

Cite this: *Chem. Sci.*, 2024, 15, 19022 All publication charges for this article have been paid for by the Royal Society of Chemistry

# Exonuclease-assisted enrichment and base resolution analysis of pseudouridine in single-stranded RNA<sup>†</sup>

Xin Fang,<sup>‡a</sup> Ziang Lu,<sup>‡a</sup> Yafen Wang,<sup>‡b</sup> Ruiqi Zhao,<sup>‡c</sup> Jing Mo,<sup>id a</sup> Wei Yang,<sup>a</sup> Mei Sun,<sup>a</sup> Xiang Zhou<sup>id ad</sup> and Xiaocheng Weng<sup>id \*ade</sup>

Pseudouridine ( $\Psi$ ) is one of the most abundant RNA modifications, playing crucial roles in various biological processes. Identifying  $\Psi$  sites is vital for understanding their functions. In this study, we proposed a novel method for identifying  $\Psi$  sites with an improved signal-to-noise ratio. This method, called RNA exonuclease-assisted identification of pseudouridine sites (RIPS), combines specific CMC-labeling of  $\Psi$  sites with an exonuclease-assisted digestion strategy for the detection of  $\Psi$  sites. Utilizing exonuclease XRN1 to digest RNA strands not labeled by CMC, RIPS significantly reduces the background signal from unlabeled strands and enhances the positive signal of  $\Psi$  sites labeled by CMC, which terminates exonuclease digestion. As a result, we can enrich  $\Psi$  sites and identify them at single-base resolution. Considering the unique functions of single-stranded RNA (ssRNA), we employed RIPS to distinguish  $\Psi$  sites in single-stranded and double-stranded regions of RNA. Our results indicated that CMC could specifically label  $\Psi$  sites in ssRNA under natural conditions, enabling RIPS to selectively identify  $\Psi$  sites in ssRNA, which may facilitate the study on the functions of  $\Psi$  sites.

Received 31st May 2024

Accepted 20th October 2024

DOI: 10.1039/d4sc03576c

rsc.li/chemical-science

## Introduction

RNA plays important roles in various biological processes. In addition to the four common bases, researchers have discovered many modified bases that are present in RNA. To date, more than 160 post-transcriptional modifications have been identified in different types of RNA.<sup>1–3</sup> Among these modifications, pseudouridine ( $\Psi$ ) is one of the most prevalent modifications, found in rRNA, tRNA and even in mRNA, where it has essential functions.<sup>4,5</sup> Many studies have demonstrated that  $\Psi$  can affect mRNA splicing, rRNA binding internal ribosome entry sites, translation fidelity, and RNA structure stability.<sup>6–10</sup> More recently, research has demonstrated that  $\Psi$  in tRNA can predict leukemic progression in myelodysplastic syndrome.<sup>11</sup>

$\Psi$  is an isomer of uridine (U) and is catalyzed by pseudouridine synthases (PUSs), which can be divided into two main groups: RNA-dependent box H/ACA ribonucleoproteins and RNA-independent PUSs.<sup>12–15</sup> The base-pairing pattern of  $\Psi$  in RNA is similar to U, but it can form an additional hydrogen bond, giving it vital roles in regulating the RNA structure.<sup>9,16,17</sup> It's well-known that the RNA secondary structure is an important regulatory element for post-transcriptional gene expression,<sup>18–20</sup> and several methods have been developed to map the RNA secondary structure, such as Keth-seq,<sup>21</sup> DMS-seq,<sup>22</sup> icSHAPE,<sup>23</sup> and EDC or CMC based probing technology.<sup>24</sup> It has been reported that the RNA structure could impact RNA modification,<sup>25</sup> while RNA modification can also regulate the RNA structure.<sup>26</sup>  $\Psi$  located in a duplex could stabilize RNA folding domains,<sup>27</sup> while  $\Psi$  sites in ssRNA, such as the loop of the hairpin, play special roles in the translation,<sup>28</sup> splicing,<sup>29</sup> and immune response.<sup>30</sup> Previous studies have revealed that three  $\Psi$  modifications ( $\Psi$ 1911,  $\Psi$ 1915, and  $\Psi$ 1917) in the 23S rRNA helix 69 could modulate conformation and structural dynamics of the helix in different ways, and the  $\Psi$ 1917C mutation may alter the RNA structure and become one of the important reasons for inhibiting the translation process in *E. coli*.<sup>31,32</sup> Therefore, distinguishing  $\Psi$  sites in ssRNA and double-stranded RNA (dsRNA) will facilitate research on the function of  $\Psi$ .

Various research tools and techniques have been developed to map  $\Psi$  sites, which included  $\Psi$ -seq, PSI-seq, Pseudo-seq, HydraPsiSeq, CeU-seq, BID-seq, PRAISE and the nanopore-

<sup>a</sup>College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers-Ministry of Education, Wuhan University, Wuhan, Hubei, 430072, P. R. China. E-mail: xcweng@whu.edu.cn

<sup>b</sup>School of Public Health, Wuhan University, Wuhan, Hubei 430071, P. R. China

<sup>c</sup>Department of Neurosurgery, Zhongnan Hospital of Wuhan University, Wuhan, Hubei 430071, P. R. China

<sup>d</sup>TaiKang Center for Life and Medical Sciences, Wuhan University, Wuhan, Hubei 430071, P. R. China

<sup>e</sup>Department of Otorhinolaryngology-Head and Neck Surgery, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, P. R. China

<sup>†</sup> Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4sc03576c>

<sup>‡</sup> These authors contributed equally to this work.



