous mRNA molecules (Figure 3C), validating the use of the quality assessment tests and confirming the potential utility of these RNAPs. While further complete characterization is required to fully assess their bioindustrial application, the identification of eight novel functional enzymes more than



**FIGURE 2** RNA polymerases (RNAPs) were functionally characterized in a cell-free coupled transcription-translation system (A). Protein coding and cognate promoter sequence pairs were inserted into screening platform vectors and incubated with SP6-rabbit reticulocyte mastermix at 30°C (B), 20°C (C), and 37°C (D). Luciferase expression was quantified 1 h post-incubation; data are expressed as a percentage of the production achieved using the control T7 RNAP. Values represent the mean + standard deviation (SD) of three independent experiments ( $n = 3$ , each performed in triplicate).

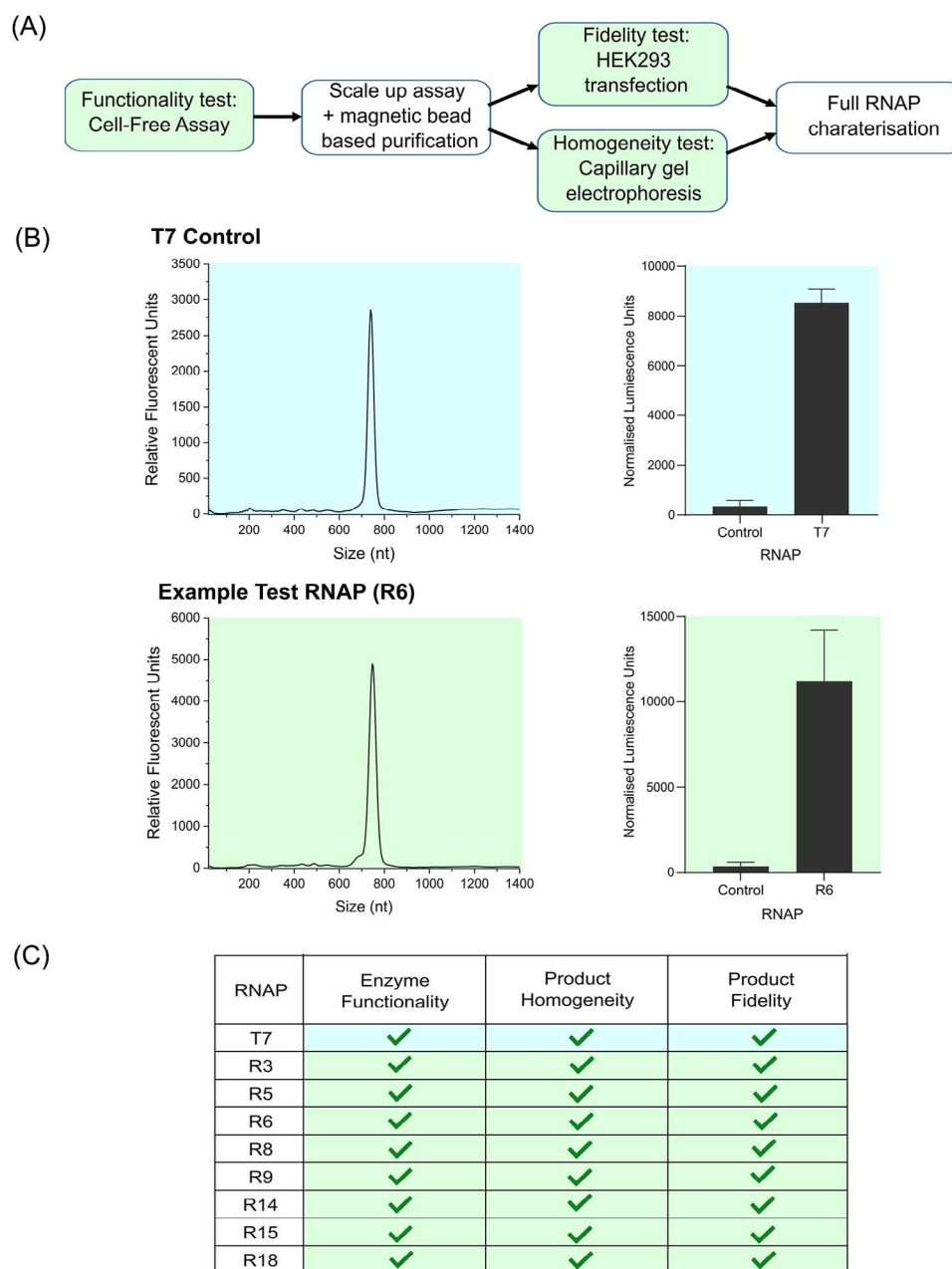
doubles the number of available RNAPs, expanding the biocatalyst solution space for mRNA manufacture by  $\approx 130\%$  (i.e., an expansion from 6 to 14 total available enzymes).

