

De novo RNA synthesis from free nucleotides by a family of conserved plant proteins

Ruonan Li^{1#}, Changzeng Yu^{1#}, Mingyue Li¹, Lingsong Gu¹, Xiechao Zhan^{*1}, Yigong Shi^{*1}

RNA synthesis generally requires a pre-existing nucleic-acid template and sometimes, an oligonucleotide primer. Here we demonstrate that a plant protein named OsNTP4, a nucleotidyltransferase-like enzyme from *Oryza sativa*, directly synthesizes single-stranded RNA of varying lengths from free nucleoside triphosphates (NTPs) in the absence of any template or primer. The RNA products have the canonical 3'-5' phosphodiester linkages and display sequence features. The synthetic activity requires divalent metal ions. Importantly, the related rice protein OsNTP10 and the *Arabidopsis thaliana* protein AtURT1 possess the same activity of *de novo* RNA synthesis, and together with OsNTP4 define a family of conserved plant enzymes. These findings reveal a direct protein-to-RNA reaction, which may have implications for our understanding of evolution and the origin of life.

<https://doi.org/10.15302/vita.2026.10.0078>

INTRODUCTION

In cells, RNA synthesis is generally coupled to a pre-existing nucleic-acid substrate. DNA-dependent or RNA-dependent polymerase uses a DNA or RNA template to specify the sequence of the RNA product, respectively¹⁻⁴; whereas terminal nucleotidyltransferases extend the 3'-ends of existing RNAs⁵⁻⁷. These reactions have established principles for chain initiation, nucleotide selection, and processive elongation^{3,6,7}. Recent studies have expanded the known repertoire of protein-directed nucleic-acid synthesis, including protein-primed and protein-templated DNA synthesis^{8,9}. However, whether a protein can assemble an RNA polymer with 3'-5' phosphodiester linkage from free nucleoside triphosphates (NTPs), without an added nucleic-acid primer or template, remains unknown.

If such a reaction could be found, it would be expected to be different from all known pathways of RNA synthesis^{2,10}, modification^{11,12}, and turnover¹³. Two known classes of enzymes catalyze related but clearly different reactions. One is the terminal nucleotidyltransferase⁵⁻⁷, which acts on a pre-existing single-stranded RNA end, thereby regulating RNA maturation, stability, and decay^{10,12,14}. The other is the polynucleotide phosphorylase, which synthesizes RNA through reverse phosphorolysis, but uses nucleoside diphosphates and different reaction chemistry^{15,16}. *De novo* RNA synthesis from NTPs would require an enzyme to initiate and extend an RNA chain with 3'-5' linkages using nucleotide monomers only.

Identification of such a reaction faces at least two challenges. First, remote evolutionary relationships are often obscured at the sequence level. Second, direct proof of *de novo* nucleic-acid synthesis by a protein can be confounded

by protein-associated nucleic acids or trace amounts of co-purifying enzymes. With the advent of AlphaFold, structure prediction and structure-based searches now offer a route around the first challenge by recognizing conserved catalytic architectures independently of sequence similarity¹⁷⁻¹⁹. The second challenge can be mitigated by a combination of biochemical purification, mass spectrometry (MS), site-directed mutagenesis, and orthogonal enzymatic assays on protein recovered from crystals. We used these strategies to search for proteins with previously unrecognized nucleotide-processing potential. The human enzyme cyclic GMP-AMP synthase (cGAS) was used as the sole structural query because its nucleotidyltransferase core constitutes a well-defined architecture for nucleotide binding and phosphodiester-bond formation²⁰⁻²².

In this study, we report three plant proteins that possess the activity of *de novo* RNA synthesis: OsNTP4 and OsNTP10 from *Oryza sativa*, and AtURT1 from *Arabidopsis thaliana*. In all three cases, the reaction proceeds in the absence of primer or template and yields RNA products with the canonical 3'-5' phosphodiester linkages. Although these proteins are already known to belong to the terminal nucleotidyltransferase family²³⁻²⁵, they are not known to possess the ability of *de novo* RNA synthesis. These proteins define a family of plant enzymes, which may have conserved family members in other species.

RESULTS

Structure-guided identification of OsNTP4

To our knowledge, cGAS or cGAS-like protein has yet to be reported in plants²⁶. We therefore sought plant proteins that may display structural homology to cGAS but with distinct

1. State Key Laboratory of Gene Expression, Key Laboratory of Structural Biology of Zhejiang, School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China. [#]These authors contributed equally to this work. ^{*}Correspondence: Xiechao Zhan (zhanxiechao@westlake.edu.cn), Yigong Shi (syg@westlake.edu.cn)

Received: September 14, 2026; Accepted: October 15, 2026; Published: October 20, 2026

nucleotide-processing activities. Toward this goal, we used human cGAS as a structural query for DALI-based searches of AlphaFold-predicted protein structures (Fig. 1a). We focused on three plant species: a model dicot *Arabidopsis thaliana*, and two major monocots *Oryza sativa* and *Zea mays*. Among the initial hits, 266 proteins, including 76 from rice, 90 from *Arabidopsis*, and 100 from maize, display structural similarity Z-score of 2.0 or higher (Fig. 1a, b). We further narrowed this set to 29 proteins, including 13 candidates from *Arabidopsis*, 10 from rice, and 6 from maize, each exhibiting strong structural similarity to human cGAS with a Z-score of 10.0 or higher (Fig. 1a, c).

We next asked whether these 29 proteins retain common sequence features within their putative catalytic regions. Human cGAS has an acidic catalytic triad comprising Glu225–Asp227–Asp319²⁷, whereas the bacterial cGAS/DncV-like enzyme DncV from *Vibrio cholerae* uses Asp131, Asp133, and Asp193²⁸. Substitution between Glu and Asp preserves an acidic environment while allowing divergence in local active-site geometry. We therefore prioritized conservation of the three acidic residues in sequence comparison and eliminated 10 proteins that violate this rule. This operation yielded 19 plant proteins for further analysis (Fig. 1a, d).

Alignment of the corresponding regions revealed two conserved acidic sequence blocks among these 19 candidates (Fig. 1e). The first block corresponds to the region containing Glu225 and Asp227 of human cGAS. Whereas cGAS carries an E ϕ D ϕ sequence in this region, the plant candidates contain a conserved D ϕ D ϕ motif, where ϕ is predominantly Ile, Leu, or Val. The second block corresponds to the region containing Asp319 of human cGAS and contains a conserved D ϕ motif in the plant proteins (Fig. 1e). Thus, despite divergence in residue identity at the first catalytic position, the selected candidates retain a three-residue acidic configuration corresponding to the catalytic triad of cGAS.

Pairwise sequence comparison among the 19 candidates led to identification of four subgroups (Fig. 1f). Despite an average of only ~15% sequence identity with human cGAS (Fig. 1d), these 19 candidates show much higher pairwise sequence identity within each subgroup, generally ranging from 40% to 80%. Pairwise structural comparisons broadly recapitulate the sequence-based organization (Fig. 1g). Proteins within the major sequence-related subgroups generally display higher pairwise Z-scores than those from different subgroups. This analysis suggests functional diversification within these candidates.

We focused on the rice protein OsNTP4 (Os3; UniProt: AOA0P0VE25). The AlphaFold model of OsNTP4 can be superimposed to the X-ray structure of human cGAS (PDB: 4KM5)²⁹ with an overall root-mean-square deviation (RMSD) of 4.623 Å over 1,182 atoms (Fig. 1h). Within the conserved active-site region, the two structures are aligned with an RMSD of 0.735 Å over 66 atoms. Consistent with the sequence alignment, the three acidic residues Asp100, Asp102, and Asp157 of OsNTP4 map precisely to the catalytic residues Glu225, Asp227, and Asp319 of human cGAS, respectively (Fig. 1h).

OsNTP4 converts free NTPs into RNase-sensitive polymers

To examine the potential enzymatic activity of OsNTP4, we

overexpressed and purified the recombinant full-length protein to homogeneity through three steps: affinity column, ion exchange chromatography, and size exclusion chromatography (Fig. 2a; Supplementary Fig. S1). The purified OsNTP4 at 5 μ M concentration was incubated with free NTPs (2 mM each for ATP, GTP, CTP, and UTP) for 12 h at 37 °C. The reaction products and the remaining nucleotides were subjected to fractionation by anion exchange chromatography using a Capto Q column. In the absence of OsNTP4, the substrate NTPs were eluted from the column by about 260–320 mM NaCl, corresponding to elution volumes of 13–16 mL (Fig. 2b). In the presence of OsNTP4, a fraction of the substrate NTPs was consumed as reflected by their diminished peaks, especially for UTP and CTP (Fig. 2c). Notably, a new peak, which is presumably the reaction product, appeared at about 680 mM NaCl, corresponding to an elution volume of ~30 mL (Fig. 2c). As shown by denaturing gel electrophoresis, the reaction product is a mixture of oligomers of varying sizes, with the majority in the range of 50–300 nucleotides (Fig. 2d).

We then examined whether OsNTP4 had any substrate preference for each of the four nucleotides. 2 mM UTP was completely consumed by OsNTP4, producing a discrete peak at the elution volume of ~30 mL (Fig. 2e). More than two-thirds of 2 mM CTP was consumed, producing a product peak at ~26 mL (Fig. 2f). About half of 2 mM ATP was converted to a broad peak with an elution volume of ~23 mL (Fig. 2g). In contrast to the other three nucleotides, only a very small fraction of GTP was consumed by OsNTP4 (Fig. 2h). The reaction products were resolved by electrophoresis on a denaturing gel (Fig. 2i). The average length for the reaction products is the longest for UTP as the substrate and shortest for GTP. The reaction products from CTP were stained weakly by SYBR Gold, limiting a direct comparison of their apparent lengths with those of the other products (Fig. 2i). Treatment by RNase I, which has broad substrate specificity, resulted in complete digestion of the reaction products from the substrate nucleotides UTP, CTP, and ATP (Fig. 2j). The majority of the reaction products from GTP were digested. The low yield, relatively short lengths, and slow digestion by RNase I of the GTP-derived reaction products may reflect the formation of peculiar structures, which in this case are most likely the G-quadruplexes. Consistent with this analysis, our reaction buffer includes 150 mM KCl, which is known to stabilize G-quadruplexes³⁰. Therefore, OsNTP4 produces RNase-sensitive nucleotide polymers from free NTPs.

We next examined whether the catalytic activity depended on the three conserved acidic residues. To this end, we generated three missense variants: D100A, D102A, and D157A, and individually purified all three variants to homogeneity (Supplementary Fig. S2a–c). Each of the three mutations resulted in abrogation of the catalytic activity of OsNTP4 (Supplementary Fig. S2a–c). Mutation of a surface residue Lys194 to Ala has little effect on the catalytic activity of OsNTP4 under the same assay conditions (Supplementary Fig. S2d).