based method.<sup>4,33–40</sup> The classical method for  $\Psi$  profiling is a chemically assisted method using *N*-cyclohexyl-*N'*-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMC) to map transcriptome-wide  $\Psi$  sites.<sup>41</sup> CMC can react with guanosine (G), U and  $\Psi$  to produce the corresponding adducts, G-CMC, U-CMC and  $\Psi$ -CMC. However, G-CMC and U-CMC adducts undergo CMC removal under alkaline treatment, while  $\Psi$ -CMC does not because it is more stable. The  $\Psi$ -CMC blocks the reverse transcriptase and prevents reverse transcription at one nucleotide 3' to the  $\Psi$ -CMC sites, enabling single-base resolution mapping of  $\Psi$ . Although this method is commonly used, it has certain limitations. It doesn't pre-enrich  $\Psi$ -containing RNA, so lower abundance  $\Psi$  sites may be missed due to the high background caused by a large number of RNA fragments without CMC labeling.  $N_3$ -CMC, a derivative of CMC, was developed to identify  $\Psi$  in RNA with higher accuracy and sensitivity, using biotin to enrich RNA containing  $\Psi$  (CeU-seq).<sup>36</sup> However, the synthesis and storage of  $N_3$ -CMC are challenging, limiting its widespread application. Recently, bisulfite-induced deletion sequencing (BID-seq) and quantitative  $\Psi$  assessment *via* bisulfite/sulfite treatment (PRAISE) were reported, which use a bisulfite-mediated reaction to quantitatively convert  $\Psi$  to a sulfite adduct of  $\Psi$  without cytidine deamination, resulting in a unique deletion signature at the  $\Psi$  sites during reverse transcription.<sup>37,38,42,43</sup> However, sufficient read coverage is required to identify low-abundance  $\Psi$  sites because of the absence of enrichment methods.

Herein, an RNA exonuclease-assisted identification of pseudouridine sites (RIPS) was developed to address the issue of the high background caused by RNA fragments not labeled by CMC. This method involved using CMC to specifically label  $\Psi$ .<sup>44,45</sup> XRN1 is a 5' exonuclease that could degrade RNA starting from its 5' end. Due to CMC steric hindrance, XRN1 digestion is terminated at one nucleotide 5' to the  $\Psi$ -CMC sites in CMC-labeled strands. As a result, RIPS can enrich  $\Psi$ -containing RNA strands labeled by CMC and identify  $\Psi$  sites with a higher signal-to-noise ratio. It is worth noting that  $\Psi$  within different RNA secondary structures may have distinct biological roles, but has not received sufficient attention. The RIPS method can distinguish  $\Psi$  in ssRNA or dsRNA because CMC can only label  $\Psi$  in ssRNA, not in dsRNA, which may provide new insights into the understanding of functions of  $\Psi$  modifications.

## Results and discussion

### Development of the RIPS strategy for $\Psi$ identification

Based on previous studies, we have developed a novel method for  $\Psi$  detection called RIPS. This method employed exonuclease to digest RNA lacking  $\Psi$  modifications that cannot be labeled by CMC. Meanwhile, CMC-labeling at the  $\Psi$  sites in  $\Psi$ -containing RNA hindered exonuclease digestion and made the digestion terminate before  $\Psi$  sites due to the steric hindrance of CMC (Fig. 1). Therefore, a large number of RNA strands that weren't labeled by CMC could be excluded from library construction and sequencing, achieving the desired enrichment effect (Fig. 2A). Furthermore, as CMC can only label  $\Psi$  sites in ssRNA but not dsRNA, RIPS can selectively identify  $\Psi$  in ssRNA with

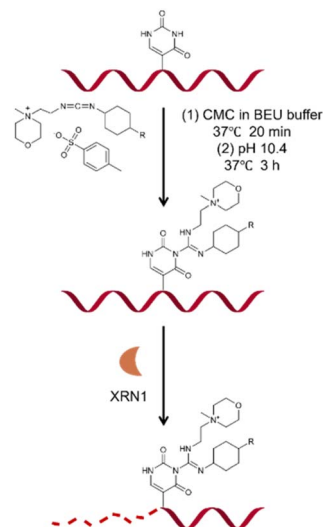


Fig. 1 Schematic illustrating the principle of CMC labeling and XRN1 exonuclease digestion.

a higher signal-to-noise ratio by digesting unlabeled dsRNA (Fig. 2B).

### The feasibility verification of the RIPS strategy

Initially, two 19-nucleotide (nt) RNAs containing a  $\Psi$  site (19-RNA- $\Psi$ ) or a U site (19-RNA-U) were incubated with CMC, respectively. Using denaturing polyacrylamide gel electrophoresis, we observed that 19-RNA- $\Psi$  migrated slower after reacting with CMC and undergoing alkaline treatment, while 19-RNA-U did not (Fig. 3A). High-resolution mass spectrometry confirmed that CMC could indeed label the  $\Psi$ -containing RNA