### 3.3 | Cognate promoter prediction is the critical limiting factor restricting further expansion of the RNAP biocatalyst solution space

The finding that  $\approx 70\%$  of characterized enzymes were non-functional in *in vitro* tests (Figure 2) indicates that the vast majority of puta-

tive RNAPs cannot be simply extracted from online databases and directly employed in mRNA manufacturing applications. Given that RNAPs are known to display highly stringent promoter recognition requirements,<sup>[26]</sup> we hypothesized that enzyme inactivity may have resulted from inaccurate predictions of cognate promoter sequences. To exemplify this, we characterized the ability of R6, the best performing polymerase in *in vitro* screens, to initiate transcription from the promoters of other functional enzymes. As shown in Figure 4A, R6 could not drive quantifiable gene expression from any of these variant elements, where even a single nucleotide change was sufficient to completely abolish transcriptional output. These data highlight that



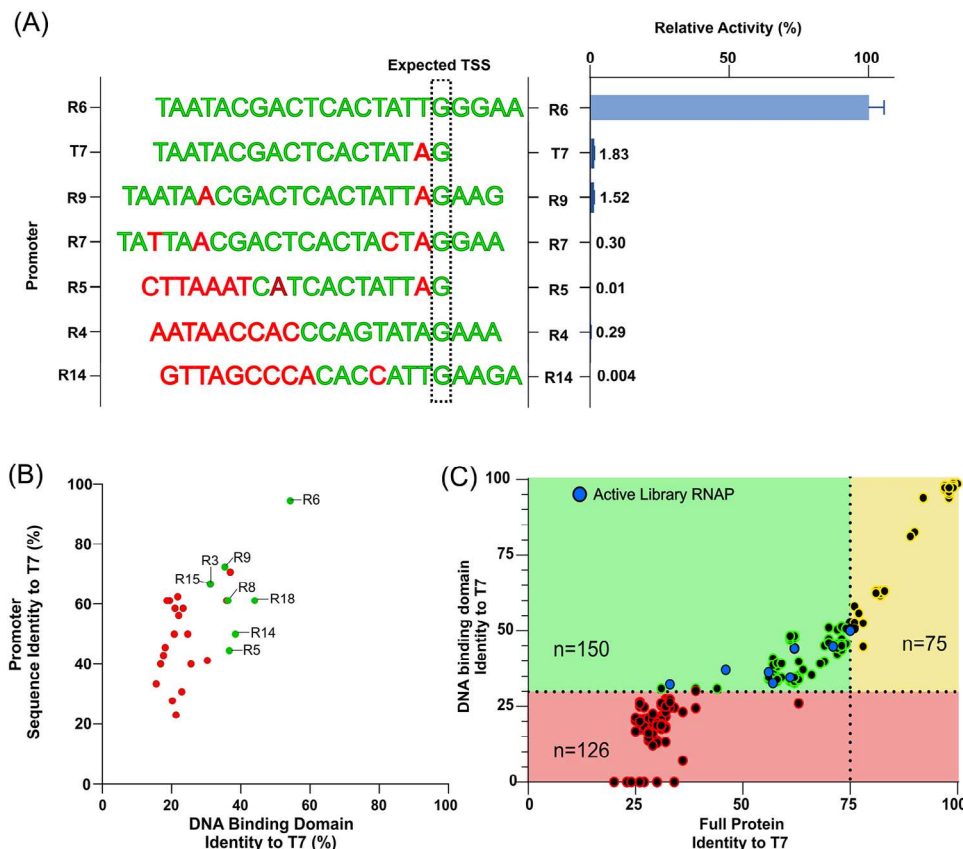


**FIGURE 3** (A) Process flow diagram for high-throughput assessment of novel RNA polymerases (RNAPs); enzymes must pass key performance tests (shown as green boxes) to progress to the next analysis step. (B) Functional RNAPs (Figure 2) were used in scaled-up cell-free coupled transcription-translation assays to produce luciferase mRNA that was subsequently purified using Oligo-(dT) magnetic beads. mRNA quality was profiled by (i) capillary gel electrophoresis analysis and (ii) transfection into HEK293 cells prior to measurement of luciferase protein levels after 48 h (values represent the mean + standard deviation [SD] of three independent experiments). Electropherogram peak distribution and luciferase enzyme activity units were compared to benchmark T7 controls (shown in blue) to confirm whether test RNAPs (example profile of R6 shown in green) had passed product homogeneity (presence of similar single sharp peak) and fidelity (measurable luciferase activity) assessments. (C) Quality analytics profile for all functional RNAPs identified in this study. Enzymes which pass all three performance tests are considered promising candidates for use in *in vitro* transcription (IVT)-based mRNA production processes.

the ability to exploit any given potential RNAP biocatalyst is heavily dependent on highly accurate definition of its cognate promoter sequence.

We reasoned that RNAP promoter prediction tools may be incapable of precisely defining new elements that are significantly divergent from currently known sequences, as evidenced by our inability to

derive cognate promoters for  $\approx 40\%$  of bioinformatically determined RNAP clusters (see Section 3.1). Existing promoter prediction algorithms are primarily trained based on known promoter sequences, and their predictive ability is accordingly limited for novel test enzymes with significantly different sequences and structures.<sup>[24,35]</sup> Indeed, given the paucity of characterized RNAP promoters, novel “test”