The reaction product is characterized by 3'–5' phosphodiester linkages

Sensitivity to RNase treatment suggests the reaction products to be RNA polymers. But the chemical identity of the

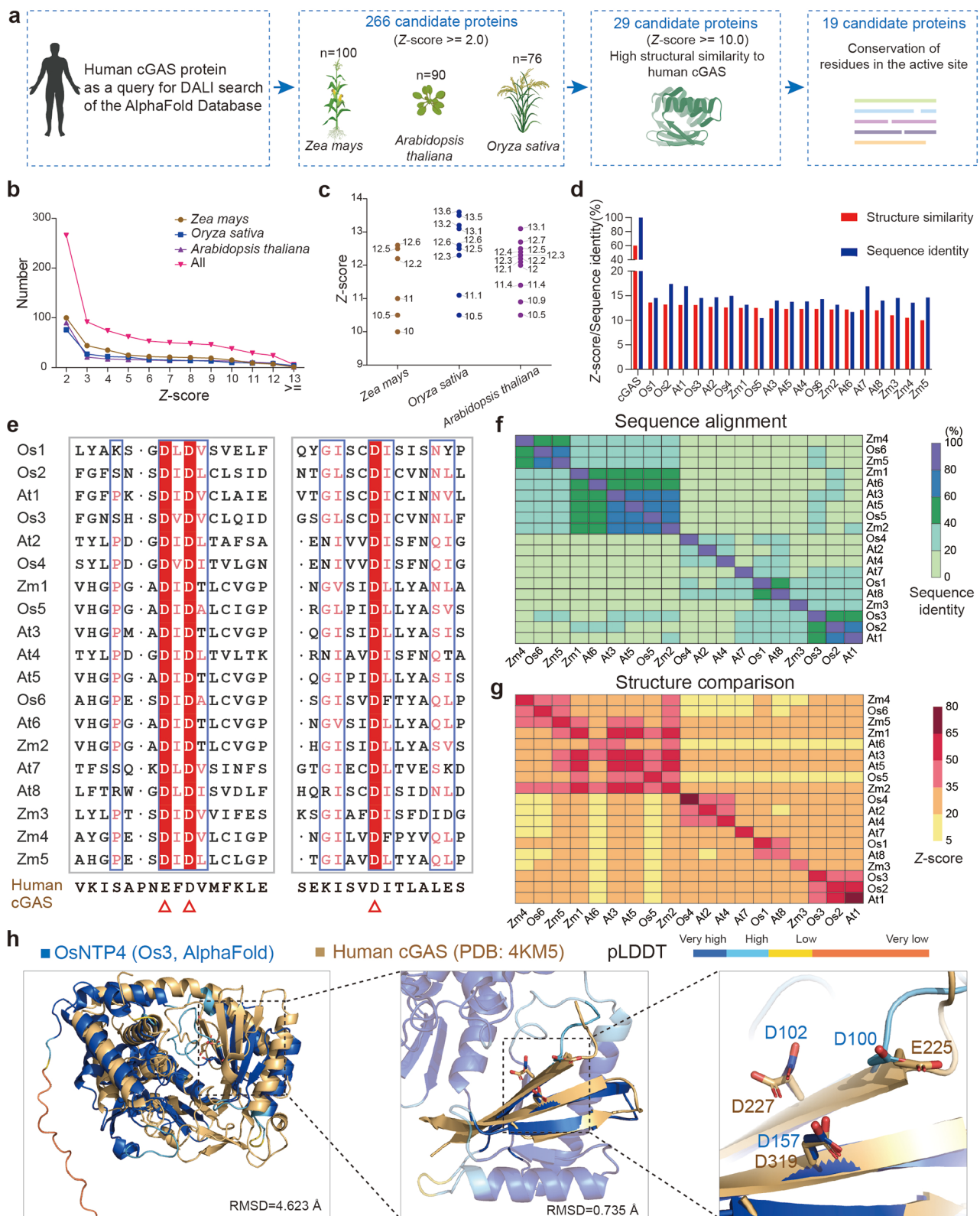


Fig. 1 Structure-guided identification of OsNTP4. **a** A structure-guided discovery pipeline. Human cGAS was used as a DALI query to search AlphaFold-predicted protein structures from *Zea mays*, *Arabidopsis thaliana*, and *Oryza sativa*. The search identified 266 structural candidates at a Z-score threshold of ≥ 2.0 . Filtering for proteins with Z-scores ≥ 10.0 yielded 29 candidates, and subsequent selection for conservation of three acidic residues corresponding to the cGAS catalytic region yielded 19 final candidates. This diagram was prepared using BioGDP⁴⁶. **b** The distribution of Z-scores among the candidate proteins. **c** The distribution of Z-scores of 10.0 or higher among the 29 candidate proteins, grouped by species. **d** Comparison of structural similarity to human cGAS, measured by Z-score (red), and sequence identity to human cGAS (blue) for the selected 19 candidates. **e** Alignment of the sequence regions encompassing the two acidic blocks that correspond to the catalytic regions of human cGAS. Red triangles mark the positions of Glu225, Asp227, and Asp319 in human cGAS. Conserved acidic

residues in the plant candidates are highlighted. **f** Pairwise comparison of sequence identity for the 19 selected plant candidates. **g** Pairwise comparison of structural similarity for the selected 19 plant candidates, calculated as Z-scores. **h** Superposition between the AlphaFold-predicted structure of OsNTP4 (Os3; colored by pLDDT) and the X-ray structure of human cGAS (PDB: 4KM5; brown). The global superposition yielded an RMSD of 4.623 Å over 1,182 aligned atoms, whereas the locally conserved catalytic core region has an RMSD of 0.735 Å over 66 aligned atoms. Enlarged views show the corresponding acidic residues in OsNTP4 (Asp100, Asp102, and Asp157) and human cGAS (Glu225, Asp227, and Asp319).

internucleotide linkage remained unclear. We therefore analyzed homopolymeric products generated from UTP, ATP, or CTP by nuclear magnetic resonance (NMR) and compared each product with the corresponding synthetic poly-(U), poly-(A), or poly-(C) reference RNA, each 50 nucleotides. The ^1H , ^1H COSY spectrum of the UTP-derived product displayed the anticipated ribose proton correlations (Fig. 3a, b). The ^1H , ^{31}P HMBC spectrum identified the linkages between the bridging phosphate and both the H3' proton of one ribose and the H5' proton of the adjacent ribose, confirming the P-O3' and P-O5' bonds that define a canonical 3'-5' phosphodiester linkage^{31,32} (Fig. 3c, d). A 2'-5' linkage would instead be expected to yield an H2'-P correlation instead of the H3'-P correlation. Equivalent COSY and HMBC correlations were observed for the ATP- and CTP-derived products (Supplementary Figs. S3, S4). These correlations are diagnostic of a 3'-5' phosphodiester linkage and demonstrate that OsNTP4 synthesizes RNA with the canonical 3'-5' backbone connectivity.

We next examined the requirement for divalent metal ions. No reaction product was detected in reactions lacking both Mg^{2+} and Mn^{2+} (Supplementary Fig. S5a). Addition of Mg^{2+} , Mn^{2+} , or both ions, enabled product formation, but with distinct substrate preferences (Supplementary Fig. S5b-d). The nucleotide UTP was efficiently converted to poly-(U) product in the presence of either Mg^{2+} or Mn^{2+} . But the synthesis of poly-(A) and poly-(G), and to a lesser extent poly-(C), appears to require the presence of Mn^{2+} (Supplementary Fig. S5b-d). Thus, OsNTP4-mediated RNA synthesis requires a divalent metal co-factor and is modulated by the chemical identity of the nucleotide substrate.

Together with the loss of activity caused by substitution of Asp100, Asp102, or Asp157, these results support a model in which the conserved acidic residues and a coordinated divalent metal promote formation of canonical RNA phosphodiester bonds. The present data establish the chemical identity of the OsNTP4 product, while leaving the molecular mechanism of chain initiation and nucleotide selection for future investigation.

RNA polymerization activity is intrinsic to OsNTP4

It is not trivial to prove that the observed RNA polymerization activity is intrinsic to OsNTP4 itself, namely, OsNTP4 does not need any nucleic-acid primer, any nucleic-acid template, or any associated protein or peptide co-factor. The catalytic domain of OsNTP4 (OsNTP4_CD, residues 36-362) was biochemically purified through three consecutive steps (Supplementary Fig. S6), each designed to remove contaminants and associated proteins or co-factors. Using MS, we determined the molecular weight of OsNTP4_CD to be 37,681 Daltons (Fig. 4a), which agrees well with its predicted molecular mass. The same MS analysis was also performed on the full-length OsNTP4 protein with similarly affirmative result

(Supplementary Fig. S7). These results indicate that OsNTP4 is free of any covalently bound co-factor.

Next, we examined whether the purified OsNTP4 protein contains any free nucleotide or oligonucleotide. We used Proteinase K to completely digest 200 μM OsNTP4_CD in a volume of 50 μL . Free nucleotides and oligonucleotides were extracted and analyzed by sequential Mono S and Capto Q column. In parallel, we subjected each of the four nucleotides (A, G, C, U), at 200 μM each in 50 μL , to be processed identically. The results are unequivocal: each free nucleotide was faithfully detected; but no trace amount of any nucleotide or any other co-factor was detected in the OsNTP4_CD sample (Fig. 4b). These data demonstrate that the purified OsNTP4_CD protein contains no detectable amount of free NTPs or RNA oligonucleotides.

Finally, we verified the intrinsic activity of OsNTP4 using protein crystallization. We crystallized OsNTP4_CD using the hanging-drop vapor-diffusion method. The crystals were harvested, extensively washed using the well buffer, and examined for the RNA synthesis activity in the presence of 2 mM UTP (Fig. 4c). OsNTP4_CD from the dissolved crystals generated a UTP-dependent high-molecular-weight product, whereas the matched noncrystalline control from the same drop failed to do so (Fig. 4d-f). Product formation was detected by dual-color fluorescence imaging and independently confirmed by Cy5 and SYBR Gold staining (Fig. 4d-f). Collectively, these orthogonal analyses argue against nucleotide carryover, covalent nucleotide modification, or a separable co-purifying activity as the source of the reaction, supporting the conclusion that RNA polymerization activity is intrinsic to OsNTP4.

De novo RNA synthesis by OsNTP10 and AtURT1

Given the unanticipated finding on OsNTP4, we speculated that *de novo* RNA synthesis may be shared by a subfamily of plant nucleotidyltransferases. To search for additional members of this subfamily, we examined sequence and structural alignment of the initially identified 19 candidate proteins (Fig. 1f, g). Two proteins OsNTP10 (Os2; UniProt: AOA0POXSC2) and AtNTP3 (At1; UniProt: O64642) share particularly high sequence and structural identities with OsNTP4. Encouraged by this result, we broadened the search by performing phylogenetic analysis on all 47 known nucleotidyltransferases from *Oryza sativa*, *Arabidopsis thaliana*, and *Zea mays*. The analysis revealed five major clades (Supplementary Fig. S8). Remarkably, OsNTP4, OsNTP10, and AtNTP3 are all assigned to Clade I, which also include two additional members from maize. All five proteins in Clade I share a conserved C-terminal nucleotidyltransferase core, including the conserved acidic residues (Supplementary Fig. S9). We over-expressed and purified OsNTP10 and AtNTP3 and examined their activity. Notably, AtNTP3 is also named AtURT1, which has been structurally resolved^{33,34}.