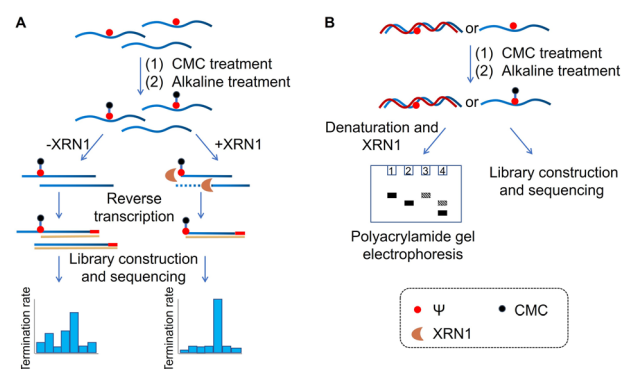
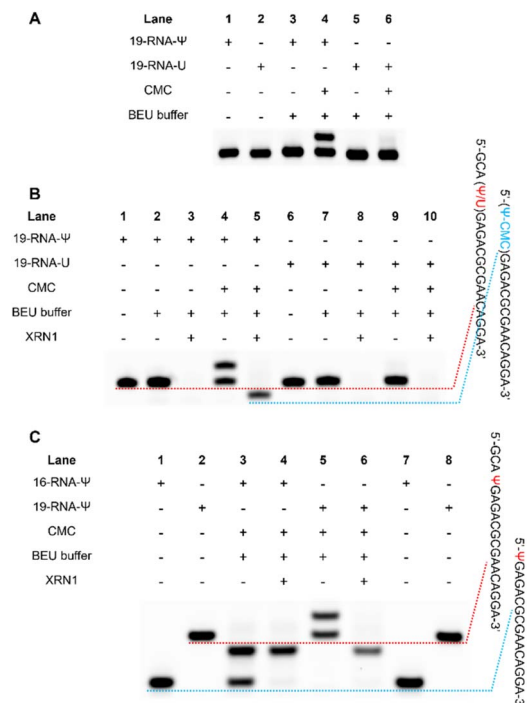


Fig. 2 Schematic of the RIPS strategy. (A) CMC was used to specifically label  $\Psi$  sites contained in RNA. And unlabeled RNA was digested by XRN1 exonuclease from the 5' end to the 3' end, while the  $\Psi$ -CMC adduct caused the digestion to terminate before  $\Psi$  sites of labeled RNA, thereby enabling  $\Psi$  detection with a higher signal-to-noise ratio. (B) RIPS strategy for the distinction of  $\Psi$  sites in single-stranded and double-stranded regions of RNA. CMC could selectively label  $\Psi$  sites in ssRNA under natural conditions, while dsRNA failed to be labeled. Selective labeling of  $\Psi$  was confirmed by gel electrophoresis analysis. And the  $\Psi$  sites in ssRNA were specifically identified by high-throughput sequencing.





**Fig. 3** Preliminary experimental validation of RIPS. (A) RNAs containing Ψ or U were treated with CMC, and separated by gel electrophoresis. (B) Preliminary validation of the feasibility of the RIPS strategy. The RNA containing Ψ or U was reacted with CMC. XRN1 was used to digest unlabeled RNA. (C) Gel electrophoresis analysis of the products of 16-RNA-Ψ and 19-RNA-Ψ reacted with CMC and treated with XRN1.

(Fig. S1†). We also optimized the conditions for the reaction of the RNA with CMC (Fig. S2†).

Subsequently we treated the above 19 nt RNA with XRN1 exonuclease after CMC labeling. Denaturing polyacrylamide gel electrophoresis revealed the emergence of a shorter RNA strand in the 19-RNA-Ψ sample treated with CMC and XRN1, which was expected to result from the termination of digestion before the Ψ sites. In contrast, the 19-RNA-U sample treated with CMC and XRN1 displayed a reduced amount of the original RNA strand (Fig. 3B).

To confirm whether exonuclease cleavage terminated before the Ψ site, a 16 nt RNA strand (16-RNA-Ψ) was synthesized, with the three bases in front of the Ψ site of the 19-RNA-Ψ removed. CMC labeling and exonuclease cleavage were conducted on both 16-RNA-Ψ and 19-RNA-Ψ. Gel electrophoresis analysis showed that the products in Lanes 3, 4, and 6 were identical, though Lane 3 still exhibited some residual RNA material (Fig. 3C). To further verify the products, mass spectrometry was conducted. The results revealed that the molecular weights of both 16-RNA-Ψ and 19-RNA-Ψ treated with CMC labeling and XRN1 digestion were 6107.9, matching the molecular weight of 16-RNA-Ψ (5855.7) plus the molecular weight of one CMC molecule (252.2) (Fig. S3†). These findings provided compelling evidence that XRN1 digestion indeed terminated at the Ψ site of 19-RNA-Ψ following CMC labeling and exonuclease treatment.

These results demonstrated that CMC effectively impeded exonuclease digestion from the 5' end by its steric hindrance,

causing the digestion to terminate before the CMC-labeled Ψ sites and producing a shorter band, while the U-containing RNA didn't.

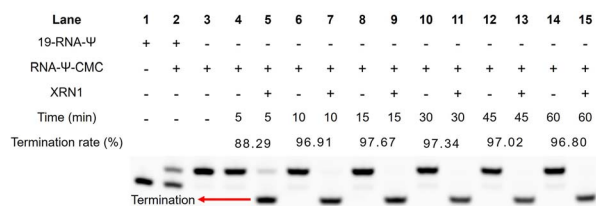
### Optimization of experimental conditions for the RIPS strategy

We further optimized the digestion conditions to ensure that CMC-unlabeled RNA strands were almost completely digested, while the CMC-labeled RNA strands could hinder digestion before the Ψ sites as much as possible. First, we optimized the concentration of exonuclease (0, 0.125, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 1, and 2 U). Gel electrophoresis was used to assess the digestion efficiency of XRN1, with the best efficiency observed at 0.25 U exonuclease (Fig. S4†). Next, we optimized the digestion time of exonuclease (0, 5, 10, 15, 30, 40, 50, and 60 min). The optimal digestion conditions for 19-RNA-Ψ were determined to be 100 ng of RNA, NEB buffer 3, 0.5 U enzyme concentration, and 5 minutes of digestion time (Fig. S4†). These experimental results demonstrated that the combination of CMC-specific labeling technology and the exonuclease digestion method provided an efficient strategy to enrich Ψ-containing RNA which was labeled by using CMC.