**FIGURE 4** (A) The ability of RNA polymerase (RNAP)6 (*Pectobacterium* phage PP74) to drive Luciferase expression from varying non-cognate promoter elements was evaluated in cell-free coupled transcription-translation assays (see Figure 2). Data are expressed as a percentage of the production achieved using the cognate RNAP-6 promoter. Values represent the mean + SD of three independent experiments ( $n = 3$ , each performed in triplicate). (B) Test library RNAP-promoter pairs were analyzed to determine relative DNA binding domain and promoter sequence identity with T7. Pairs that were found to be active or inactive in functional characterization tests are shown as green and red dots respectively. (C) The entire theoretical RNAP biocatalyst solution space was analyzed to identify promising future targets for in vitro characterization (green section). Enzymes predicted to have incorrect promoter definitions or similar bioproduction phenotypes to T7 are shown in the red and yellow sections respectively. Functional RNAPs identified in this study are shown as blue dots.

enzymes may recognize sequence motifs and architectures that are (i) substantially different to those used to train/design current algorithms, and accordingly (ii) beyond the predictive capabilities of available tools. Rationalizing that divergence in promoter structure/sequence would be underpinned by differences at the amino acid level, we investigated whether enzyme inactivity was associated with DNA binding domain sequences that varied significantly to those of well-studied biocatalysts. As shown in Figure 4B, 8/10 polymerases that share relatively high DNA binding domain sequence identity with T7 ( $>30\%$ ) were found to be active, while all 16 enzymes that share relatively low similarity ( $<30\%$ ) were non-functional. This indicates that accurate prediction and assessment of new RNAPs may not be possible when DNA binding domain sequence similarity to well-studied enzymes falls below a certain threshold. In contrast, cognate promoter sequence similarity with T7 promoter was not a good predictor of RNAP functionality, where 6/8 active and 12/16 inactive elements shared between 40% and 65% sequence identity with T7 (Figure 2B). We therefore concluded that comparative amino acid sequence analysis with T7 is currently the most effective method to refine libraries

of putative RNAPs by identifying enzymes that are unlikely to be functional in in vitro testing.

Our findings indicate that potential RNAPs can be efficiently screened in silico, where enzymes that share below  $\approx 30\%$  DNA-binding domain sequence identity with T7 are unlikely to be functional in vitro owing to incorrect promoter definition. However, as shown in Figure 4C, this cut-off removes  $\approx 36\%$  of the theoretical biocatalyst solution space for mRNA production. Of the remaining 225 polymerases, 75 share relatively high overall protein sequence identity ( $>75\%$ ) with T7. Such enzymes are considered unlikely to exhibit substantial differences to T7 in key performance criteria such as enzyme processivity and product-related impurity generation. Accordingly, only 150 RNAPs are predicted to be both active in vitro and potentially display novel, desirable bioproduction functionalities (including the eight we have identified in this study). This analysis therefore highlights 142 promising additional biocatalyst targets for future investigation, including 88 that do not share high sequence identity ( $>75\%$ ) with either the six previously characterized RNAPs or the eight enzymes identified in this study (listed in Table S1). However, it also suggests

