Gain-of-function reporters for analysis of mRNA 3'-end formation: Design and optimization

Natoya Peart^{1,2,3} and Eric J Wagner^{1,2,3}

¹Department of Biochemistry and Molecular Biology, The University of Texas Medical School at Houston, Houston, TX, ²Graduate Program in Biochemistry and Molecular Biology, The University of Texas Graduate School of Biomedical Sciences, Houston, TX, and ³Department of Biochemistry and Molecular Biology, The University of Texas Medical Branch, Galveston, TX

BioTechniques 60:137–140 (March 2016) doi 10.2144/000114390

Keywords: 3'-end processing; reporter assay; gain-of-function

The concept of mRNA 3'-end formation as a static, minimally regulated housekeeping process has undergone a paradigm shift. Many recent studies have shown that accurate and efficient 3'-end formation of mRNA is highly regulated and that dysregulation of this process is a hallmark of several diseases. While there are many global analysis methods for monitoring altered mRNA processing, methods for investigating specific RNA 3'-end processing events in cells have not significantly changed. Here we describe a facile gain-of-function cellular reporter for the analysis of mRNA 3'-end formation as an alternative to approaches that are technically challenging or use radioactivity. We also offer suggestions for optimization of our approach and enhancement of its reproducibility.

The formation of the 3' end of mRNA is a complex and highly regulated process that is essential for the generation of mature mRNAs (1). In eukaryotic cells, the process involves the recognition of *cis* sequence elements by several *trans*-acting protein factors (2). A key *cis* element essential for this process is the hexanucleotide polyadenylation signal (PAS). Recognition of the PAS, typically AAUAAA, is aided by the presence of other *cis* sequence elements that stimulate cleavage and subsequent polyadenylation of the mRNA. These sequences may include a loosely defined upstream sequence element (USE), a G/U-rich or G-rich downstream sequence element (DSE), and the actual cleavage site itself (3). The combination of these *cis*

regulatory elements creates a biosynthetic context that determines whether an mRNA will be efficiently processed (3–5). Perturbation of this context is evident in the etiologies of a variety of human diseases (6).

Cell-based reporters can be used to analyze mutations in disease-causing genes for their effects on mRNA 3'-end formation, as well as to understand the biological mechanism of 3'-end formation. Reporter systems have been successfully used to study various RNA 3'-end formation events, including the Nrd1/Nab3/Sen2 complex in *Saccharomyces cerevisiae* transcription termination (7), the *cis* and *trans* requirements of histone mRNA 3'-end formation (8), the role of the Integrator complex in small nuclear RNA

(snRNA) 3'-end formation (9,10), and more recently, the effects of the nuclear cap binding complex on stimulation of 3'-end processing of several RNA families (11). In the case of snRNA or histone pre-mRNA processing, the *cis*-acting signals are compact, making reporter design straightforward; however, the processing signals required for cleavage and polyadenylation are much more diverse and cover larger ranges (5). Here, we describe the construction of an effective gain-of-function reporter system to analyze pre-mRNA processing and provide an assessment of the reporter's capabilities and limitations.

To demonstrate the design and utility of an enhanced green fluorescence protein (referred to here as GFP) expression-based assay to monitor mRNA 3'-end processing, we generated a transcriptional readthrough reporter (Figure 1A). This reporter (termed PAS-GFP) is constructed in the pcDNA3.1 (+) vector (Invitrogen, Carlsbad, CA) such that the sequences to be analyzed are cloned upstream of a GFP open reading frame (ORF), which itself is followed by a bovine growth hormone (BGH) PAS. The expectation is that if RNA polymerase II encounters a functional cleavage and PAS prior to transcription of the GFP-encoding mRNA, no fluorescence will be observed. As a positive control, we tested the late SV40 PAS, which has been well-characterized for cleavage and polyadenylation efficiency (12). We inserted a 233-nucleotide sequence that included the SV40 PAS, cleavage site, and DSE into the PAS-GFP reporter. Importantly, we also created two additional negative control reporters where random sequences devoid of defined cleavage and polyadenylation elements were also inserted upstream of the GFP ORF (Figure 1A). All reporter constructs as well as an empty PAS-GFP vector were transiently transfected into HeLa or HEK293T cells (ATCC, Manassas, VA), and both protein and total RNA were isolated for analysis 48 h post-transfection (as described in References 9 and 10). An additional plasmid encoding HA-tagged mCherry was cotransfected to control for transfection efficiency. We

METHOD SUMMARY

We have designed a transcriptional readthrough reporter for gain-of-function detection of mRNA 3'-end processing by enhanced GFP expression. Our method is facile and easily customizable for the visual and non-radioactive analysis of *cis* and *trans* mRNA regulation, in which altered 3'-end formation disrupts normal physiological processes.