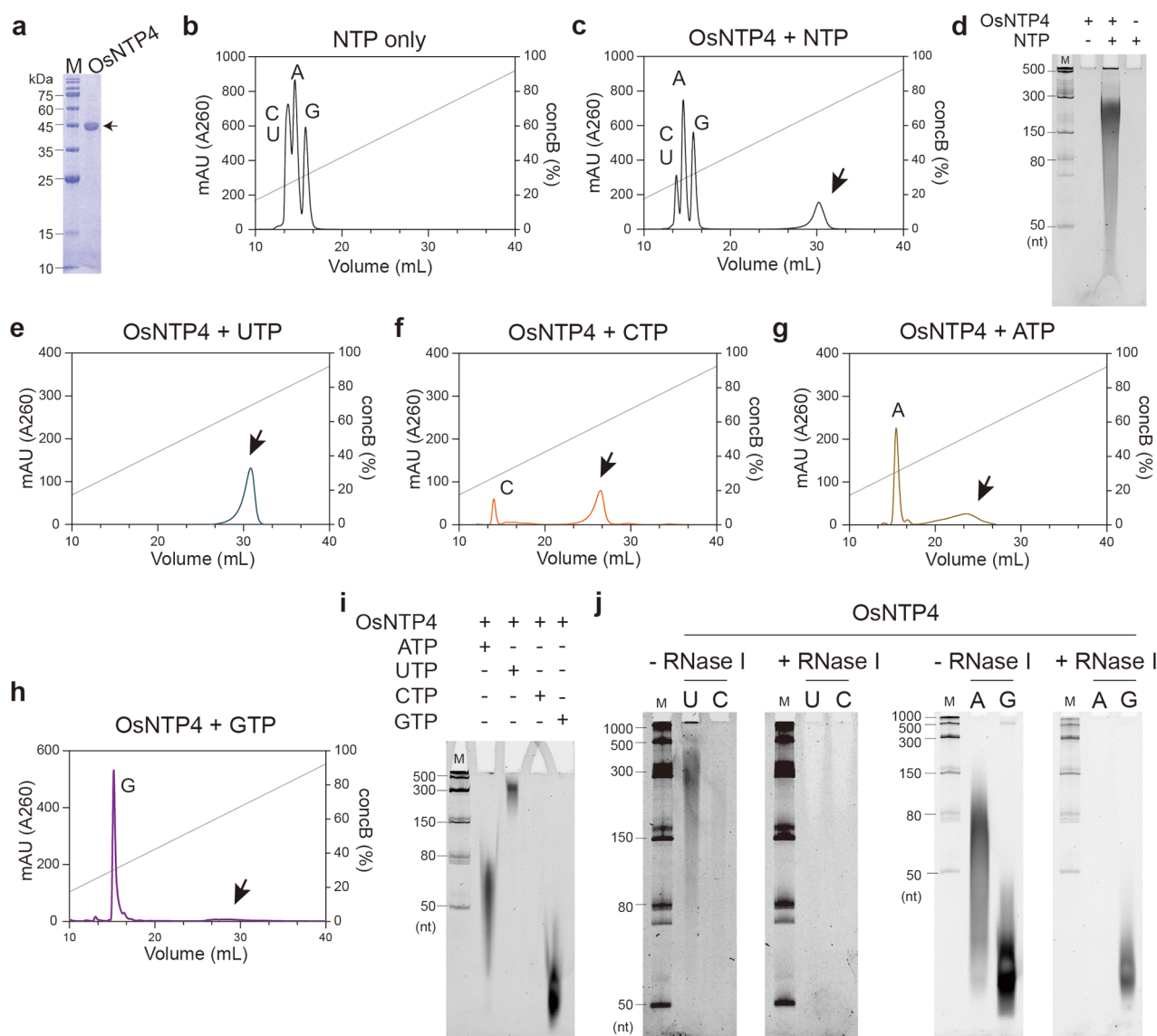


Fig. 2 OsNTP4 generates RNase-sensitive polymers from free NTPs. **a** The purified full-length OsNTP4 protein. Shown here is an SDS-PAGE gel stained by Coomassie blue. **b** No reaction product is detectable in the absence of OsNTP4. Shown here is the ion-exchange chromatogram of the control reaction that contained NTP only. **c** Production of reaction products by OsNTP4 in the presence of NTPs. Shown here is the ion-exchange chromatogram of the reaction containing purified OsNTP4 and all four NTPs. The reaction product is indicated by the arrowhead. **d** Analysis of the reaction products by denaturing gel electrophoresis. High-molecular-weight products are detected only in the reaction containing both OsNTP4 and NTPs. **e** Production of reaction products by OsNTP4 in the presence of UTP. The reaction product is indicated by the arrowhead. **f** Production of reaction products by OsNTP4 in the presence of CTP. **g** Production of reaction products by OsNTP4 in the presence of ATP. **h** Production of reaction products by OsNTP4 in the presence of GTP. **i** Denaturing gel analysis of the reaction products generated with individual NTP substrates. **j** The reaction products are sensitive to RNase I treatment. Products were analyzed before and after RNase I treatment.

We first examined OsNTP10 (Fig. 5a; Supplementary Fig. S10). In reactions lacking added RNA or DNA, OsNTP10 generated RNA polymers in the presence of ATP, UTP, or CTP, as shown by the well-defined peaks for the reaction products (Fig. 5b–d). The use of UTP was the most efficient among the three nucleotides. Similar to OsNTP4, OsNTP10 also consumed GTP, but the reaction products appeared as a shallow hump in the chromatogram of Capto Q column (Fig. 5e). Upon electrophoresis, the reaction products from ATP, UTP, or CTP were detected as high-molecular-weight species on denaturing gels, which were completely digested by RNase I

(Fig. 5f). In contrast, the reaction products from GTP appeared as RNA polymers of less than 50 nucleotides, which were partially degraded by RNase I (Fig. 5f). We next examined the chemical linkage in the UTP-derived homopolymeric product using NMR analysis. The COSY and HMBC correlations are diagnostic of a 3′–5′ phosphodiester linkage, establishing that the UTP-derived OsNTP10 product contains the canonical RNA backbone connectivity (Fig. 5g). Therefore, OsNTP10 has the same biochemical activity as OsNTP4 and converts free NTPs into RNase-sensitive RNA polymers.

We next tested AtURT1, previously characterized as an RNA

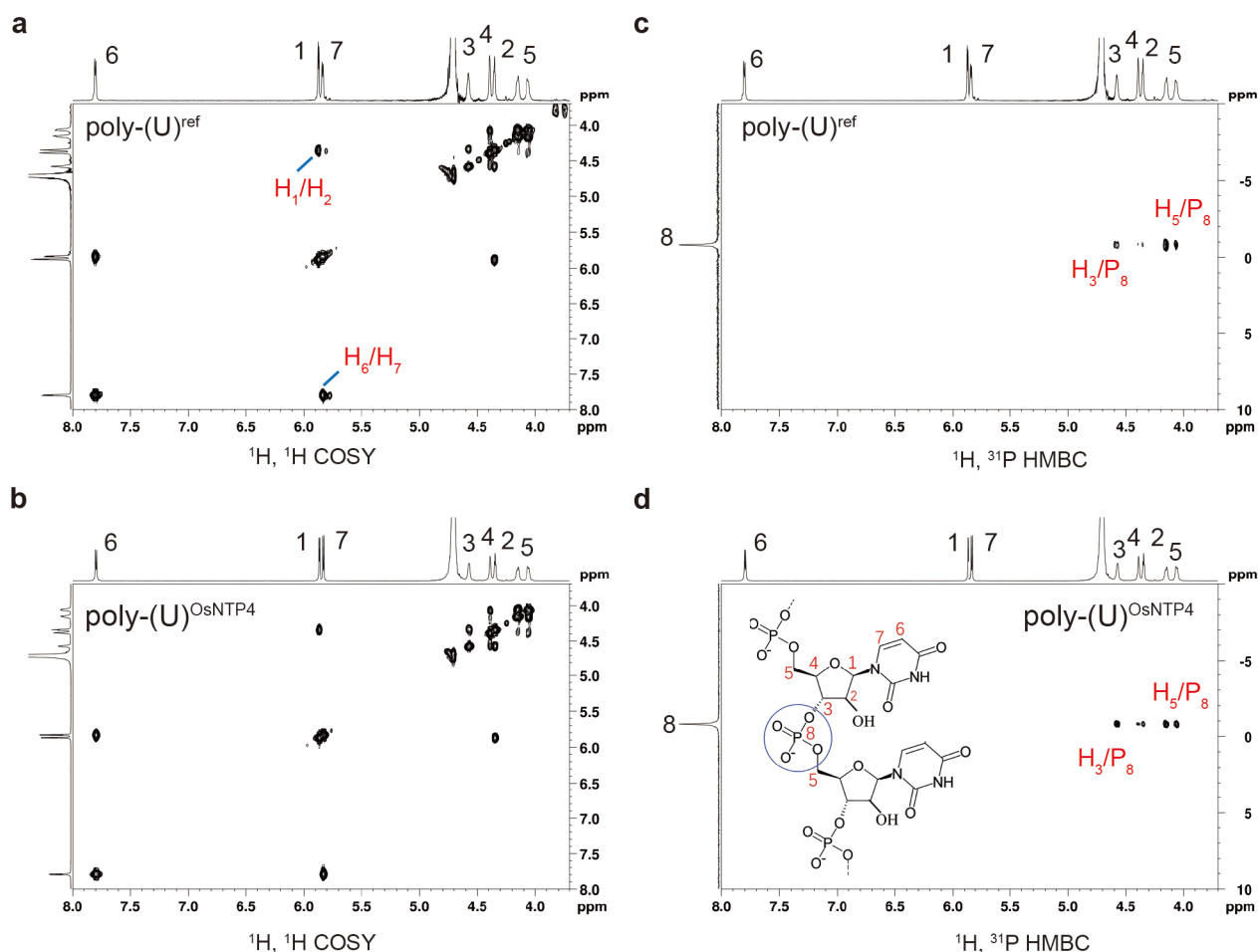


Fig. 3 The OsNTP4 reaction products, RNA polymers from UTP only, have a canonical 3'-5' phosphodiester linkage as revealed by NMR analysis. **a** The ^1H , ^1H COSY spectra of a reference poly-(U) sample. **b** The ^1H , ^1H COSY spectra of the reaction products of OsNTP4 in the presence of UTP only. **c** The ^1H , ^{31}P HMBC spectra of a reference poly-(U) sample. **d** The ^1H , ^{31}P HMBC spectra of the reaction products of OsNTP4 in the presence of UTP only. Red labels identify correlations between the bridging phosphate (P8) and H3' and H5' ribose protons, indicating a 3'-5' phosphodiester linkage.

3'-end uridylyl transferase³⁵. The catalytic domain of AtURT1 (AtURT1_CD, residues 410–764) was purified to homogeneity following three purification steps (Fig. 6a; Supplementary Fig. S11). In the presence of all four nucleotides, AtURT1_CD generated reaction products that were eluted at about 30 mL on Capto Q column (Fig. 6b). These RNA polymers display a size range from less than 50-nt to more than 150-nt (Fig. 6c). In the presence of 2 mM UTP, AtURT1_CD generated a range of reaction products with distinct elution volumes (Fig. 6d). AtURT1_CD also exhibited activities for all other three nucleotides, with GTP generating shorter RNA oligomers than UTP or ATP, whereas CTP-derived products could not be conclusively assessed because of their weak staining (Fig. 6e). These results show that AtURT1 also exhibits *de novo* RNA synthesis activity in the absence of a primer, a template, or any added nucleic-acid substrate.

We then compared *de novo* RNA synthesis with extension of a pre-existing RNA. Consistent with previous reports^{34,35}, AtURT1 extended a Cy5-labeled U15 RNA substrate in an NTP-dependent manner (Supplementary Fig. S12a), OsNTP4 extended a Cy5-labeled MiR-158a RNA (Supplementary Fig. S12b), and OsNTP10 similarly extended Cy5-labeled U15 RNA (Supplementary Fig. S12c). The three proteins displayed

different extension patterns with individual NTPs, indicating distinct nucleotide preferences. Together, these results show that both *de novo* RNA synthesis from free NTPs and RNA 3'-end tailing are shared activities of the three plant NTP proteins.