that >120 potentially useful enzymes are currently difficult to exploit, highlighting promoter prediction capability as the key limiting factor preventing comprehensive exploitation of the theoretical biocatalyst solution space for mRNA production. Although the cognate promoters of individual polymerases can be elucidated via non-bioinformatic laboratory techniques,<sup>[18]</sup> these time-intensive methods are intractable when testing multiple enzymes in parallel. Accordingly, full exploration of the putative RNAP biocatalyst solution space to optimize mRNA production processes will likely require significant advancements in phage promoter prediction tools.

### 3.4 | Novel identified RNAPs enhance the biocatalyst solution space for IVT-based mRNA production

To validate that novel RNAPs identified via our HT cell-free screening platform have utility in mRNA manufacturing processes, we recombinantly produced polymerases R5 and R6 in *E. coli*. These polymerases were chosen to represent highly and moderately active enzymes, where R6 (*Pectobacterium* phage PP74) was previously shown to be the best performing RNAP in all temperatures tested, and R5 (*Dickeya* phage Mysterion) drove relatively low-to-medium levels of transcription across varying reaction conditions (Figure 2). Polymerases were overexpressed in 0.5 L scale production processes and purified using His-tag affinity and size exclusion chromatographic operations. Purified recombinant RNAPs were then utilized in IVT reactions to manufacture *Cypridina* luciferase (CLuc) mRNA. As shown in Figure 5, both enzymes drove significant levels of Cluc expression, validating their function as biocatalysts for synthetic mRNA production. To evaluate enzyme robustness, we tested the performance of each RNAP at a range of pH (predicted optimum  $\pm 1$ ) and temperature (predicted optimum  $\pm 5^\circ\text{C}$ ) set-points. Both RNAPs were functional across all conditions tested, where R5 performance was relatively constant, and R6 activity increased with temperature. The latter highlights that expected phage host growth temperatures are not directly predictive of optimal in vitro reaction conditions for recombinant RNAPs.

As shown in Figure 5, utilization of R6 at “optimal” reaction parameters (pH 7.9,  $35^\circ\text{C}$ ) facilitated mRNA product titers >60% of that achieved when using NEB recombinant T7 at recommended conditions (pH 7.9,  $37^\circ\text{C}$ ). Although R6 drove higher levels of gene transcription than T7 in our cell-free system, it is not surprising that T7s relative activity was enhanced in IVT processes given that NEB T7 is a highly pure engineered enzyme with fully optimized reaction conditions. Indeed, we anticipate that rational protein engineering/evolution, coupled with improved purification techniques and reaction parameters (e.g., optimized  $\text{MgCl}_2$  concentration), will significantly increase R6s biocatalytic activity in IVT-based mRNA production. Irrespective of this, by initially facilitating product yields equivalent to  $\approx 61\%$  of that achieved by optimized T7, R6 is considered a highly active biocatalyst.

