FEBS Letters 589 (2015) 279–284

FEBS Letters

journal homepage: www.FEBSLetters.org

Uncapped 5′ ends of mRNAs targeted by cytoplasmic capping map to the vicinity of downstream CAGE tags

Daniel L. Kiss [a,b], Kenji Oman [a,c], Ralf Bundschuh [a,c,d,e], Daniel R. Schoenberg [a,b,*]

[a] Center for RNA Biology, The Ohio State University, Columbus, OH 43210, United States
[b] Department of Molecular and Cellular Biochemistry, The Ohio State University, Columbus, OH 43210, United States
[c] Department of Physics, The Ohio State University, Columbus, OH 43210, United States
[d] Department of Chemistry & Biochemistry, and Division of Hematology, The Ohio State University, Columbus, OH 43210, United States
[e] Department of Internal Medicine, The Ohio State University, Columbus, OH 43210, United States

ARTICLE INFO

Article history:
Received 1 October 2014
Revised 9 December 2014
Accepted 10 December 2014
Available online 23 December 2014

Edited by Michael Ibba

Keywords:
mRNA capping
Cytoplasm
CAGE
Uncapped 5′ end
Cytoplasmic capping
Mammalian cell

ABSTRACT

In mammalian transcriptomes approximately 25% of 5′ ends determined by Capped Analysis of Gene Expression (CAGE) map to locations within spliced exons. The current study sought to determine if the cytoplasmic capping complex participates in generating these downstream CAGE tags. 5′-RACE was used to amplify the uncapped ends of target transcripts that accumulate when cytoplasmic capping is blocked. Sequencing of these RACE products mapped the positions of uncapped ends either exactly to or just downstream of archived CAGE tags. These findings support a role for cytoplasmic capping in generating the downstream capped ends identified by CAGE.

1. Introduction

The 5′ ends of all mRNAs and lncRNAs have a methylguanosine (m7G) 'cap' that is added co-transcriptionally as the first step of post-transcriptional processing. This served as the basis for development of Capped Analysis of Gene Expression (CAGE) as a method of identifying transcription start sites. However, in the first broad scale application of CAGE to mammalian transcriptomes a large number of capped ends were identified that did not match to known sites of transcription initiation [1]. At the time, we were studying the decay of non-sense-containing β-globin mRNA in erythroid cells [2]. In these cells, non-sense-containing human β-globin mRNA is cleaved by SMG6 to generate metastable intermediates [3] that were described as having a 5′ cap [4]. The concept of capped decay products raised several questions, including whether these indeed had an m7G cap and how this might be added when all of the proteins that catalyze capping were thought to be present only in the nucleus. Further complicating matters was the fact that capping enzyme transfers covalently-bound GMP onto RNA with a 5′-diphosphate end, and there was no known mechanism for generating this substrate from 5′-monophosphate RNA. In pursuing this question we identified the cytoplasmic complex that includes capping enzyme and a kinase that generates a capping substrate from 5′-monophosphate RNA [5]. We subsequently showed that the cytoplasmic capping complex assembles on adapter protein Nck1, a protein with 3 SH3 domains and a single SH2 domain that functions primarily in transducing receptor tyrosine kinase signaling [6]. Importantly, Nck1 is only found in the cytoplasm, and the binding of capping enzyme and the 5′-kinase to adjacent SH3 domains juxtaposes these two critical enzymes in a manner that facilitates cytoplasmic capping.

In [7], cytoplasmic capping targets were identified by the susceptibility of uncapped 5′ ends to in vitro degradation by Xrn1 when recovered from cells expressing a dominant negative form of capping enzyme, termed K294A, that was modified to restrict its distribution to the cytoplasm. Three classes of targets were identified by position-dependent changes in probe intensity on human exon arrays, two of which accumulate uncapped forms

Abbreviation: CAGE, Capped Analysis of Gene Expression

* Corresponding author at: Department of Molecular and Cellular Biochemistry, The Ohio State University, 333 Hamilton Hall, 1645 Neil Ave., Columbus, OH 43221, United States. Fax: +1 614 292 4118.
E-mail address: schoenberg.3@osu.edu (D.R. Schoenberg).

http://dx.doi.org/10.1016/j.febslet.2014.12.009

when cytoplasmic capping is blocked. The accumulation of uncapped forms of these mRNAs was confirmed by 4 independent methods; increased susceptibility to in vitro degradation by Xrn1, selective recovery of uncapped RNAs following ligation of an RNA adapter and hybridization to a biotin-tagged antisense DNA, selective exclusion from a cap affinity column containing a heterodimer of eIF4E bound to eIF4G, and the appearance of products by 5′-RACE only when cytoplasmic capping is blocked. The latter proved to be particularly important in that it identified full-length transcripts and 5′-truncated forms of the same RNAs.