Nucleotide features of OsNTP4-generated RNAs

OsNTP4 displays clear substrate preferences for *de novo* synthesis of RNA. We characterized the NTP-derived products generated by OsNTP4. The RNA products were first polyadenylated using poly-(A) polymerase. The polyadenylated RNA was then subjected to reverse transcription and strand switching, followed by barcoded PCR amplification and rapid sequencing adapter attachment using an Oxford Nanopore cDNA-PCR barcoding workflow. The resulting libraries were sent to nanopore sequencing. The sequenced RNA products were analyzed for sequence features. Because of the uncharacterized consequences of the conformation and length of diverse RNA products on the multiple steps of treatment before DNA sequencing, the reported sequence features only faithfully reflect those of the sequenced DNA molecules, not the original intact RNA polymers.

The sequenced RNA products display a broad length distri-

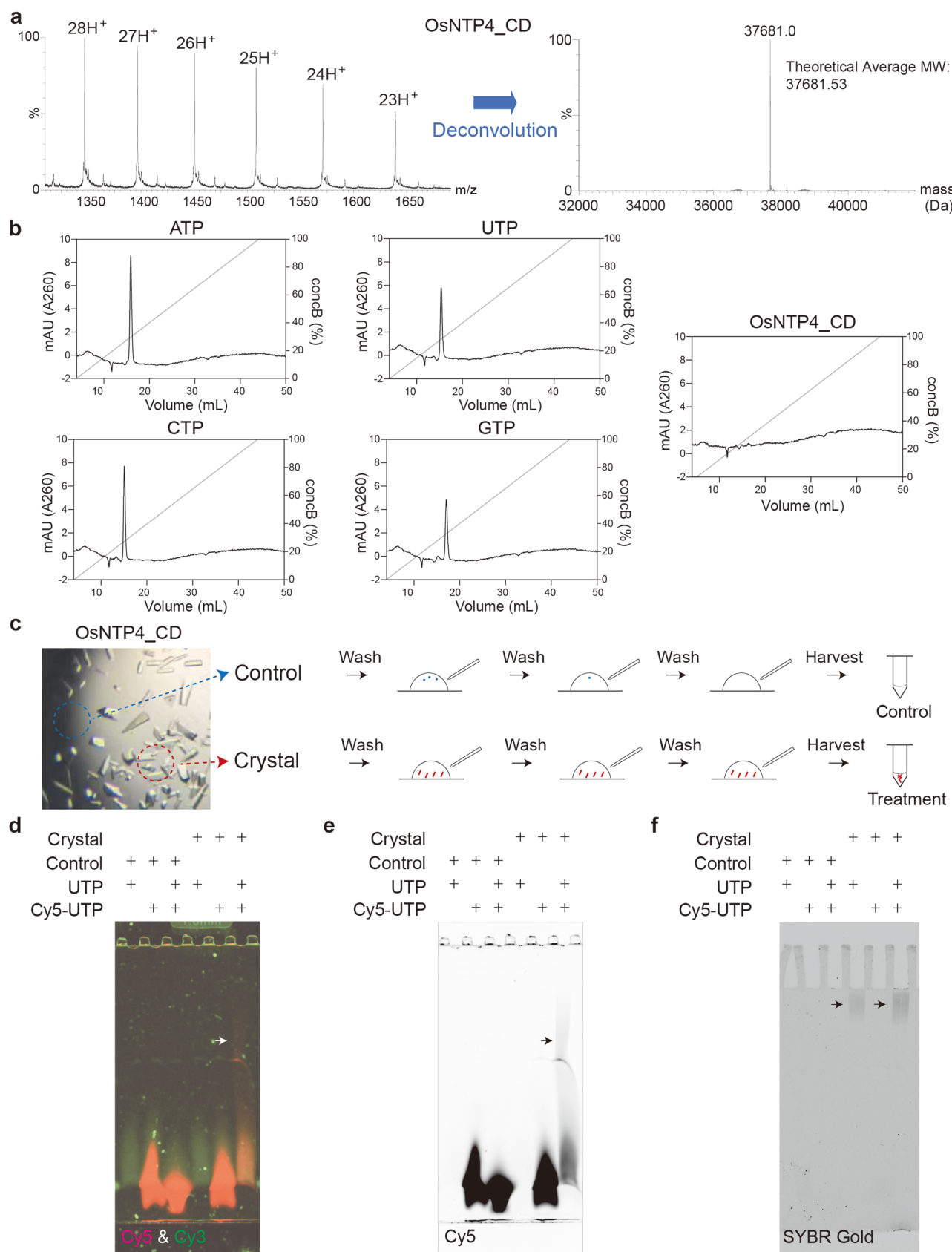


Fig. 4 RNA polymerization activity is intrinsic to OsNTP4. **a** The recombinant OsNTP4_CD protein contains no covalently linked nucleotide or any other cofactor. Shown here is a mass spectrum of OsNTP4_CD which indicates a mass of 37,681.0 Da, consistent with the calculated mass of 37,681.53 Da. **b** The recombinant OsNTP4_CD protein contains no detectable free nucleotides. Proteinase K-treated samples were analyzed by sequential Mono-S and Capto-Q ion-exchange chromatography. ATP, UTP, CTP, and GTP standards were identically processed as

OsNTP4_CD. The nucleotide standards yielded distinct peaks, whereas no corresponding nucleotide peak was detected in the OsNTP4_CD sample. **c** A workflow diagram for harvesting OsNTP4_CD crystals. The crystals and noncrystalline materials of equal volumes were collected from the same drop, washed three times, and harvested for activity assays. **d** OsNTP4_CD from the dissolved crystals exhibited *de novo* RNA synthesis activity. Shown here is a dual-color fluorescence analysis of UTP-dependent product formation by the dissolved crystals and noncrystalline materials (control). **e** Cy5 fluorescence analysis of the same reactions as in panel (d). **f** SYBR Gold staining of the reaction products. Arrows indicate high-molecular-weight RNA products.

bution, ranging from approximately 20 to 593 nucleotides, with 11 sequences exceeding 400 nucleotides in length (Fig. 7a). Sequence complexity analysis reveals variation in the overall nucleotide composition and local sequence organization of the products, as assessed by Shannon entropy, linguistic complexity, purine-to-pyrimidine compositional ratio, and dinucleotide standard deviation (Fig. 7b). The Shannon entropy values range from 1.0 to 2.0 bits, with a median of 1.2 bits, indicating that the sequences are generally not dominated by a single nucleotide and retain appreciable overall compositional diversity. Linguistic complexity values are relatively low (0.007–0.125), with a paired median difference of –0.012 relative to the null model, suggesting that the observed sequences exhibited a lower diversity of local short-sequence patterns. The compositional ratio (purine-to-pyrimidine) displays a median $\log_2$ value of –2.536, corresponding to an approximately 5.8-fold excess of pyrimidines over purines, indicating a strong overall preference for pyrimidine, which, together with the nucleotide composition analysis, was primarily attributable to U enrichment. Dinucleotide standard deviation values are uniformly low (approximately 0.09–0.16), closely resembling those obtained from the null model and indicating limited dinucleotide-specific organization. Together, these analyses reveal that the products are neither simple homopolymers nor compositionally random: they exhibit a pronounced pyrimidine-associated preference and reduced local subsequence complexity.

Analysis of nucleotide composition shows that U is the most abundant base (56.0%) among the sequenced RNA products, whereas C, G, and A are present at intermediate frequencies of 13.4–16.4% (Fig. 7c). Homopolymer tracts are a recurrent feature of the sequenced RNA products, with distinct length distribution of the homopolymers of the four nucleotides (Fig. 7d). The nucleotide length of the longest homopolymer for U, A, G, and C are 125, 49, 30, and 24, respectively (Fig. 7d). Motif discovery analysis identified two recurrent sequence motifs beyond simple homopolymer tracts (Fig. 7e). Among the non-homopolymeric 6-mers, GGGUU (motif 1) and UCCUCU (motif 2) are present 260 and 250 times, respectively, in the sequenced RNA products. These motifs provide a descriptive signature of the sequenced product pool. Together, these preliminary data show that OsNTP4 produces a diverse set of RNA products with varying length, base composition, and motif features. This preliminary analysis provides a starting point for future experiments that may faithfully determine the features of the RNA products.

DISCUSSION

Our results identify OsNTP4 as a plant protein that generates RNA polymers from free NTPs in the absence of a nucleic-acid primer, a nucleic-acid template, or any other added nucleic-acid substrate. OsNTP4 does not need to be activated by any protein or nucleic-acid co-factor. Its catalytic activity requires

divalent metal ions. Importantly, the RNA polymers have the canonical 3'–5' phosphodiester linkages characteristic of RNA. Together, these findings define an NTP-to-RNA reaction that differs from established nucleotide polymerization pathways (Fig. 8). Template-dependent polymerases require a nucleic-acid template^{1–4}, whereas terminal nucleotidyltransferases extend pre-existing RNA 3' ends^{5–7}. Polynucleotide phosphorylase catalyzes the reverse phosphorolysis reaction using NDPs^{15,16}. In contrast, cGAS and 2'–5'-oligoadenylate synthetase (OAS) require activation by dsDNA or dsRNA, respectively, and produce specific oligonucleotide signals rather than RNA polymers^{20,22,36} (Fig. 8). OsNTP4 therefore defines a unique free-NTP-to-RNA reaction that is distinct from that of any of these enzyme classes. We further demonstrate that the rice protein OsNTP10 and the *Arabidopsis* protein AtURT1 both have the same activity as OsNTP4. Collectively, these findings show that a family of plant proteins can catalyze RNA synthesis from free nucleotide substrates and expand the known biochemical routes by which RNA can be generated³⁷.

The RNA-world hypothesis proposes that RNA plays a central role in early information storage and catalysis^{38,39}. RNA can encode heritable sequence information while also adopting structures that support diverse catalytic and regulatory functions. In this study, we demonstrate *de novo* synthesis of RNA by proteins using only nucleotide substrates. This activity expands the known biochemical repertoire of RNA polymer formation. Importantly, the RNA products display some sequence features, although the level of base preference and sequence specificity fall short of what would be expected for specific RNA sequences. Nonetheless, the diversity of RNA sequences itself is a source of functional selection and perhaps evolution. Enzymes that ligate and/or exchange RNA fragments further increase their functional diversity^{40,41}. Furthermore, there may be proteins capable of making very specific RNA oligonucleotides.

We used several orthogonal approaches to examine whether the observed activity is intrinsic to OsNTP4, because this is central to the claim of *de novo* RNA synthesis. The OsNTP4 preparation contained no detectable free nucleotide, and intact-protein MS further revealed no covalent adduct. Importantly, extensively washed OsNTP4 crystals retained *de novo* RNA synthesis, while matched noncrystalline material did not. These experiments strongly argue that the activity originates from OsNTP4 rather than from detectable nucleotide carryover or a separable co-purifying activity.