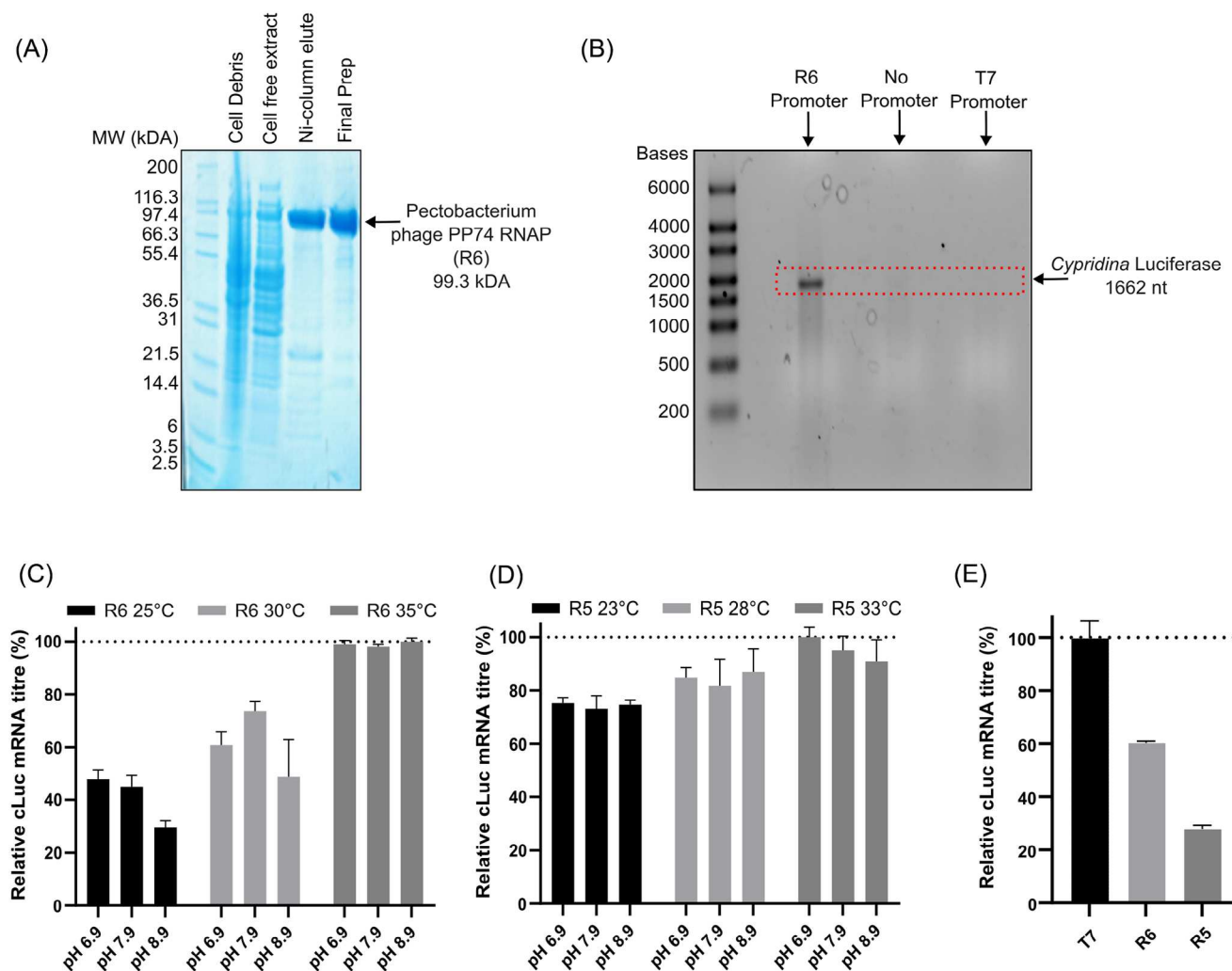
Polymerase R5 enabled product titers  $\approx 47\%$  of that achieved using R6, suggesting that the comparative performance of novel RNAPs in cell-free testing platforms is broadly predictive of their relative ability

to maximize mRNA yields in IVT manufacturing processes. Accordingly, we concluded that the additional six novel enzymes identified in this study (*Erwinia amylovora* phage Era103, *Pectobacterium* phage DUPP II, *Pectobacterium* phage Jarilo, *Delftia* phage IME-DE1, *Pseudomonas* phage Henninger, and *Citrobacter* phage CR8) are also likely to facilitate moderate-to-high mRNA production yields. While we cannot currently comment on the relative ability of these new polymerase to enhance product quality, previous work suggests they will generate variable levels of product-related impurities, such as dsRNA and truncated species.<sup>[18–20,30,36]</sup> Indeed, this new library is particularly likely to exhibit differential bioproduction phenotypes, given that they were specifically selected based on sharing minimal amino acid sequence similarities. We therefore conclude that addition of these eight novel functional enzymes to the RNAP biocatalyst solution space will significantly enhance IVT-based mRNA manufacturing optimization strategies. We note that industrialization of these enzymes will be dependent on additional factors, particularly the ability to manufacture them at high yields in large-scale production processes and their relative long-term stability.<sup>[31]</sup> While the two RNAPs tested here could be produced using standard upstream and downstream unit operations, full validation studies for all new enzymes will need to include comprehensive manufacturability assessments.