Western blots

DNA

Bacterial Colonies

NIR

Proteins

Blue Light

GFP

Image Capture & Analysis

Colorimetric

RGB

Visible

Ultraviolet

RNA

Fluorescent

westerns

Multiplex

Fluorescent

stain-less gels

Immunoblots

Gel Imaging

Arrays

Protein blots

Unlimited Imaging!

An Analytik Jena Company

UVP, LLC USA: (800) 452-6788 | (909) 946-3197
Cambridge UK: +44(0)1223 420022

uvp.com/ChemiDoc-It2delivers

A.

B.

C.

D.

Figure 1. PAS-GFP reporter analysis using an SV40 polyadenylation signal (PAS). (A) Schematic of the transcriptional readthrough reporter designed to study mRNA 3'-end formation. The GFP open reading frame (ORF) without a native AUG start codon is placed downstream of either the SV40 PAS (SV40-GFP) or random sequences derived from the pUC vector (RSeq-GFP). (B) Western blot of lysates from HeLa cells (left) and HEK293T cells (right) transfected with the reporters shown in panel A. The Western blots were probed for GFP protein using α GFP. HAM-Cherry protein expressed from a co-transfected HAMCherry plasmid used to normalize transfection was probed with α HA, and tubulin used as a loading control was probed with α -tub. These controls were used in all blots throughout this study. (C) Quantification of the GFP signal in the Western blot of the HeLa lysates shown in panel B (left) was performed using Image J. (D) qRT-PCR analysis of RNA isolated from the HeLa cells transfected as those analyzed by Western blotting in panel B (left). Error bars show standard deviation from three independent experiments.

observed that only the insertion of the SV40 PAS into the reporter prevented GFP protein expression (Figure 1B). Quantification of GFP expression using ImageJ analysis correlated with qRT-PCR measurement of GFP mRNA (Figure 1, C and D). It is noteworthy that both of the random sequences slightly reduced the expression of the GFP reporter compared with the empty PAS-GFP vector. This is not surprising as some level of translation inhibition would be expected from the insertion of an ~220-nucleotide sequence upstream of the GFP start codon. Along the same lines, it is important to consider that not all random sequences are inert and that they may contain cryptic regulatory elements; therefore, it is recommended to test multiple negative control sequences.

To demonstrate sensitivity of the reporter to mutations of the *cis* regulatory element for 3'-end processing, we introduced either a three-nucleotide mutation (AAUAAA to GGGAAA) into the PAS [PAS null (PN)], a deletion of the DSE [DSE null (DN)], or a mutation of the cleavage site [cleavage null (CN)] (Figure 2A). The most significant readthrough was observed after mutation of the PAS and to a lesser extent the DSE (Figure 2B). It was not unexpected

that the cleavage site mutation did not result in significant readthrough as it is thought to be the least critical element, since alternative cleavage sites can be utilized if one is mutated (3). Importantly, we observed strong correlation between GFP protein expression and GFP mRNA levels, demonstrating that GFP expression is reflective of mRNA production (Figure 2, C and D). Finally, 3' rapid amplification of cDNA ends (RACE) and sequencing confirmed that the SV40 PAS located upstream of GFP and the BGH PAS located downstream of GFP represent the only two cleavage and polyadenylation events, as predicted (Figure 2E). While GFP expression is representative of transcriptional readthrough, we did observe slightly greater sensitivity when measuring readthrough using qRT-PCR, where we measured the GFP mRNA levels normalized to the neomycin resistance mRNA encoded elsewhere within the PAS-GFP plasmid.