Subsequently, we assessed the termination efficiency of XRN1 digestion for CMC-labeled RNA. CMC-labeled RNA was separated from other unlabeled RNA by HPLC (Fig. S5†). The recovered CMC-labeled RNA strand (RNA-Ψ-CMC) was digested by XRN1 with a time gradient (0, 5, 10, 15, 30, 45, and 60 min). Compared to the control group, which did not undergo digestion, the CMC-labeled RNA strand demonstrated a 95% probability of impeding exonuclease cleavage after 10 minutes (Fig. 4), indicating that the enzymatic digestion strategy enriched CMC-labeled RNA and enabled high-sensitivity detection of Ψ.

### Ψ identification by the RIPS strategy using high-throughput sequencing

In order to verify that the exonuclease digestion does stop before the Ψ sites, and that the method could indeed be used for the identification of the Ψ sites with a higher signal-to-noise ratio, we used a longer 60 nt RNA containing a Ψ site (60-RNA-Ψ) to assess CMC labeling efficiency and exonuclease digestion efficiency through library sequencing analysis. The samples treated with CMC were subjected to an exonuclease digestion



**Fig. 4** Termination rate of XRN1 digestion influenced by the Ψ-CMC adduct. The CMC-labeled RNA was separated by HPLC and subjected to XRN1 digestion. RNA was separated by gel electrophoresis and the termination rate was calculated by comparing the signals of shortened RNA and untreated RNA.



time gradient (15, 30, and 60 min). The gel electrophoresis analysis showed that a shortened chain was produced by XRN1 digestion as previously described (Fig. 5A). The sequencing results demonstrated that obvious reverse transcription termination occurred at one nucleoside 3' to the  $\Psi$  sites in the samples treated with CMC, with a termination rate of less than 30%. Meanwhile, the samples treated with both CMC and XRN1 were found to indeed undergo reverse transcription termination before the  $\Psi$  sites, with a higher termination efficiency compared to CMC-only treated samples. The reverse transcription termination efficiency of CMC and XRN1 treated samples exceeded 55% (Fig. 5B and Dataset S1†). This evidence supported the notion that the exonuclease could digest CMC-unlabeled RNA within 15 min of digestion time, and that CMC-labeling in RNA could terminate digestion of XRN1 in front of the  $\Psi$  sites, which could thereby improve the reverse transcription termination rate of the  $\Psi$  site and enhance the signal-to-noise ratio. The results indicated the feasibility and reliability of the RIPS method.

### RIPS strategy identifies $\Psi$ sites in total RNA

Subsequently, we sought to detect several  $\Psi$  sites located in rRNA and even mRNA using RIPS (Table S3†). We used real-time quantitative PCR (qPCR) to quantify the termination rate, designing two pairs of primers, with one pair of primers (FP1 and RP1) located on either side of the  $\Psi$  site and the other pair

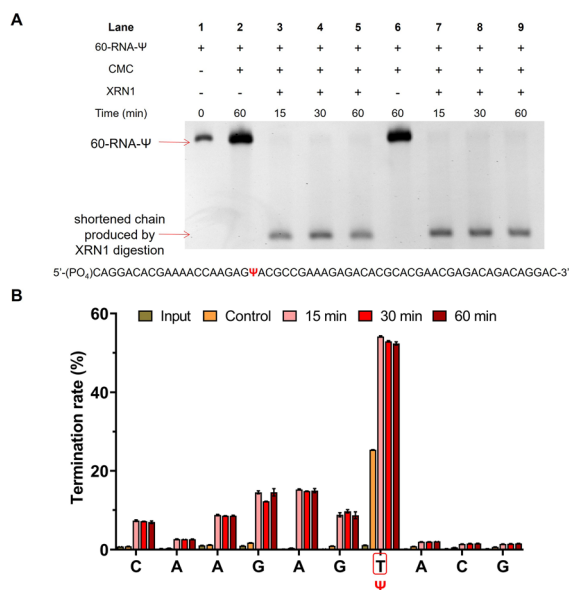


Fig. 5 RIPS strategy identified  $\Psi$  sites with a higher signal-to-noise ratio. (A) Gel electrophoresis analysis exhibited the efficiency of XRN1 digestion toward 60-RNA- $\Psi$ . (B) Validation of the RIPS strategy for  $\Psi$  site identification of 60-RNA- $\Psi$  by high-throughput sequencing. RNA was treated with XRN1 for different times (0, 15, 30, and 60 min). Input: 60-RNA- $\Psi$  input. Control: 60-RNA- $\Psi$  treated with CMC only. 15 min: 60-RNA- $\Psi$  treated with CMC and XRN1 for 15 min. 30 min: 60-RNA- $\Psi$  treated with CMC and XRN1 for 30 min. 60 min: 60-RNA- $\Psi$  treated with CMC and XRN1 for 60 min. The position of  $\Psi$  was marked with a red frame ( $n = 2$  biologically independent experiments).