## 4 | CONCLUDING REMARKS

The eight novel sequence-diverse functional RNAPs identified in this study substantially increases the number of biocatalysts available for mRNA production. Although further work is required to comprehensively define their relative performance characteristics, particularly their associated impurity generation profiles, this expansion of the biocatalyst solution space significantly enhances design options for molecule-, process-, and application-specific optimization of mRNA product yield and quality. This improved flexibility will become increasingly useful as mRNA product lines continue to diversify toward large, complex molecules that pose new manufacturing challenges.<sup>[11,12,14–16]</sup> Our combined in silico and in vitro analysis of the theoretical biocatalyst solution space (i) showed that full exploitation of potential RNAPs is restricted by cognate promoter prediction capabilities, but, also (ii) identified a panel of enzymes that are particularly promising for future investigation.

The screening platform we have developed can be utilized to rapidly test additional RNAP libraries. For example, this will be useful for responsive functional characterization of putative enzymes that are newly-deposited in online protein databases each year. Moreover, it can be applied to significantly increase the throughput for testing rationally engineered enzyme variants. Protein engineering is a promising strategy to enhance the performance of candidate RNAPs, for example, by reducing the impurity generation profile of an enzyme that naturally exhibits high processivity rates. Our platform will permit initial functional screening of such libraries without requiring time-intensive processes associated with manufacturing each RNAP individually. Further, it could be utilized for product-RNAP matchmak-



**FIGURE 5** RNA polymerase (RNAP) 5 and 6 were recombinantly produced in *E. coli* and purified using His-tag affinity and size exclusion chromatography (A). Purified enzymes were used to manufacture luciferase mRNA in in vitro transcription (IVT) reactions, and production of full-length product was verified by gel electrophoresis analysis (B). IVT production processes utilizing RNAP6 (C) and RNAP5 (D) were performed in varying reaction conditions and resulting mRNA titers were quantified using nanodrop spectrophotometry. Using identified optimal reaction parameters for each enzyme, the relative performance of recombinant polymerases was evaluated compared to an NEB T7 control (E). Data in C and D are expressed as a percentage of the production achieved using optimal reaction parameters for each enzyme. Data in (E) are expressed as a percentage of the production achieved using T7. In (C), (D), and (E) values represent the mean + standard deviation (SD) of three independent experiments ( $n = 3$ , each performed in triplicate).

ing, to align mRNA molecule-specific features (e.g., length, GC content, secondary structure) with enzyme-specific performance characteristics. This may be particularly applicable to development of small-batch customized mRNA therapeutics that require cost-effective methods to optimize production of novel sequences.<sup>[1,37]</sup> However, to fully harness its potential for such applications the screening system will require further development, particularly the creation of additional product quality analytics modules that can quantify critical impurity levels such as dsRNA.<sup>[38]</sup> In conclusion, by more than doubling the number of available polymerases, and providing associated methods to select and screen further new enzymes, this study has facilitated a significant expansion of the RNAP biocatalyst solution space, enhancing strategies to optimize and troubleshoot IVT-based mRNA production processes.

#### AUTHOR CONTRIBUTIONS

**Edward Curry:** Conceptualization; methodology; analysis; investigation; manuscript preparation. **Svetlana Sedelnikova:** Investigation; analysis. **John Rafferty:** Investigation; analysis. **Martyn Hulley:** Conceptualization; analysis. **Melinda Pohle:** Investigation; analysis. **George Muir:** Investigation; analysis. **Adam Brown:** Conceptualization; methodology; analysis; manuscript preparation; funding acquisition.

#### ACKNOWLEDGMENTS

This study was supported by AstraZeneca and the Biotechnology and Biological Sciences Research Council (Grant No. BB/T508664/1).

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.



## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Edward Curry  <https://orcid.org/0000-0002-5123-8866>