We wondered if the uncapped ends of shorter transcripts that appeared by 5′-RACE of RNA from capping inhibited cells might correspond to downstream capping sites identified by CAGE. Using positional data of CAGE tags from ENCODE [8] we designed primers to several of the transcripts for which shortened forms appeared by 5′-RACE [7]. We show that uncapped ends that accumulate when cytoplasmic capping is blocked map either at or near CAGE tags, thus providing the first direct evidence of a functional role for cytoplasmic capping in generating this form of transcriptome diversity.

2. Materials and methods

2.1. Downstream CAGE tag correlations

Poly-A +/−, hg19-aligned cytoplasmic CAGE tags were downloaded from the UCSC FTP data server [9] for the K562 cell line (Table 1). Reads were combined across the 4 samples (one poly-A−, three poly-A+ samples), and genomic coordinates for 5′-ends of reads were mapped to transcript coordinates for Gencode v19 transcripts [10], downloaded from the UCSC Table Browser [11]. Transcripts in the top quartile of total CAGE expression (which had no significant difference in total CAGE tags across categories, compared using Student's *t*-test) were classified as recapping targets or controls based on [7], and as containing a downstream CAGE tag when at least one location downstream of the annotated translation start site had a minimum CAGE coverage of 10 reads. Fisher's exact test was performed on the distribution of the number of transcripts across these categories using the R Statistical Computing Package, version 3.1.1 [12].

2.2. Cell culture and preparation of cytoplasmic RNA

Tetracycline-inducible U2OS cells stably transfected with pcDNA4/TO/myc-K294-ΔNLS+NES-Flag (K294A) were cultured in McCoy's medium (Gibco) supplemented with 10% fetal bovine serum [5,7]. 3×10^6 log-phase cells were split into 150 mm tissue culture dishes followed 24 h later by addition of 1 μg/ml doxycycline to induce K294A. The medium was removed 24 h later, the cells were rinsed twice with ice-cold phosphate buffered saline and suspended with a cell scraper. These were recovered by centrifugation for 5 min at 1000×*g*, the pellet was suspended in 5 volumes of lysis buffer (20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.2% NP-40, 80 U/ml RNaseOUT (Invitrogen)) and incubated on ice for 10 min with gentle agitation. Nuclei

Table 1
CAGE libraries used.

wgEncodeRikenCageK562CytosolPapAln.bam	Fig. 1A
wgEncodeRikenCageK562CytosolPapAlnRep1.bam	Fig. 1A
wgEncodeRikenCageK562CytosolPapAlnRep2.bam	Fig. 1A
wgEncodeRikenCageK562CytosolPamAln.bam	Fig. 1A
wgEncodeRikenCageAlignmentsGm12878CytosolLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsGm12878NucleolusTotal	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsGm12878NucleusLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsH1hescCellLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsHepg2CytosolLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsHepg2NucleolusTotal	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsHepg2NucleusLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsHuvecCytosolLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562ChromatinTotal	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562CytosolLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562CytosolLongpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562NucleolusTotal	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562NucleoplasmTotal	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562NucleusLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562NucleusLongpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsK562PolysomeLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsNhekCytosolLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsNhekNucleusLongnonpolya	Figs. 1B–D, 3 and 4
wgEncodeRikenCageAlignmentsProstateCellLongnonpolya	Figs. 1B–D, 3 and 4

Table 2
Oligonucleotides used.

Oligonucleotide name	Sequence	Use(s)
RNA_RACE_Adaptor	GUUCAGAGUUCUACAGUCCGACGAUC	Ligation reaction
RACE_Forward	GTTCAGAGTTCTACAGTCCGACGATC	1° and 2° nested PCR
ZNF207_nTail (1725-1699)Rev	GTCCAATAGTTCCTGGTATGTGGGAAG	cDNA synthesis and 1° nested PCR
SARS_3′ (1838-1815)Rev	ATCAATGATGGGTCCCTATGCCCA	cDNA synthesis and 1° nested PCR
ITGB1_sTail (3678-3653)Rev	GGGCAACTCAAATGGTGAGAAGTAAA	cDNA synthesis and 1° nested PCR
ZNF207-A (326-301)Rev	ACCTGCATGCAATGAATAGCTAAGC	2° nested PCR
ZNF207-B (649-627)Rev	TGGCATTAATGGAGGTATGCCT	2° nested PCR
ZNF207-C (1192-1169)Rev	GTACTATTTAAGGGTTTGAAATC	2° nested PCR
SARS (160-134)Rev	CTTGAAGCGCTTCTCCTGCGTCTCTC	2° nested PCR
ITGB1-A (469-446)Rev	AATTTTAATGTAAATGTCTGTGG	2° nested PCR
ITGB1-B (1705-1681)Rev	TTTCATTTGTATTATCCCTCTTCC	2° nested PCR
ITGB1-C (2296-2273)Rev	CCAATAAGAACAATTCCAGCAAC	2° nested PCR