of primers (FP2 and RP2) located downstream of the  $\Psi$  site (Table S2†). As shown in Fig. 6A, only cDNAs without reverse transcription termination were successfully amplified by FP1 and RP1, while cDNAs both with or without reverse transcription termination could be successfully amplified by FP2 and RP2. We used Input as the control and divided the relative content of FP1 and RP1 amplification by the relative content of FP2 and RP2 amplification ( $2^{-(\Delta Ct_1 - \Delta Ct_2)}$ ). The termination rate was calculated using  $1 - 2^{-(\Delta Ct_1 - \Delta Ct_2)}$ . The previous conditions for  $\Psi$  detection on RNA from HEK293T cells resulted in incomplete digestion, likely due to the more complex structure of the samples and the broader variety of RNA species. To address this, we re-optimized the experimental conditions for these complex RNA samples. The optimal XRN1 digestion conditions were determined to be 100 ng of RNA, NEB buffer 1, 1 U enzyme concentration, and 5 minutes of digestion time (Fig. S6†). Initially, we digested the input samples without CMC labeling to eliminate the potential influence of the RNA structure and other variables on the exonuclease digestion process (Fig. S7†). The results showed that most of the unlabeled RNA was effectively digested. Subsequently, we successfully identified several known  $\Psi$  sites in rRNA and mRNA with a higher signal-to-noise ratio (Fig. 6B).<sup>36,37</sup> Furthermore, our analysis of the termination rate at 28S-U3074 revealed negligible differences with or without XRN1 treatment, supporting the efficacy of RIPS in detecting  $\Psi$  sites in genuine samples while enhancing the signal clarity.

### RIPS strategy enables selective identification of $\Psi$ sites in single-stranded RNA

In addition, we also conducted a preliminary study of  $\Psi$  sites in different structural regions of RNA. RNA has a complex

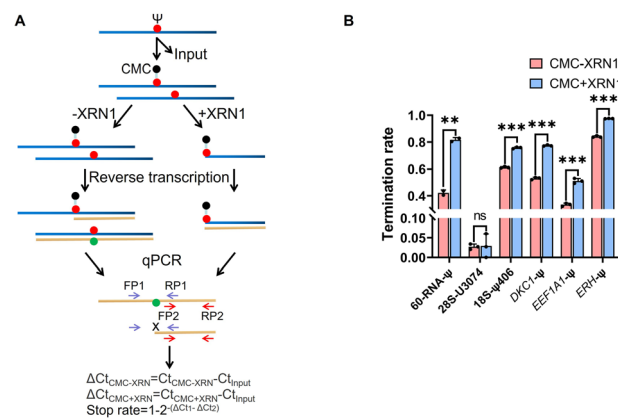


Fig. 6 RIPS strategy identified  $\Psi$  sites in total RNA with a higher signal-to-noise ratio. (A) Schematic for the calculation of the termination rate of  $\Psi$  sites by qPCR. Two pairs of primers were designed (FP1 and RP1, and FP2 and RP2). Using Input as the control, the termination rate was calculated using  $1 - 2^{-(\Delta Ct_1 - \Delta Ct_2)}$ .  $\Delta Ct_1$  was defined as the relative content of the treated samples when amplified with FP1 and RP1 primers.  $\Delta Ct_2$  was defined as the relative content of the treated samples when amplified with FP2 and RP2. (B) Termination rate of 60-RNA- $\Psi$  and several known sites in rRNA and even mRNA using RIPS or only CMC labeling. The specific sites are listed in Table S3.† Error bars indicate mean  $\pm$  s.d. for 3 technical replicates.



secondary or tertiary structure, and  $\Psi$  sites located in various RNA structural regions may possess different biological functions. In ssRNA, the hydrogen atoms at the N3 position of  $\Psi$  bases don't participate in base-pairing, allowing CMC to specifically label  $\Psi$ . However, when the  $\Psi$  sites are located in dsRNA, the hydrogen atoms at the N3 position would be involved in hydrogen bond formation, and thus prevent CMC from labelling the  $\Psi$  sites. To examine the labeling efficiency of CMC to ssRNA and dsRNA, we designed two partly complementary strands to block the  $\Psi$  sites in 19-RNA- $\Psi$ . The 19-RNA- $\Psi$  was annealed with the two complementary strands to form dsRNA, in which the  $\Psi$  sites were blocked in the double-stranded regions (Fig. S8<sup>†</sup>). The single-stranded 19-RNA- $\Psi$  and dsRNA were both treated with CMC and analyzed by denaturing polyacrylamide gel electrophoresis. During the subsequent denaturing polyacrylamide gel electrophoresis, the RNA secondary structure was opened, and it could be seen that only the 19-RNA- $\Psi$  that was treated with CMC showed two electrophoretic bands, while the blocking-RNA that was treated with CMC displayed only one band (Fig. 7A), indicating that only the  $\Psi$  in the single-stranded regions could be labeled by CMC.

We then selected the previously used 60-RNA- $\Psi$  and designed a series of partly complementary RNA to perform XRN1 digestion, library construction and sequencing. After annealing, the  $\Psi$  sites were blocked in the double-stranded regions. We finally selected the RNA-blocking 17, as it had best blocking efficiency (Fig. S9<sup>†</sup>). The single-stranded 60-RNA- $\Psi$  and double-stranded RNA were treated with CMC. Then both RNAs were denatured and treated with XRN1. We found that the 60-RNA- $\Psi$  in dsRNA was completely digested by XRN1, because the  $\Psi$ -containing RNA failed to be labeled by CMC. However, the single-stranded 60-RNA- $\Psi$  was only partly digested by XRN1

and displayed a shortened chain, suggesting that the digestion was resisted by CMC labeling (Fig. 7B). This result further demonstrated that only  $\Psi$  in the single-stranded regions could be specifically labeled by CMC, while  $\Psi$  in the double-stranded regions could not. We further performed additional library construction and sequencing analysis of  $\Psi$  in the ssRNA or dsRNA under CMC labeling. The results showed that reverse transcription termination was apparent at  $\Psi$  sites in single-stranded regions but not in double-stranded regions (Fig. 7C and Dataset S2<sup>†</sup>), proving that RIPS could be used to distinguish  $\Psi$  sites located at different RNA secondary structures.