## REFERENCES

- Qin, S., Tang, X., Chen, Y., Chen, K., Fan, N., Xiao, W., Zheng, Q., Li, G., Teng, Y., Wu, M., & Song, X. (2022). mRNA-based therapeutics: Powerful and versatile tools to combat diseases. *Signal Transduction and Targeted Therapy*, 7, 1–35.
- Webb, C., Ip, S., Bathula, N. V., Popova, P., Soriano, S. K. V., Ly, H. H., Eryilmaz, B., Nguyen Huu, V. A., Broadhead, R., Rabel, M., Villamagna, I., Abraham, S., Raeesi, V., Thomas, A., Clarke, S., Ramsay, E. C., Perrie, Y., & Blakney, A. K. (2022). Current status and future perspectives on mRNA drug manufacturing. *Molecular Pharmaceutics*, 19, 1047–1058.
- Al Fayed, N., Nassar, M. S., Alshehri, A. A., Alnefaie, M. K., Almughem, F. A., Alshehri, B. Y., Alawad, A. O., & Tawfik, E. A. (2023). Recent advancement in mRNA vaccine development and applications. *Pharmaceutics*, 15, 1–24.
- Liu, C., Shi, Q., Huang, X., Koo, S., Kong, N., & Tao, W. (2023). mRNA-based cancer therapeutics. *Nature Reviews Cancer*, 23, 526–543.
- Vavilis, T., Stamoula, E., Ainatoglou, A., Sachinidis, A., Lamprinou, M., Dardalas, I., & Vizirianakis, I. S. (2023). mRNA in the context of protein replacement therapy. *Pharmaceutics*, 15, 166.
- Guilleres, J., Lopez, P. J., Proux, F., Launay, H., & Dreyfus, M. (2005). A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 5958–5963.
- Lyon, S., & Gopalan, V. (2018). A T7 RNA polymerase mutant enhances the yield of 5'-thienoguanosine-initiated RNAs. *ChemBioChem*, 19, 142–146.
- Dousis, A., Ravichandran, K., Hobert, E. M., Moore, M. J., & Rabideau, A. E. (2023). An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nature Biotechnology*, 41, 560–568.
- Wu, M. Z., Asahara, H., Tzertzinis, G., & Roy, B. (2020). Synthesis of low immunogenicity RNA with high-temperature in vitro transcription. *RNA*, 26, 345–360.
- Fischer, S., Handrick, R., & Otte, K. (2015). The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances*, 33, 1878–1896.
- Bai, Y., Liu, D., He, Q., Liu, J., Mao, Q., & Liang, Z. (2023). Research progress on circular RNA vaccines. *Frontiers in Immunology*, 13, 1–12.
- Qu, L., Yi, Z., Shen, Y., Lin, L., Chen, F., Xu, Y., Wu, Z., Tang, H., Zhang, X., Tian, F., Wang, C., Xiao, X., Dong, X., Guo, L., Lu, S., Yang, C., Tang, C., Yang, Y., Yu, W., ..., Wei, W. (2022). Circular RNA vaccines against SARS-CoV-2 and emerging variants. *Cell*, 185, 1728–1744.
- Zhu, F., Ma, S., Wen, H., Rao, M., Zhang, P., Cui, Y., Yang, H., Tan, C., Chen, J., & Pan, P. (2023). Development of a novel circular mRNA vaccine of six protein combinations against *Staphylococcus aureus*. *Journal of Biomolecular Structure & Dynamics*, 41, 10525–10545.
- Blakney, A. K., Ip, S., & Geall, A. J. (2021). An update on self-amplifying mRNA vaccine development. *Vaccines*, 9, 1–26.
- Pourseif, M. M., Masoudi-Sobhanzadeh, Y., Azari, E., Parvizpour, S., Barar, J., Ansari, R., & Omid, Y. (2022). Self-amplifying mRNA vaccines: Mode of action, design, development and optimization. *Drug Discovery Today*, 27, 1–19.
- Chen, T. H., Potapov, V., Dai, N., Ong, J. L., & Roy, B. (2022). N 1-methylpseudouridine is incorporated with higher fidelity than pseudouridine in synthetic RNAs. *Scientific Reports*, 12, 1–13.
- Tihanyi, B., & Nyitray, L. (2020). Recent advances in CHO cell line development for recombinant protein production. *Drug Discovery Today: Technologies*, 38, 25–34.
- Lu, X., Wu, H., Xia, H., Huang, F., Yan, Y., Yu, B., Cheng, R., Drulis-Kawa, Z., & Zhu, B. (2019). Klebsiella phage KP34 RNA polymerase and its use in RNA synthesis. *Frontiers in Microbiology*, 10, 2487.
- Zhu, B., Tabor, S., Raytcheva, D. A., Hernandez, A., King, J. A., & Richardson, C. C. (2013). The RNA polymerase of marine cyanophage Syn5. *Journal of Biological Chemistry*, 288, 3545–3552.
- Xia, H., Yu, B., Jiang, Y., Cheng, R., Lu, X., Wu, H., & Zhu, B. (2022). Psychrophilic phage VSW-3 RNA polymerase reduces both terminal and full-length dsRNA byproducts in in vitro transcription. *RNA Biology*, 19, 1130–1142.
- Mu, X., Greenwald, E., Ahmad, S., & Hur, S. (2018). An origin of the immunogenicity of in vitro transcribed RNA. *Nucleic Acids Research*, 46, 5239–5249.
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47, 636–641.
- Steinberger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35, 1026–1028.
- Lavigne, R., Sun, W. D., & Volckaert, G. (2004). PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics*, 20, 629–635.
- Sampaio, M., Rocha, M., Oliveira, H., Dias, O., & Valencia, A. (2019). Predicting promoters in phage genomes using PhagePromoter. *Bioinformatics*, 35, 5301–5302.
- Rong, M., Biao, H., McAllister, W. T., & Durbin, R. K. (1998). Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 515–519.
- Conrad, T., Plumbom, I., Alcobendas, M., Vidal, R., & Sauer, S. (2020). Maximizing transcription of nucleic acids with efficient T7 promoters. *Communications Biology*, 3, 439.
- Deich, C., Cash, B., Sato, W., Sharon, J., Aufdembrink, L., Gaut, N. J., Heili, J., Stokes, K., Engelhart, A. E., & Adamala, K. P. (2023). T7Max transcription system. *Journal of Biological Engineering*, 17, 1–14.
- Sari, Y., Sousa Rosa, S., Jeffries, J., & Marques, M. P. C. (2024). Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production. *Scientific Reports*, 14, 9655.
- Wang, G., Cheng, R., Chen, Q., Xu, Y., Yu, B., Zhu, B., Yin, H., & Xia, H. (2022). mRNA produced by VSW-3 RNAP has high-level translation efficiency with low inflammatory stimulation. *Cell Insight*, 1, 100056.
- Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. *Frontiers in Bioengineering and Biotechnology*, 9, 630551.
- Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in *Escherichia coli*. *Protein Journal*, 32, 419–425.
- Egorova, T., Shuvalova, E., Mukba, S., Shuvalov, A., Kolosov, P., & Alkalaeva, E. (2021). Method for rapid analysis of mutant rna polymerase activity on templates containing unnatural nucleotides. *International Journal of Molecular Sciences*, 22, 5186.
- Cui, M., Wong, O., Li, Q., & Wang, W. (2023). An assay method for characterizing bacteriophage T7 RNA polymerase activity by transcription-translation (TX-TL) system. *Journal of Biochemistry*, 173, 343–352.
- Shujaat, M., Jin, J. S., Tayara, H., & Chong, K. T. (2022). iProm-phage: A two-layer model to identify phage promoters and their types using a convolutional neural network. *Frontiers in Microbiology*, 13, 1061122.