Compared with ATP-, UTP-, and CTP-derived products, the reaction with GTP as the sole substrate produced a substantially weaker signal and predominantly short species that migrated below ~50 nt on denaturing PAGE. These products were only partially sensitive to RNase I treatment. Their low abundance currently precludes further structural characterization by NMR. Nonetheless, the RNA products from GTP

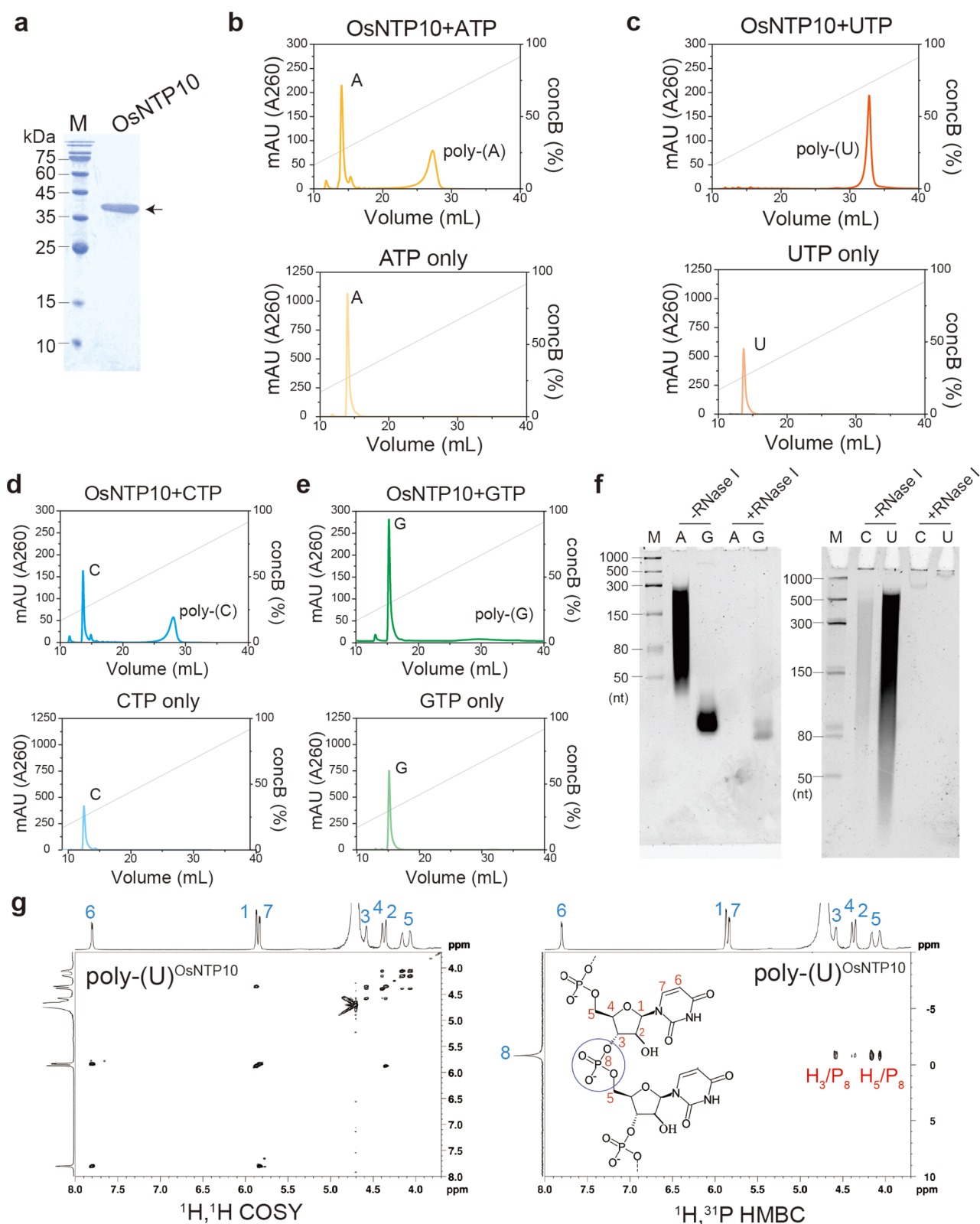


Fig. 5 OsNTP10 generates RNase-sensitive RNA polymers from free NTPs. **a** The purified full-length OsNTP10 protein. Shown here is an SDS-PAGE gel stained by Coomassie blue. **b** Production of reaction products by OsNTP10 in the presence of ATP. Shown here are the ion-exchange chromatograms of the reaction in the presence (upper panel) or absence (lower panel) of OsNTP10. **c** Production of reaction products by OsNTP10 in the presence of UTP. **d** Production of reaction products by OsNTP10 in the presence of CTP. **e** Production of reaction products by OsNTP10 in the presence of GTP. **f** The reaction products are sensitive to RNase I treatment. Products were analyzed on denaturing gels before and after RNase I treatment. **g** The reaction products of OsNTP10, poly-(U) polymers, have a canonical 3'-5' phosphodiester linkage as revealed by NMR analysis. Shown here are the ^1H , ^1H COSY spectra (left panel) and ^1H , ^{31}P HMBC spectra (right panel) of the reaction products.

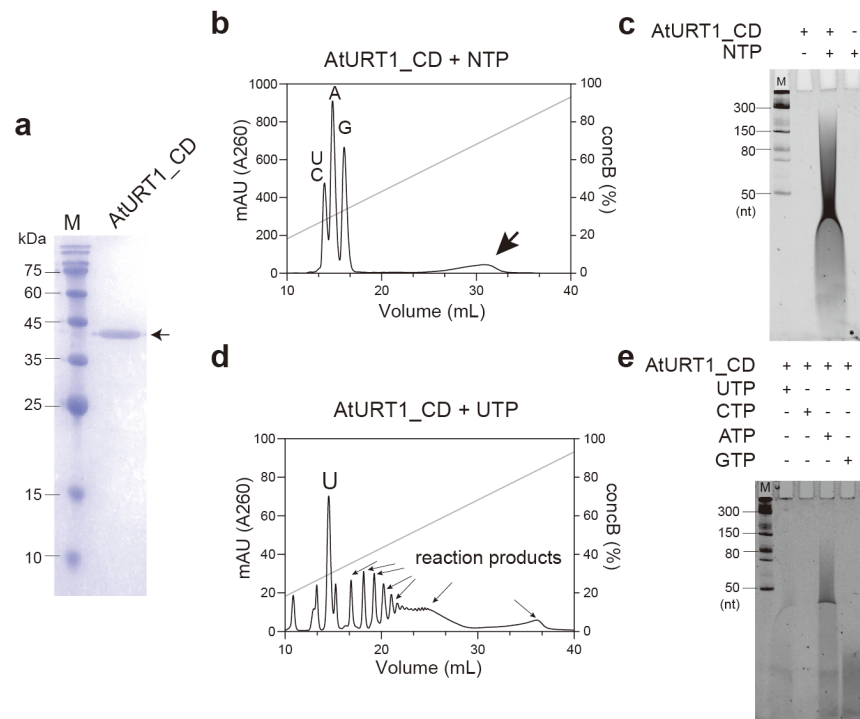


Fig. 6 AtURT1 catalytic domain generates RNA polymers from free NTPs. **a** The purified AtURT1_CD protein. Shown here is an SDS-PAGE gel stained by Coomassie blue. **b** Production of reaction products by AtURT1_CD in the presence of NTPs. Shown here is the ion-exchange chromatogram of the reaction containing purified AtURT1_CD and all four NTPs. The reaction product is indicated by the arrowhead. **c** Analysis of the reaction products by denaturing gel electrophoresis. **d** Production of reaction products by AtURT1_CD in the presence of UTP. The reaction products appear as multiple distinct peaks on the ion-exchange chromatogram. **e** Denaturing gel analysis of the reaction products generated with individual NTP substrates.

alone most likely have the same canonical 3'-5' phosphodiester linkages as those from other nucleotides. First, the majority of the GTP-derived RNA products are sensitive to treatment by RNase I, which cleaves the 3'-5' phosphodiester bonds through a transesterification mechanism^{42,43}. Second, in the sequencing results of NTP-derived RNA products, the guanine nucleotide is incorporated together with the other three nucleotides most likely through the same catalytic chemistry. Why are the GTP-derived RNA products shorter than those from other nucleotides and relatively resistant to RNase I treatment? The reason may be that poly-(G) forms G-quadruplexes, which have a stable conformation and may negatively affect RNA synthesis and digestion. Notably, the reaction buffer in our assay contains 150 mM KCl, which is known to stabilize formation of the G-quadruplex³⁰. Thus, G-quadruplex formation provides a plausible explanation for both the shorter products and their partial nuclease resistance.

While this manuscript was under submission, we noted the publication of a study reporting primer- and template-independent RNA polymerization by human TENT4B⁴⁴. Although the overall conclusions are similar between our study and the published study⁴⁴, there are important differences. In the absence of MS or NMR analysis, the human protein TENT4B was suggested to synthesize poly-(A) and poly-(G) polymers in the absence of a primer or template. We used a structure-guided search to identify plant candidates with a cGAS-related catalytic scaffold. We detected *de novo* RNA synthesis by OsNTP4 in April 2025 and took time to prove the activ-

ity being intrinsic to OsNTP4 through an array of biochemical and biophysical methods. We also established the 3'-5' phosphodiester linkage for the reaction products by NMR. Last but not least, each of the three plant enzymes in our study has been proven to be capable of synthesizing RNA oligomers of diverse sequences. Nonetheless, the parallel observation of potentially *de novo* RNA synthesis in a mammalian terminal nucleotidyltransferase highlights the potential breadth of this unexpected activity.

Recent studies of defense-associated reverse transcriptases (DRTs) provide important precedents for noncanonical protein-directed DNA synthesis. DRT9 uses an associated noncoding RNA as a template and protein residues as putative primers to generate poly-(dA) sequence, constituting protein-primed but RNA-templated DNA synthesis⁸. In contrast, Drt3b produces a protein-primed poly-(dAdC) strand in the absence of a nucleic-acid template, with active-site residues directing alternating nucleotide selection⁹. These systems demonstrate that proteins can contribute to unusual modes of DNA priming and sequence specification, but they differ from our study in polymer identity, nucleotide substrate, and product chemistry. Under the same conditions in which RNA is *de novo* synthesized, OsNTP4, OsNTP10, and AtURT1 failed to generate a detectable amount of DNA products from dNTPs. Therefore, our study defines a distinct RNA-level activity: plant nucleotidyltransferases can generate RNA polymers of diverse sequences from free NTPs without an added nucleic-acid primer or template.