Finally, we designed another ssRNA (32-RNA- $\Psi$ ) containing a  $\Psi$  and mixed it with the double-stranded 60-RNA- $\Psi$ , to examine whether our method could selectively identify  $\Psi$  sites in ssRNA from a mixture of ssRNA and dsRNA. The mixture was incubated with CMC and treated with XRN1. Denaturing polyacrylamide gel electrophoresis revealed four bands following CMC treatment, suggesting that single-stranded 32-RNA- $\Psi$  was specifically labeled by CMC (Fig. 8A). After introducing XRN1, a shortened band appeared and the top band disappeared, indicating that 32-RNA- $\Psi$  was successfully labeled by CMC and the  $\Psi$ -CMC adduct successfully resisted XRN1 digestion, while the double-stranded 60-RNA- $\Psi$  failed to be labeled and was almost completely digested by XRN1. RNA-blocking 17 remained undigested by XRN1 due to the absence of phosphate modification at its 5' end which prevented XRN1 from digesting this RNA strand. Subsequently, we performed library construction and sequencing analysis of these sets of sample Lanes 4, 5,

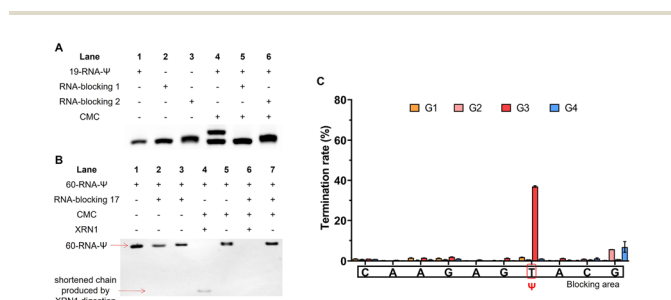


Fig. 7 Distinguishing of  $\Psi$  sites in single-stranded and double-stranded regions of RNA by the RIPS strategy. (A) RNA-blocking 1 or 2 was annealed with the 19 nt RNA to form dsRNA, allowing the  $\Psi$  sites to be blocked in the double-stranded structure, and the blocked or unblocked oligonucleotide chain was incubated with CMC under natural conditions. (B) RNA-blocking 17 was annealed with the 60 nt RNA and treated with CMC, the same as (A). After denaturation treatment, RNA was further treated with XRN1. (C) High-throughput sequencing validated that CMC only labeled  $\Psi$  sites in single-stranded regions of RNA. G1: single-stranded 60-RNA- $\Psi$  input. G2: double-stranded 60-RNA- $\Psi$  input. G3: single-stranded 60-RNA- $\Psi$  treated with CMC. G4: double-stranded 60-RNA- $\Psi$  treated with CMC. The position of  $\Psi$  was marked with a red frame and the partial blocking area was marked with a black square ( $n = 2$  biologically independent experiments).

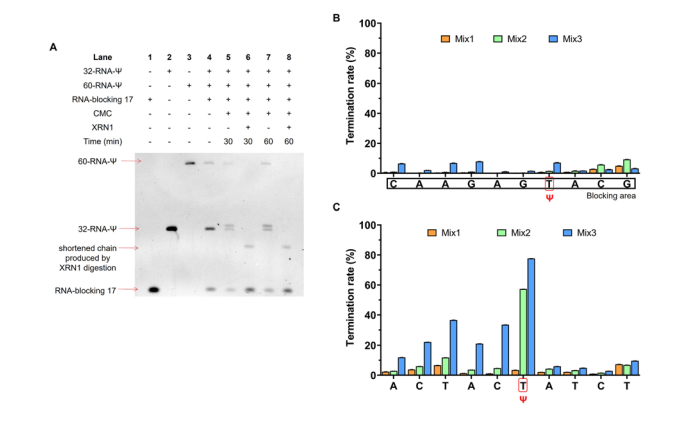


Fig. 8 RIPS strategy selectively identified  $\Psi$  sites in ssRNA from a mixture of ssRNA and dsRNA. (A) Single-stranded 32-RNA- $\Psi$  and double-stranded 60-RNA- $\Psi$  were mixed and treated with CMC under natural conditions. After that, the mixture was further treated with XRN1 under denaturing conditions. RNAs were analyzed by gel electrophoresis. (B) Reverse transcription termination rate of bases near the modified base of 60-RNA- $\Psi$  in the mixture was analyzed by sequencing. (C) Reverse transcription termination rate of bases near the modified base of 32-RNA- $\Psi$  in the mixture was analyzed by sequencing. Mix1: single-stranded 32-RNA- $\Psi$  and double-stranded 60-RNA- $\Psi$  mixture input. Mix2: single-stranded 32-RNA- $\Psi$  and double-stranded 60-RNA- $\Psi$  mixture treated with CMC only. Mix3: single-stranded 32-RNA- $\Psi$  and double-stranded 60-RNA- $\Psi$  mixture treated with CMC and XRN1. The position of  $\Psi$  was marked with a red frame and the partial blocking area was marked with a black square ( $n = 2$  biologically independent experiments).



6, 7 and 8. By analyzing the sequencing results of 60-RNA- $\Psi$ , we found a low reverse transcription termination rate of modification sites under CMC treatment, suggesting that it failed to be labeled by CMC (Fig. 8B). In contrast, the sequencing results of 32-RNA- $\Psi$  exhibited a high reverse transcription termination rate of  $\Psi$  in treated samples, and XRN1 treatment indeed improved the termination rate of  $\Psi$  (Fig. 8C and Dataset S3†). These results further confirmed that CMC can indeed be used to distinguish  $\Psi$  in single- and double-stranded structural regions. Our exonuclease-assisted method significantly improved the accuracy and sensitivity of  $\Psi$  identification. We believe that our preliminary results will contribute to the subsequent studies about identifying  $\Psi$  sites in different structural regions of RNA.