One possible concern for a transcriptional readthrough reporter with large inserts upstream is the location and context of the start codon. The constructs shown in Figure 2A all have the optimized GFP start codon mutated and rely on a start codon fortuitously located within

Figure 2. Effect of start codon context on polyadenylation signal (PAS)-GFP reporter analysis. (A) Schematic of SV40 mutant GFP reporter constructs with mutations of the PAS from AUAAAA to GGGAAA (PN), deletion of the downstream sequence element (DSE) (DN), and mutation of the cleavage site from CA to CG (CN). (B) Western blot of lysates from HeLa cells (left) and HEK293T cells (right) transfected with the reporters shown in panel A. (C) Quantification of the GFP signal in the Western blot of the HeLa lysates shown in panel B (left) performed using Image J. (D) qRT-PCR analysis of RNA isolated from the HeLa cells transfected with the reporters shown in panel A. Error bars show standard deviation from three independent experiments. (E) Ethidium bromide-stained agarose gel (top) showing products of 3' rapid amplification of cDNA ends (RACE) of RNA isolated from HeLa cells transfected with the reporters shown in panel A. Sequence chromatogram (bottom) obtained by 3' RACE showing that the SV40 PAS in the reporter uses the cleavage site previously reported in the literature. (F) Schematic of the transcriptional readthrough reporter designed to study mRNA 3'-end formation. The GFP open reading frame (ORF) with a native AUG start codon was placed downstream of the SV40 PAS constructs. (G) Western blot of lysates from HeLa cells transfected with reporter constructs containing the SV40 PASs shown in panel A placed into the reporter plasmid with the native AUG start codon of GFP shown in panel F.

FROM THE
INVENTORS OF THE

1978... The **BEADBEATER**

1993... The **Mini BEADBEATER**

NOW

2015... The **SONI BEAST**
CELL & TISSUE DISRUPTOR

The **NEW SONI BEAST**

- 10X FASTER CELL LYSIS
- SMALL SAMPLE SIZE
- SIMPLE OPERATION
- LARGE CAPACITY

SCAN FOR DETAILS

BIOSPEC PRODUCTS
WWW.BIOSPEC.COM

the SV40 polyadenylation signal. We redesigned the PAS-GFP reporter such that the GFP ORF cloned downstream of the *NotI* site maintains its native AUG (Figure 2F). The same SV40 PAS and its mutants were then tested in this new context to determine if the presence of an optimal codon would alter the results. We did not detect any difference in the effects of the SV40 mutants; however, we did observe preferential usage of the native AUG of the GFP by Western blot analysis (Figure 2G) despite the presence of the upstream start codon. This is most likely because the *NotI* site flanking the endogenous GFP start codon is GC-rich, similar to a Kozak sequence.

Results from recent global analyses of altered RNA 3'-end formation events have uncovered a process that is highly dynamic and subject to regulation (13–15, reviewed in Reference 16). However, these approaches must be followed up using specific reagents to provide a mechanistic understanding of how RNA processing takes place. Typical reporters used by others to investigate mRNA 3'-end formation (17,18) place the 3'-end processing elements downstream of a heterologous ORF (e.g., luciferase or β -globin) and then measure either the protein or mRNA expression. These reporters have proven to be robust, and the primary strength of this design compared with the transcriptional readthrough reporter described in this study is that 3'-end processing occurs in a more native context, which is downstream of an ORF. However, the potential limitation of reporters with 3'-end processing signals placed downstream lies in the interpretation of loss-of-function events. Mutations that are introduced into these reporters that result in failure to produce the reporter product may be confounded by other factors, such as instability of the transcribed RNA due to altered 3' UTR content. This concern is mitigated by the design of our transcriptional readthrough reporters. However, the PAS-GFP reporter described here is also not without limitation. As previously mentioned, cryptic regulatory elements in the 5'-UTR may impair reporter expression. We addressed one of these limitations by exploring the importance of the start codon context (Figure 2). Additional considerations when designing a transcriptional readthrough reporter are the presence of upstream ORFs (uORFs), secondary struc-

tures limiting start codon recognition, and the potential use of a non-AUG start codon (19,20); each of these factors may affect translation of the downstream reporter. These constraints may also place an upper limit on the size of the construct to be used.

Nonetheless, the customizable, gain-of-function PAS-GFP reporter described here offers an immunological and visual output to allow for a case-by-case analysis of mRNA 3'-end formation. Moreover, this reporter is readily adaptable for more complicated model systems, including in vivo expression for tissue-specific analysis of mRNA 3'-end formation and genome-wide CRISPR screening to identify novel regulatory factors.