De novo RNA synthesis by OsNTP4 depends on the

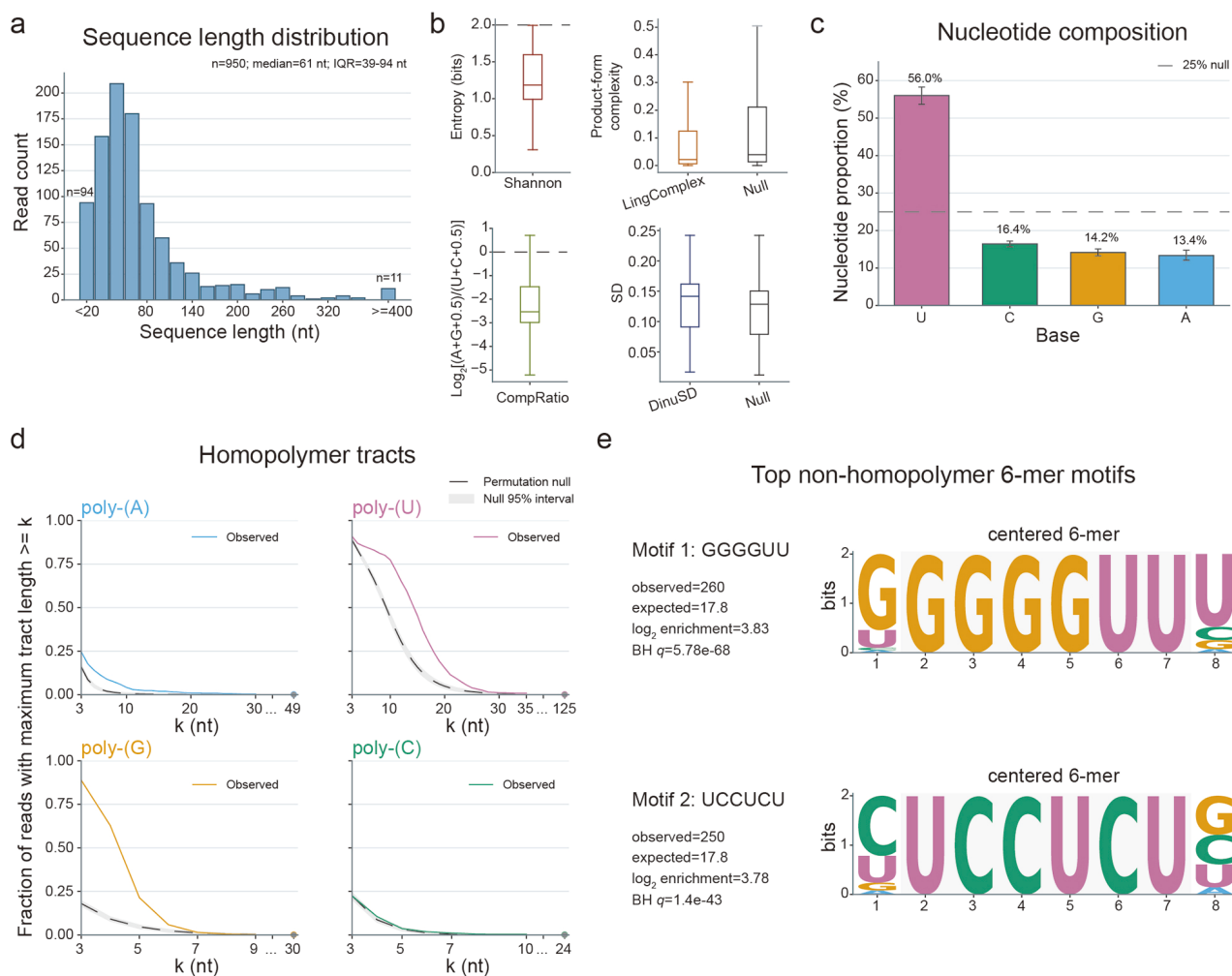


Fig. 7 Sequence features of RNA products derived from OsNTP4. **a** Length distribution of the sequenced RNA products. Shown here is the length distribution of 950 RNA sequences. Following addition of poly(A) tails, the reaction products of OsNTP4 using NTPs were subjected to reverse transcription, tagging, PCR, and sequencing. **b** The sequenced RNA products display sequence diversity. Shown here are four indicators of sequence complexity for the sequenced RNA products. These four indicators are: Shannon entropy, linguistic complexity (LingComplex), purine-to-pyrimidine compositional ratio (CompRatio), and the standard deviation of normalized dinucleotide frequencies (DinuSD). Boxes indicate the interquartile range, center lines indicate medians, and whiskers extend to 1.5 times the interquartile range. The dashed line indicates the uniform zero-order null expectation of 25% for each nucleotide. **c** Nucleotide composition of the sequenced RNA products. **d** Length distribution of homopolymers in the sequenced RNA products. Shown here are the length distributions for U, A, C, and G (colored lines, from left to right). As the control, the dashed gray lines and shaded regions show the mean and 95% interval under a random permutation of all nucleotides that preserve the length of each RNA sequence and mononucleotide counts. **e** Two most strongly enriched non-homopolymeric 6-mer motifs identified among the RNA sequences. Observed counts, permutation-derived expected counts, $\log_2$ enrichment values, and Benjamini-Hochberg (BH)-adjusted q values are shown.

conserved acidic residues Asp100, Asp102, and Asp157 and on divalent metal ions. This requirement is consistent with a metal-assisted phosphoryl-transfer mechanism organized by the acidic catalytic pocket⁴⁵. The distinct product profiles obtained with ATP, UTP, CTP, and GTP further indicate that the active site does not treat all NTPs equivalently. Consistent with these observations, analysis of the AlphaFold-predicted model identifies a charged pocket surrounding the conserved catalytic residues that may accommodate an incoming NTP (Supplementary Fig. S13a). We also used AlphaFold to predict two additional structures of OsNTP4: bound to an RNA penta-nucleotide alone and bound to an RNA penta-nucleotide in the presence of UTP (Supplementary Fig. S13b, c). Placement of UTP and a short RNA oligonucleotide within the active site region positions the RNA 3'-OH near the UTP α -phosphate, suggesting a geometrically plausible configuration for phosphodiester-bond formation (Supplementary Fig. S13d). Despite these speculations, important mechanistic questions remain to be resolved. In particular, it is not yet known how OsNTP4 initiates the first phosphodiester bond, whether initiation requires a transient dinucleotide intermediate, how the enzyme controls product length, or how product release is coupled to continued synthesis. Structural analysis of OsNTP4 bound to substrates, reaction intermediates or RNA products will be required to address these questions.

The activity is not confined to OsNTP4. OsNTP10, a second rice NTP-family protein, also generates nucleotide-depen-

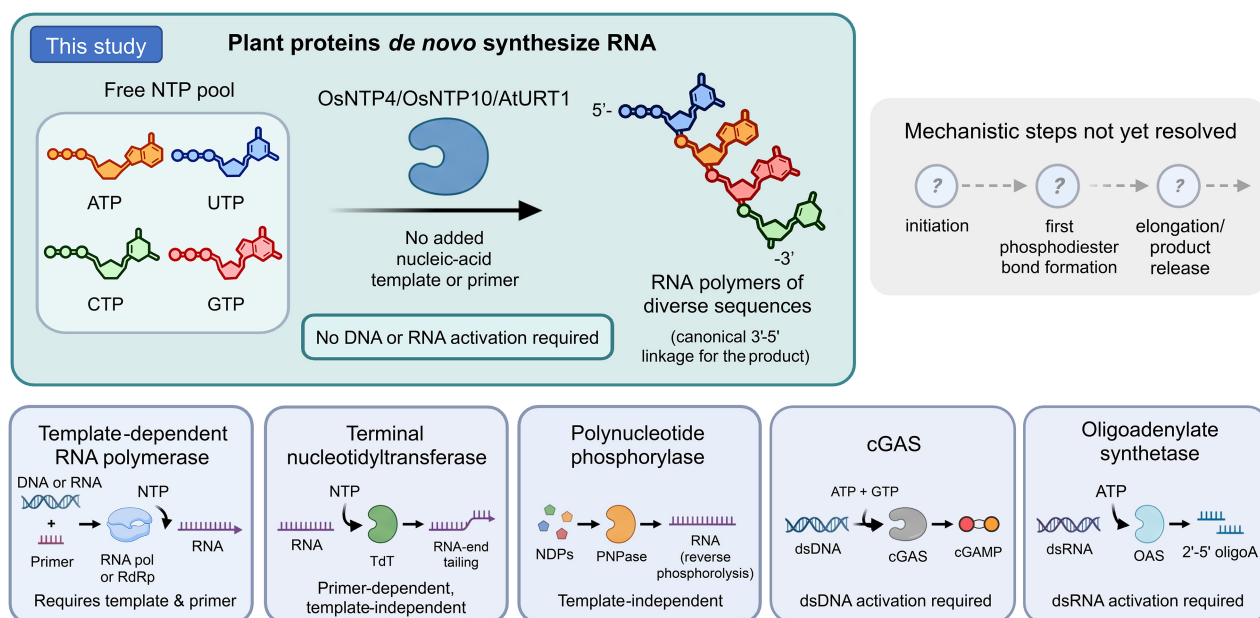


Fig. 8 A summary of the plant enzymes capable of *de novo* RNA synthesis and their comparison with other related enzymes. OsNTP4, OsNTP10, and AtURT1 generate RNA polymers of diverse sequences from free NTPs in reactions lacking an added nucleic-acid primer or template (upper panel). These RNA polymers contain the canonical 3'-5' phosphodiester linkages. The molecular steps underlying initiation, first phosphodiester-bond formation, elongation, and product release remain to be resolved. This reaction contrasts with that by any of the other five established nucleotide-processing enzyme classes (lower panel). Template-dependent RNA polymerases use NTPs and require both a nucleic-acid template and primer. Terminal nucleotidyltransferases use NTPs to extend pre-existing RNA 3' ends without a template. Polynucleotide phosphorylase synthesizes RNA from NDPs by reverse phosphorolysis. cGAS uses ATP and GTP in a dsDNA-activated reaction to produce cGAMP, whereas OAS uses ATP in a dsRNA-activated reaction to generate 2'-5' oligoadenylates.

dent polymeric products in the absence of added RNA or DNA. AtURT1, previously characterized as an RNA 3'-end uridylyl transferase^{34,35}, likewise supports *de novo* RNA synthesis under these conditions. OsNTP4 also extends a pre-existing RNA substrate, demonstrating that it retains RNA-tailing activity. These results suggest that RNA-tailing and direct RNA synthesis are alternative catalytic outputs of a related plant nucleotidyltransferase scaffold. Differences in NTP-dependent product profiles among OsNTP4, OsNTP10, and AtURT1 indicate that this shared capacity has diversified rather than being mechanistically identical in every member.

The structure-guided search also points to a broader, largely uncharacterized landscape of plant NTP proteins. In addition to OsNTP4, OsNTP10, and AtURT1, the search recovered 16 additional candidate proteins with detectable structural similarity to cGAS and conserved acidic residues in the putative catalytic region (Fig. 1). Phylogenetic analysis further resolved the NTP-family proteins from the three plant species into five major clades (Supplementary Fig. S8). The present data do not establish *de novo* RNA synthesis for these additional candidates: structural similarity and conservation of an acidic carboxylate configuration nominate them for functional analysis but do not define their substrates, products, or biological roles. Systematic characterization across clades will be needed to determine whether direct RNA synthesis is restricted to a subset of Clade I proteins or represents a more broadly diversified activity within the plant NTP family. Notably, preliminary sequencing analysis further showed that OsNTP4 produces a heterogeneous RNA population with broad distributions of length, base-composition, and

homopolymer. These data provide an initial description of the products made in a mixed-NTP reaction.

The biological function of direct RNA synthesis by plant NTP proteins remains open. The present study establishes biochemical capacity, but does not identify endogenous RNA products or define the cellular conditions under which this reaction occurs. Resolving these questions will require analysis of catalytically inactive plant alleles, identification of OsNTP4-associated RNAs *in vivo*, and direct measurement of RNA species that depend on OsNTP4 or related proteins. Nonetheless, the discovery that three plant NTP proteins can generate RNA from free NTPs expands the functional repertoire of nucleotidyltransferases. It is very likely that such enzymes are also present in species other than plants. OsNTP4 provides a tractable starting point for defining the mechanism, regulation, and possible biological consequences of this unexpected route from protein to RNA.