## Conclusions

In conclusion, we established a new  $\Psi$  identification method, a chemical and exonuclease-assisted method (RIPS). This method employs CMC to specifically label  $\Psi$  sites, followed by exonuclease digestion. This strategy efficiently enriches  $\Psi$ , enhancing the signal-to-noise ratio and detection sensitivity. Our results demonstrated that XRN1 digestion could efficiently improve the reverse transcription termination rate of CMC-labeled RNA, reducing the background signal and enhance the positive signal, which indicated that RIPS is an efficient and meaningful technique for the identification of  $\Psi$ . In addition, RIPS can distinguish  $\Psi$  located in ssRNA and dsRNA. It is believed that RIPS is also a way of thought for future studies on  $\Psi$  mapping across different structural regions of RNA. Moreover, RIPS can provide an effective idea for the accurate mapping of  $\Psi$ , which can contribute to our understanding of  $\Psi$  functions in biology and medicine.

## Data availability

Sequencing data have been deposited into the Gene Expression Omnibus (GEO). The accession number is GSE229317. To review GEO accession GSE229317, go to <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE229317>. Enter token enejsygrfcpdef into the box.

## Author contributions

X. F., and X. W. conceived the idea for this study. X. F. designed the experiments. X. F., R. Z., and M. S. verified and developed the method, and performed all the experiments. Z. L. designed and performed the bioinformatics analysis. Y. W., J. M., and W. Y. contributed to the design of the experiments. X. F., and X. W. wrote the manuscript with help from all authors.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported financially by the Ministry of Science and Technology (2023YFC3402200), the National Natural Science Foundation of China (92253202 and 22177087 to X. W.; 22377096 to Y. W.; 92153303 and 21721005 to X. Z.), and the Fundamental Research Funds for the Central Universities (2042023kfyq05 and 2042023kf0118).

## Notes and references

- 1 P. Boccaletto, M. A. Machnicka, E. Purta, P. Piatkowski, B. Baginski, T. K. Wirecki, V. de Crecy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm and J. M. Bujnicki, *Nucleic Acids Res.*, 2018, **46**, D303–D307.
- 2 X. Y. Li, S. Q. Ma and C. Q. Yi, *Curr. Opin. Chem. Biol.*, 2016, **33**, 108–116.
- 3 Y. F. Wang, X. Zhang, H. Liu and X. Zhou, *Chem. Soc. Rev.*, 2021, **50**, 13481–13497.
- 4 S. Schwartz, D. A. Bernstein, M. R. Mumbach, M. Jovanovic, R. H. Herbst, B. X. Leon-Ricardo, J. M. Engreitz, M. Guttman, R. Satija, E. S. Lander, G. Fink and A. Regev, *Cell*, 2014, **159**, 148–162.
- 5 J. H. Ge and Y. T. Yu, *Trends Biochem. Sci.*, 2013, **38**, 210–218.
- 6 N. M. Martinez, A. Su, M. C. Burns, J. K. Nussbacher, C. Schaening, S. Sathe, G. W. Yeo and W. V. Gilbert, *Mol. Cell*, 2022, **82**, 645–659.
- 7 K. Jack, C. Bellodi, D. M. Landry, R. O. Niederer, A. Meskauskas, S. Musalgaonkar, N. Kopmar, O. Krasnykh, A. M. Dean, S. R. Thompson, D. Ruggero and J. D. Dinman, *Mol. Cell*, 2011, **44**, 660–666.
- 8 A. T. Yu, J. H. Ge and Y. T. Yu, *Protein Cell*, 2011, **2**, 712–725.
- 9 E. Kierzek, M. Malgowska, J. Lisowiec, D. H. Turner, Z. Gdaniec and R. Kierzek, *Nucleic Acids Res.*, 2014, **42**, 3492–3501.
- 10 A. Baudin-Baillieu, C. Fabret, X. H. Liang, D. Piekna-Przybylska, M. J. Fournier and J. P. Rousset, *Nucleic Acids Res.*, 2009, **37**, 7665–7677.
- 11 N. Guzzi, S. Muthukumar, M. Ciesla, G. Todisco, P. C. T. Ngoc, M. Madej, R. Munita, S. Fazio, S. Ekstrom, T. Mortera-Blanco, M. Jansson, Y. Nannya, M. Cazzola, S. Ogawa, L. Malcovati, E. Hellstrom-Lindberg, M. Dimitriou and C. Bellodi, *Nat. Cell Biol.*, 2022, **24**, 299–306.
- 12 J. Ofengand, *FEBS Lett.*, 2002, **514**, 17–25.
- 13 F. Spenkuch, Y. Motorin and M. Helm, *RNA Biol.*, 2014, **11**, 1540–1554.
- 14 T. Kiss, E. Fayet-Lebaron and B. E. Jady, *Mol. Cell*, 2010, **37**, 597–606.
- 15 A. C. Rintala-Dempsey and U. Kothe, *RNA Biol.*, 2017, **14**, 1185–1196.
- 16 W. A. Decatur and M. J. Fournier, *Trends Biochem. Sci.*, 2002, **27**, 344–351.
- 17 J. G. Arnez and T. A. Steitz, *Biochemistry*, 1994, **33**, 7560–7567.
- 18 Y. L. Ding, Y. Tang, C. K. Kwok, Y. Zhang, P. C. Bevilacqua and S. M. Assmann, *Nature*, 2014, **505**, 696–700.