Author contributions

N.P. performed the experiments and analyzed the data, wrote the manuscript. E.J.W. conceived the project, and contributed to the design of the experiments and writing of the manuscript.

Acknowledgments

This work is supported by funding from the Muscular Dystrophy Foundation (MDA 202141), Friends of FSH Research, and a Schissler Foundation Fellowship for Translational Studies of Common Human Diseases.

Competing interests

The authors declare no competing interests.

References

1. Chan, S., E.-A. Choi, and Y. Shi. 2011. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA*. 2:321-335.
2. Mandel, C.R., Y. Bai, and L. Tong. 2008. Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.* 65:1099-1122.
3. Tian, B. and J.H. Graber. 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA*. 3:385-396.
4. Wilusz, J., T. Shenk, Y. Takagaki, and J.L. Manley. 1990. A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. *Mol. Cell. Biol.* 10:1244-1248.
5. Hu, J., C.S. Lutz, J. Wilusz, and B. Tian. 2005. Bioinformatic Identification of Candidate Cis-Regulatory Elements Involved in Human mRNA Polyadenylation. *RNA* 11:1485-1493.
6. Danckwardt, S., M.W. Hentze, and A.E. Kulozik. 2008. 3[prime] end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* 27:482-498.
7. Steinmetz, E.J., N.K. Conrad, D.A. Brow, and J.L. Corden. 2001. RNA-binding protein Nrd1

directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413:327-331.

8. Yang, X.-C., B.D. Burch, Y. Yan, W.F. Marzluff, and Z. Dominski. 2009. FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs. *Mol. Cell* 36:267-278.
9. Albrecht, T.R. and E.J. Wagner. 2012. snRNA 3' End Formation Requires Heterodimeric Association of Integrator Subunits. *Mol. Cell. Biol.* 32:1112-1123.
10. Chen, J., B. Waltenspiel, W.D. Warren, and E.J. Wagner. 2013. Functional analysis of the integrator subunit 12 identifies a microdomain that mediates activation of the Drosophila integrator complex. *J. Biol. Chem.* 288:4867-4877.
11. Hallais, M., F. Pontvianne, P.R. Andersen, M. Clerici, D. Lener, N.E.H. Benbahouche, T. Gostan, F. Vandermoere, et al. 2013. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* 20:1358-1366.
12. Zarkower, D. and M. Wickens. 1988. A functionally redundant downstream sequence in SV40 late pre-mRNA is required for mRNA 3'-end formation and for assembly of a precleavage complex in vitro. *J. Biol. Chem.* 263:5780-5788.
13. Xia, Z., L.A. Donehower, T.A. Cooper, J.R. Neilson, D.A. Wheeler, E.J. Wagner, and W. Li. 2014. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun.* 5:5274.
14. Shepard, P.J., E.-A. Choi, J. Lu, L.A. Flanagan, K.J. Hertel, and Y. Shi. 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*. 17:761-772.
15. Gruber, A.R., G. Martin, P. Müller, A. Schmidt, A.J. Gruber, R. Gumieny, N. Mittal, R. Jayachandran, et al. 2014. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat Commun.* 5:5465.
16. Shi, Y. and J.L. Manley. 2009. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 29:889-897.
17. Gehring, N.H., U. Frede, G. Neu-Yilik, P. Hundsdoerfer, B. Vetter, M.W. Hentze, and A.E. Kulozik. 2001. Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat. Genet.* 28:389-392.
18. Mayr, C. and D.P. Bartel. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138:673.
19. Araujo, P.R., K. Yoon, D. Ko, A.D. Smith, M. Qiao, U. Suresh, S.C. Burns, and L.O.F. Penalva. 2012. Before It Gets Started: Regulating Translation at the 5' UTR. *Comp. Funct. Genomics* 2013:475731.
20. Ivanov, I.P., A.E. Firth, A.M. Michel, J.F. Atkins, and P.V. Baranov. 2011. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* 39:4220-4234.

Received 25 September 2015; accepted 11 November 2015.

Address correspondence to Eric J. Wagner, Department of Biochemistry and Molecular Biology, University of Texas Medical School, Houston, TX. E-mail: ejwagner@utmb.edu

To purchase reprints of this article, contact: biotechniques@fosterprinting.com