36. Zhu, B., Hernandez, A., Tan, M., Wollenhaupt, J., Tabor, S., & Richardson, C. C. (2015). Synthesis of 2'-fluoro RNA by Syn5 RNA polymerase. *Nucleic Acids Research*, 43, e94–e94.
37. Weber, J. S., Carlino, M. S., Khattak, A., Meniawy, T., Ansstas, G., Taylor, M. H., Kim, K. B., McKean, M., Long, G. V., Sullivan, R. J., Faries, M., Tran, T. T., Cowey, C. L., Pecora, A., Shaheen, M., Segar, J., Medina, T., Atkinson, V., Gibney, G. T., ..., Zaks, T. (2024). Individualised neoantigen therapy mRNA-4157 (V940) plus pembrolizumab versus pembrolizumab monotherapy in resected melanoma (KEYNOTE-942): A randomised, phase 2b study. *The Lancet*, 403, 632–644.
38. Nelson, J., Sorensen, E. W., Mintri, S., Rabideau, A. E., Zheng, W., Besin, G., Khatwani, N., Su, S. V., Miracco, E. J., Issa, W. J., Hoge, S., Stanton, M. G., & Joyal, J. L. (2020). Impact of mRNA chemistry and manufacturing process on innate immune activation. *Science Advances*, 6, 6893.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Curry, E., Sedelnikova, S., Rafferty, J., Hulley, M., Pohle, M., Muir, G., & Brown, A. (2024). Expanding the RNA polymerase biocatalyst solution space for mRNA manufacture. *Biotechnology Journal*, 19, e2400012.  
<https://doi.org/10.1002/biot.202400012>