- 19 Y. Wan, K. Qu, Z. Q. Ouyang, M. Kertesz, J. Li, R. Tibshirani, D. L. Makino, R. C. Nutter, E. Segal and H. Y. Chang, *Mol. Cell*, 2012, **48**, 169–181.
- 20 R. C. Spitale, P. Crisalli, R. A. Flynn, E. A. Torre, E. T. Kool and H. Y. Chang, *Nat. Chem. Biol.*, 2013, **9**, 18–20.
- 21 X. C. Weng, J. Gong, Y. Chen, T. Wu, F. Wang, S. X. Yang, Y. S. Yuan, G. Z. Luo, K. Chen, L. L. Hu, H. H. Ma, P. L. Wang, Q. F. C. Zhang, X. Zhou and C. He, *Nat. Chem. Biol.*, 2020, **16**, 489–492.
- 22 S. Rouskin, M. Zubradt, S. Washietl, M. Kellis and J. S. Weissman, *Nature*, 2014, **505**, 701–705.
- 23 R. C. Spitale, R. A. Flynn, Q. C. Zhang, P. Crisalli, B. Lee, J. W. Jung, H. Y. Kuchelmeister, P. J. Batista, E. A. Torre, E. T. Kool and H. Y. Chang, *Nature*, 2015, **519**, 486–490.
- 24 P. Y. Wang, A. N. Sexton, W. J. Culligan and M. D. Simon, *RNA*, 2019, **25**, 135–146.
- 25 T. M. Carlile, N. M. Martinez, C. Schaening, A. Su, T. A. Bell, B. Zinshteyn and W. V. Gilbert, *Nat. Chem. Biol.*, 2019, **15**, 966–974.
- 26 C. J. T. Lewis, T. Pan and A. Kalsotra, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 202–210.
- 27 J. Karijolich, C. Q. Yi and Y. T. Yu, *Nat. Rev. Mol. Cell Biol.*, 2015, **16**, 581–585.
- 28 P. S. Ray, J. Jia, P. Yao, M. Majumder, M. Hatzoglou and P. L. Fox, *Nature*, 2009, **457**, 915–919.
- 29 Z. S. Liu, Q. Liu, X. F. Yang, Y. Y. Zhang, M. Norris, X. X. Chen, J. Cheema, H. K. Zhang and Y. L. Ding, *Genome Biol.*, 2021, **22**, 11.
- 30 F. Heil, H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner and S. Bauer, *Science*, 2004, **303**, 1526–1529.
- 31 J. Jiang, R. Aduri, C. S. Chow and J. SantaLucia, *Nucleic Acids Res.*, 2014, **42**, 3971–3981.
- 32 A. Liiv, D. Karitkina, U. Maivali and J. Remme, *BMC Mol. Biol.*, 2005, **6**, 18.
- 33 A. F. Lovejoy, D. P. Riordan and P. O. Brown, *PLoS One*, 2014, **9**, e110799.
- 34 T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli and W. V. Gilbert, *Nature*, 2014, **515**, 143–146.
- 35 V. Marchand, F. Pichot, P. Neybecker, L. Ayadi, V. Bourguignon-Igel, L. Wacheul, D. L. J. Lafontaine, A. Pinzano, M. Helm and Y. Motorin, *Nucleic Acids Res.*, 2020, **48**, e1110.
- 36 X. Y. Li, P. Zhu, S. Q. Ma, J. H. Song, J. Y. Bai, F. F. Sun and C. Q. Yi, *Nat. Chem. Biol.*, 2015, **11**, 592–597.
- 37 Q. Dai, L. S. Zhang, H. L. Sun, K. Pajdzik, L. Yang, C. Ye, C. W. Ju, S. Liu, Y. R. Wang, Z. Zheng, L. D. Zhang, B. T. Harada, X. Y. Dou, I. Irklyenko, X. R. Feng, W. Zhang, T. Pan and C. He, *Nat. Biotechnol.*, 2023, **41**, 344–354.
- 38 M. Zhang, Z. Jiang, Y. Ma, W. Liu, Y. Zhuang, B. Lu, K. Li, J. Peng and C. Yi, *Nat. Chem. Biol.*, 2023, **19**, 1185–1195.
- 39 S. H. Huang, W. Zhang, C. D. Katanski, D. Dersh, Q. Dai, K. Lolans, J. Yewdell, A. M. Eren and T. Pan, *Genome Biol.*, 2021, **22**, 330.
- 40 S. Tavakoli, M. Nabizadeh, A. Makhamreh, H. Gamper, C. A. McCormick, N. K. Rezapour, Y. M. Hou, M. Wanunu and S. H. Rouhanifard, *Nat. Commun.*, 2023, **14**, 334.
- 41 A. V. Bakin and J. Ofengand, *Methods Mol. Biol.*, 1998, **77**, 297–309.
- 42 V. Khoddami, A. Yerra, T. L. Mosbrugger, A. M. Fleming, C. J. Burrows and B. R. Cairns, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 6784–6789.
- 43 A. M. Fleming, A. Alenko, J. P. Kitt, A. M. Orendt, P. F. Flynn, J. M. Harris and C. J. Burrows, *J. Am. Chem. Soc.*, 2019, **141**, 16450–16460.
- 44 C. W. Q. Koh, Y. T. Goh and W. S. S. Goh, *Nat. Commun.*, 2019, **10**, 5636.
- 45 K. S. Mironov, M. Shumskaya and D. A. Los, *Biochimie*, 2020, **177**, 63–67.