MATERIALS AND METHODS

Plasmid construction

The codon-optimized coding sequences of the full-length *Oryza sativa* NTP4 (OsNTP4; residues 1–367), OsNTP10 (residues 1–372), and *Arabidopsis thaliana* URT1 (AtURT1; residues 1–764) were commercially synthesized and cloned into pET-28a (+) vectors with N-terminal 6×His tag. The DNA fragments encoding the AtURT1 catalytic domain (AtURT1_CD; residues 410–764) and the OsNTP4 catalytic domain (OsNTP4_CD; residues 36–362) were amplified from the respective full-length plasmids by PCR. AtURT1_CD was cloned into pET-28a (+) with an N-terminal 6×His tag.

OsNTP4_CD was cloned into pET-29a (+) with an N-terminal 8×His-SUMO tag. All plasmids were verified by DNA sequencing.

Protein expression and purification

Expression plasmids were transformed into *Escherichia coli* BL21(DE3). Cells were cultured in LB medium at 37 °C to an optical density at 600 nm of approximately 1.0. Protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), followed by incubation at 18 °C for 16 h. Harvested cells were resuspended in the lysis buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 0.5 mM DTT, lysed by sonication, and clarified by centrifugation at 20,000× g for 1 h at 4 °C. The supernatant was applied to a Ni-NTA column (QIAGEN, 30230) equilibrated in the lysis buffer. The column was washed using 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20 mM imidazole, and 0.5 mM DTT. Bound proteins were eluted using 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 300 mM imidazole, and 0.5 mM DTT. The eluted proteins were diluted two-fold using 25 mM Tris-HCl, pH 8.0 and 5 mM DTT. The samples were further purified on a HiTrap Heparin HP column (Cytiva, 17040701) equilibrated with 25 mM Tris-HCl, pH 8.0 and 5 mM DTT. Bound proteins were eluted with a linear NaCl gradient to 1 M NaCl in the same buffer. Fractions containing the target protein were pooled and concentrated before size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva, 17517501) equilibrated using 25 mM Tris-HCl, pH 8.0, 350 mM NaCl, and 5 mM DTT. Peak fractions were pooled for biochemical assays.

For OsNTP4_CD, the eluate from Ni-NTA column was diluted two-fold using 25 mM Tris-HCl, pH 8.0, and 5 mM DTT, and applied to a Source 15Q chromatography (Cytiva, 90100465). The bound protein was eluted using a linear gradient of 0–1 M NaCl in 25 mM Tris-HCl, pH 8.0 and 5 mM DTT. The N-terminal 8×His-SUMO fusion tag was removed by incubation with 6×His-tagged 3C protease at 42 μg protease per mg target protein for 30 min at room temperature. The reaction mixture was applied to Ni-NTA resin to remove the cleaved tag, uncleaved fusion protein, and the 6×His-tagged protease. Tag-free OsNTP4_CD in the flow-through was further purified by a second step of Source 15Q chromatography (Cytiva, 90100465). Finally, the peak fractions were concentrated and applied to size-exclusion chromatography (SEC) on a Superdex 200 Increase 10/300 GL column (Cytiva, 17517501) equilibrated in 25 mM Tris-HCl, pH 8.0, 350 mM NaCl and 5 mM DTT. Peak fractions from SEC were pooled and concentrated for crystallization.

In vitro RNA-polymerization assay and anion-exchange analysis

Unless otherwise indicated, each RNA-polymerization reaction was assembled in a volume of 100 μL, which contained 5 μM purified protein, 2 mM each of four NTPs (ATP, GTP, CTP, and UTP), 100 mM HEPES, pH 7.5, 150 mM KCl, 2.5 mM MgCl₂ and 2 mM MnCl₂. Reactions were incubated at 37 °C for 12 h. For assays with individual nucleotides, ATP, UTP, CTP, or GTP was added separately at 2 mM. As a negative control, the reaction was carried out in the absence of added protein.

Due to prolonged incubation at 37 °C, some proteins in the reaction mixture precipitated out of solution. Each reaction

was clarified by centrifugation. The supernatant was diluted to 500 μL using RNase-free water and separated on a Capto HiRes Q 5/50 anion-exchange column (Cytiva, 29275878) equilibrated with 25 mM Tris-HCl, pH 8.0. Products were eluted with a 0–100% linear gradient of buffer B containing 25 mM Tris-HCl, pH 8.0 and 1 M NaCl at a flow rate of 0.8 mL/min over 40 mL. The reaction products were monitored at a wavelength of 260 nm.

RNA 3'-tailing assay

The RNA-tailing reaction was performed in a volume of 100 μL, which contained 5 μM purified protein (OsNTP4, OsNTP10, or AtURT1_CD), 500 nM 5'-Cy5-labeled RNA substrate, 100 mM HEPES, pH 7.5, 150 mM KCl, 2.5 mM MgCl₂, 2 mM MnCl₂, and, where indicated, 2 mM ATP, UTP, CTP, or GTP. As a negative control, the reaction was carried out in the absence of added nucleotides. Samples were incubated at 37 °C for 2 h, clarified by centrifugation, mixed with an equal volume of 2× RNA loading buffer, and analyzed by denaturing urea-polyacrylamide gel electrophoresis (urea-PAGE).

Denaturing urea-PAGE and fluorescence imaging

The RNA products were analyzed on 8% or 12% denaturing polyacrylamide gels containing 8 M urea. Gels were run in 1× TBE at 250 V for 20 or 50 min, as appropriate for gel percentage and product size. For total nucleic-acid detection, gels were stained with SYBR Gold for 5 min and imaged using UV excitation at 302 nm. Cy5-labeled RNA products were visualized with an Amersham Imager 680. Cy3 and Cy5 signals were acquired in the 520 and 630 nm channels, respectively.

NMR analysis

All NMR experiments were performed at 25 °C on a Bruker NEO 800 MHz NMR spectrometer equipped with a QCI-P Cryoprobe. For 1D ¹H and ³¹P experiments, 64k complex data points were acquired with 16 and 128 scans, respectively. Data of 2D ¹H, ¹H COSY employing the “cosygpprqf” pulse sequence were collected with 4 scans × 2048 data points (F2) × 160 increments (F1). Data of 2D ¹H, ³¹P HMBP using the “hmbcgpplndqf” pulse sequence were collected with 8 scans × 4096 data points (F2) × 128 increments (F1).

Crystallization of OsNTP4_CD

OsNTP4_CD crystals were grown at 20 °C by the hanging-drop vapor-diffusion method. The protein solution and the well buffer were mixed in an equal-volume fashion. The well buffer contained 17% (w/v) PEG 4000, 0.2 M Tris-HCl, pH 7.4, 0.1 M MgCl₂ and 0.1 M ammonium acetate.

RNA-polymerization assay using OsNTP4_CD crystals

The hanging drop that contained crystals was diluted by 20 μL well buffer. Eight aliquots, each 0.36 μL of the crystal-containing solution from the diluted hanging drop were manually collected. The crystals were washed three times using the well buffer and transferred to microcentrifuge tubes. Matched control aliquots were collected from crystal-free regions of the same drop and processed identically.

Crystal-derived and control samples were incubated in 15 μL reactions containing 100 mM HEPES, pH 7.5, 150 mM KCl,

2.5 mM MgCl₂, 2 mM MnCl₂, 2 mM UTP, 50 nM Cy5-UTP (B8333, APE × BIO) and RNase-free water. The reaction was allowed to proceed at 37 °C for 6 h. Products were resolved on 12% denaturing urea-PAGE and imaged on an Amersham Imager 680 in the Cy5, Cy3 and SYBR Gold channels.

Intact-protein MS

Intact-protein mass analysis was performed using a Waters UPLC system coupled to a Xevo G2 time-of-flight mass spectrometer with an electrospray ionization source operated in positive-ion mode. Protein samples were separated on a BEH C4 column (50 × 2.1 mm, 1.7 μm) at 40 °C. Mobile aqueous phase A contained 0.1% formic acid, and mobile organic phase B was acetonitrile containing 0.1% formic acid. The flow rate was 0.3 mL/min. The gradient was held at 5% B for 1 min, increased from 5% to 95% B over 4 min, and held at 95% B for 2 min.

Each sample was injected at 1 μL. Data were acquired over m/z 200–3,000 using the following source settings: capillary voltage, 3 kV; sample cone voltage, 50 V; source offset, 30 V; source temperature, 140 °C; desolvation temperature, 600 °C; nebulizer pressure, 35 psi; cone-gas flow, 50 L/h; and desolvation-gas flow, 800 L/h. Data were acquired and deconvoluted using MassLynx version 4.2.

Free nucleotide detection in OsNTP4_CD

Purified OsNTP4_CD (200 μM) was incubated with Proteinase K (0.1 mg/mL) in a final volume of 50 μL at 37 °C for 30 min. In parallel, four nucleotide-standard controls were prepared under identical conditions, each containing 200 μM of a single nucleotide standard (A, U, C, or G) and 0.1 mg/mL Proteinase K in a final volume of 50 μL. Each reaction was diluted with 440 μL RNase-free water, followed by the addition of 10 μL 1 M sodium acetate and 1 μL glacial acetic acid. Samples were vortexed to adjust the final pH to approximately 3.0 and centrifuged at 12,000× g for 10 min at 4 °C. The supernatants were collected and loaded onto a Mono S column (Cytiva, 17516801) pre-equilibrated with buffer A (25 mM Tris-HCl, pH 8.0). The flow-through fractions were subsequently loaded directly onto a Capto Q column (Cytiva, 29275878) pre-equilibrated with buffer A. Bound species were eluted using a linear gradient from 0% to 100% buffer B (25 mM Tris-HCl, pH 8.0, 1 M NaCl). Absorbance was monitored continuously at 260 and 280 nm.

Phylogenetic and sequence analyses

Protein sequences of the NTP-family members from *A. thaliana*, *O. sativa* and *Z. mays* were retrieved from Ensembl Plants⁴⁷. Phylogenetic analysis was performed in MEGA version 7.0.26⁴⁸. The maximum-likelihood tree was generated using the Jones-Taylor-Thornton substitution model with gamma-distributed rate variation across five categories⁴⁹. Gaps and missing sites were treated by partial deletion with a 95% site-coverage cutoff. The analysis included 47 amino acid sequences and 80 aligned positions. Bootstrap support was estimated from 50 replicates, and tree searching was performed with the nearest-neighbor-interchange heuristic.

For sequence comparison, amino-acid sequences were aligned by ClustalW in MEGA version 7.0.26 using default settings⁵⁰. Sequence conservation and secondary-structure

annotations were rendered with ESPrnt version 3.2⁵¹.

Sequence analysis for the OsNTP4 reaction products

The reaction mixture containing 5 μM OsNTP4 and 2 mM NTP was incubated at 37 °C for 12 h. The reaction mixture was centrifuged at 12,000× g for 10 min, and the supernatant was collected. The reaction products in the supernatant were purified using the RNA clean beads (NOVOPROTEIN, N243-01B). Following RNA quantitation, an aliquot of 5 μg purified RNA products was incubated with 2.5 units of the *E. coli* Poly-(A) Polymerase (M0276S, NEB) and 1 mM ATP at 37 °C for 30 min. The poly-(A)-added RNA products were re-purified using the RNA clean beads and used for library preparation using the cDNA-PCR Sequencing V14-Barcoding kit (SQK-PCB114.24). Sequencing was performed on a PromethION 2 Solo (Nanopore) sequencer.

Sequence pattern analysis

Reads were oriented to the RNA sense-strand direction before analysis. The barcode, sequencing adapter, VNP, SSP, and detected terminal poly-(A) sequences were removed. The final dataset contained 950 sequences. Sequence lengths were summarized in 20-nt bins, with all sequences longer than 400 nt pooled into a single > 400-nt category. Sequence complexity was quantified using Shannon entropy,

$H = -\sum_i p_i \log_2 p_i$; linguistic complexity, calculated as the product of the observed-to-maximum possible distinct k -mer ratios for $k = 1 - 6$; the purine-to-pyrimidine compositional ratio, $\log_2 \left[\frac{A + G + 0.5}{C + U + 0.5} \right]$; and the standard deviation of the normalized frequencies of all 16 dinucleotides. Nucleotide composition was calculated separately for each read and compared with a uniform zero-order null model in which $P(A) = P(G) = P(U) = P(C) = 0.25$. Overall composition was tested using a chi-square test, and 95% confidence intervals were obtained by resampling reads with replacement 1,000 times. Homopolymer tracts were evaluated separately for A, G, C, and U using the survival statistic $P(L_{\max} \geq k)$, where $L_{\max}$ is the longest base-specific tract in a read. The null distribution was defined by a sequence-specific zero-order permutation model in which nucleotide order was randomized independently within each read while preserving its exact length and mononucleotide counts. Exact conditional probabilities under this model were calculated for each read and averaged across reads; 95% null intervals were derived from the corresponding Bernoulli variances. Directional 6-mer motifs were counted in all overlapping sequence windows, excluding homopolymer-containing windows, and compared with a uniform zero-order null under which every 6-mer had probability 4^{-6} per window. Motif enrichment was expressed as $\log_2 \left[\frac{(O + 0.5)}{(E + 0.5)} \right]$ where O and E are the observed and expected counts. Statistical significance was assessed from read-level observed-minus-expected differences using two-sided normal tests, followed by *Benjamini-Hochberg* correction across all 4,096 6-mers. All motifs were ranked by their *Benjamini-Hochberg* adjusted P -values. Sequence logos were generated from eight-nucleotide contexts comprising the centered 6-mer and one flanking nucleotide on each side. All sequences and figures are reported in RNA notation, with U

used in place of T. Recurrent motifs were identified using the MEME Suite⁵².

DATA AVAILABILITY

All data necessary to evaluate the conclusions of this study are provided in the main text and Supplementary Materials. Additional data and materials are available upon request from the corresponding authors.

ACKNOWLEDGMENTS

We would like to thank Wenwen Zhang, Xiaoyan Xu, and Dr. Shan Feng of the Mass Spectrometry & Metabolomics Core Facility, Dr. Xiaohuo Shi and Dr. Xingyu Lu of the Instrumentation and Service Center for Molecular Sciences, Yifei Zhao of the Protein Characterization and Crystallography Facility, and Zhiquan Gao and Ping Deng of the Genomics Core Facility for their technical support. This work was supported by funds from the Pioneer and Lingyan R&D Program of Zhejiang (2024SSY50029 to Y.S.), the National Natural Science Foundation of China (32571475 to X.Z.) and the research funds from Westlake University to Y.S.

AUTHOR CONTRIBUTIONS

X.Z. and Y.S. conceived the project. R.L. executed the project. R.L., C.Y., M.L., and L.G. performed the experiments. All authors contributed to data analysis. R.L., X.Z., and Y.S. wrote the manuscript. Y.S. supervised the project.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- Steitz, T.A. The structural basis of the transition from initiation to elongation phases of transcription, as well as translocation and strand separation, by T7 RNA polymerase. *Curr. Opin. Struct. Biol.* **14**, 4–9 (2004).
- Ahlquist, P. RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science* **296**, 1270–1273 (2002).
- Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science* **292**, 1863–1876 (2001).
- te Velthuis, A.J.W. Common and unique features of viral RNA-dependent polymerases. *Cell. Mol. Life Sci.* **71**, 4403–4420 (2014).
- Warkocki, Z., Liudkovska, V., Gewartowska, O., Mroczek, S. & Dziembowski, A. Terminal nucleotidyl transferases (TENTs) in mammalian RNA metabolism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, 20180162 (2018).
- Liudkovska, V. & Dziembowski, A. Functions and mechanisms of RNA tailing by metazoan terminal nucleotidyltransferases. *Wiley Interdiscip. Rev. RNA* **12**, e1622 (2021).
- Schmidt, M.J. & Norbury, C.J. Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. *Wiley Interdiscip. Rev. RNA* **1**, 142–151 (2010).
- Tang, S. et al. Protein-primed homopolymer synthesis by an antiviral reverse transcriptase. *Nature* **643**, 1352–1362 (2025).
- Deng, P.J., Lee, H., Armijo, C., Wang, H.Q. & Gao, A. Protein-templated synthesis of dinucleotide repeat DNA by an antiphage reverse transcriptase. *Science* **392**, 1274–1281 (2026).
- Nguyen, D. et al. A polyadenylation-dependent 3' end maturation pathway is required for the synthesis of the human telomerase RNA. *Cell Rep.* **13**, 2244–2257 (2015).
- Roundtree, I.A., Evans, M.E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
- Kadaba, S. et al. Nuclear surveillance and degradation of hypomodified initiator tRNA^{Met} in *S. cerevisiae*. *Genes Dev.* **18**, 1227–1240 (2004).
- Sieburth, L.E. & Vincent, J.N. Beyond transcription factors: roles of mRNA decay in regulating gene expression in plants. *F1000Res.* **7**, 1940 (2018).
- LaCava, J. et al. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
- Cameron, T.A., Matz, L.M. & De Lay, N.R. Polynucleotide phosphorylase: not merely an RNase but a pivotal post-transcriptional regulator. *PLoS Genet.* **14**, e1007654 (2018).
- Grunberg-Manago, M., Ortiz, P.J. & Ochoa, S. Enzymic synthesis of polynucleotides. I. Polynucleotide phosphorylase of *Azotobacter vinelandii*. *Biochim. Biophys. Acta* **20**, 269–285 (1956).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Varadi, M. et al. AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).
- Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
- Sun, L.J., Wu, J.X., Du, F.H., Chen, X. & Chen, Z.J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
- Kranzusch, P.J. cGAS and CD-NTase enzymes: structure, mechanism, and evolution. *Curr. Opin. Struct. Biol.* **59**, 178–187 (2019).
- Hornung, V., Hartmann, R., Ablasser, A. & Hopfner, K.P. OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat. Rev. Immunol.* **14**, 521–528 (2014).
- Yang, H.Q. et al. Identification and expression profiling of *Oryza sativa* nucleotidyl transferase protein (NTP) genes under various stress conditions. *Gene* **628**, 93–102 (2017).
- Zhu, L.R. et al. Crystal structure of *Arabidopsis* terminal uridylyl transferase URT1. *Biochem. Biophys. Res. Commun.* **524**, 490–496 (2020).
- Chen, J.W., Li, X.Z., Dong, X.X. & Wang, X.Y. Functions and mechanisms of RNA tailing by nucleotidyl transferase proteins in plants. *Front. Plant Sci.* **15**, 1452347 (2024).
- Jenson, J.M. & Chen, Z.J. cGAS goes viral: a conserved immune defense system from bacteria to humans. *Mol. Cell* **84**, 120–130 (2024).
- Kranzusch, P.J. et al. Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* **158**, 1011–1021 (2014).
- Ming, Z.H. et al. Crystal structure of the novel di-nucleotide cyclase from *Vibrio cholerae* (DncV) responsible for synthesizing a hybrid cyclic GMP-AMP. *Cell Res.* **24**, 1270–1273 (2014).
- Kranzusch, P.J., Lee, A.S.Y., Berger, J.M. & Doudna, J.A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.* **3**, 1362–1368 (2013).
- Lyu, K.X., Chow, E.Y.C., Mou, X., Chan, T.F. & Kwok, C.K. RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Res.* **49**, 5426–5450 (2021).
- Luy, B. & Marino, J.P. ¹H-³¹P CPMG-correlated experiments for the assignment of nucleic acids. *J. Am. Chem. Soc.* **123**, 11306–11307 (2001).
- Gronenborn, A.M., Kimber, B.J., Clore, G.M. & McLaughlin, L.W. A nuclear magnetic resonance study of the ribotrinucleoside diphosphate UpUpC. *Nucleic Acids Res.* **11**, 5691–5699 (1983).
- Hu, Q. et al. Molecular mechanism underlying the di-uridylation activity of *Arabidopsis* TUTase URT1. *Nucleic Acids Res.* **50**, 10614–10625 (2022).
- Tu, B. et al. Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in *Arabidopsis*. *PLoS Genet.* **11**, e1005119 (2015).
- Scheer, H. et al. The TUTase URT1 connects decapping activators and prevents

- the accumulation of excessively deadenylated mRNAs to avoid siRNA biogenesis. *Nat. Commun.* **12**, 1298 (2021).
36. Donovan, J., Dufner, M. & Korennykh, A. Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc. Natl. Acad. Sci. USA* **110**, 1652–1657 (2013).
 37. Wiegand, D.J. et al. Template-independent enzymatic synthesis of RNA oligonucleotides. *Nat. Biotechnol.* **43**, 762–772 (2025).
 38. Robertson, M.P. & Joyce, G.F. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012).
 39. Gilbert, W. Origin of life: the RNA world. *Nature* **319**, 618 (1986).
 40. Lincoln, T.A. & Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232 (2009).
 41. Vaidya, N. et al. Spontaneous network formation among cooperative RNA replicators. *Nature* **491**, 72–77 (2012).
 42. Rodríguez, S.M. et al. Nonspecific base recognition mediated by water bridges and hydrophobic stacking in ribonuclease I from *Escherichia coli*. *Protein Sci.* **17**, 681–690 (2008).
 43. Giurgiu, C., Li, L., O'Flaherty, D.K., Tam, C.P. & Szostak, J.W. A mechanistic explanation for the regioselectivity of nonenzymatic RNA primer extension. *J. Am. Chem. Soc.* **139**, 16741–16747 (2017).
 44. Nagpal, N., Tai, A.K., Fong, Y.W. & Agarwal, S. Primer- and template-independent RNA polymerization by terminal nucleotidyltransferase TENT4B. *Nucleic Acids Res.* **54**, gkag864 (2026).
 45. Steitz, T.A., Smerdon, S.J., Jäger, J. & Joyce, C.M. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* **266**, 2022–2025 (1994).
 46. Jiang, S. et al. Generic Diagramming Platform (GDP): a comprehensive database of high-quality biomedical graphics. *Nucleic Acids Res.* **53**, D1670–D1676 (2025).
 47. Bolser, D.M., Staines, D.M., Perry, E. & Kersey, P.J. Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. *Methods Mol. Biol.* **1533**, 1–31 (2017).
 48. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
 49. Jones, D.T., Taylor, W.R. & Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275–282 (1992).
 50. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
 51. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
 52. Bailey, T.L., Johnson, J., Grant, C.E. & Noble, W.S. The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.15302/vita.2026.10.0078>.

Correspondence and requests for materials should be addressed to Xiechao Zhan, and Yigong Shi.

Reprints and permission information is available at <https://www.vita-journal.com/>.

© The Author(s) 2026. Published by Higher Education Press. This is an Open Access article distributed under the terms of the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).