

A position-specific 3'UTR sequence that accelerates mRNA decay

Rene Geissler and Andrew Grimson

Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

ABSTRACT

The 3' untranslated regions (3'UTRs) of mammalian mRNAs direct an extensive range of alternative post-transcriptional outcomes, including regulation of mRNA decay and translation, contributing significantly to overall gene regulation. However, our knowledge of the underlying sequences and mechanisms is incomplete. We identified a novel 3'UTR sequence motif in mammals that targets mRNAs for transcript degradation. The motif is found in hundreds of mRNAs and is enriched in transcripts encoding regulatory proteins, such as transcription and signaling factors. Degradation of mRNAs containing the motif is mediated by the CCR4-NOT deadenylation complex. We identified hnRNPs A1 and A2/B1 as *trans* factors that directly bind to the motif, indicating a novel role for these proteins in deadenylation. Interestingly, a genome-wide analysis of the impact of this new regulatory pathway showed that the most active motifs are located within the 5' and 3'-terminal portions of 3'UTRs, whereas elements in the center tend to be inactive. The highly position-specific function of the motif adds a new layer of regulation to gene expression mediated by 3'UTRs.

ARTICLE HISTORY

Received 8 July 2016
Accepted 8 August 2016

KEYWORDS

3'UTR; CCR4-NOT; *cis*-regulatory element; deadenylase complex; hnRNP A2/B1 and A1; mRNA decay; post-transcriptional gene regulation

The control of gene expression is fundamental to cell function. Although much of this control occurs at the level of transcription, post-transcriptional control is both prevalent and consequential.¹ In mammals, the 3' untranslated region of an mRNA (3'UTR) plays a major role in determining the mRNA's post-transcriptional fate.² Mammalian 3'UTRs contain almost as much conserved sequence as that ascribed to transcriptional regulatory elements, yet much of this conservation cannot be attributed to known regulatory mechanisms. Furthermore, the degree of sequence conservation is greater for mammalian 3'UTRs than for those of other eukaryotes.^{3,4} Notably, transcriptome-wide mapping of protein binding sites demonstrates that 3'UTRs are bound by a diverse set of several hundred different proteins, most of which are uncharacterized.^{5,6} Therefore, it is clear that 3'UTR regulation involves myriad *cis* sequences and *trans* factors, many of which remain to be discovered and characterized. Thus, an essential component of understanding 3'UTR-mediated gene regulation is to decipher the mechanistic details of specific regulatory pathways triggered by *cis*-regulatory sequences within them.

We set out to understand the cellular function and underlying molecular mechanism of a novel 3'UTR motif (UAAC/GUUAU), which drew our attention for its striking prevalence (~7% of mammalian 3'UTRs contain one or more copy) and its strong conservation. We demonstrated that the motif is repressive and established that the core 8 nucleotide sequence is both necessary to mediate regulation and sufficient when tested in many different sequence contexts. Moreover, the motif is active only within 3'UTRs and is ineffective in either introns or 5'UTRs. We found that regulation triggered by the motif acts by accelerated transcript degradation, via mRNA deadenylation mediated by the CCR4-NOT complex, the major mRNA

deadenylase complex (Fig. 1).⁷ We purified *trans* factors that recognize the motif and identified the hnRNPs A1 and A2/B1, which we confirmed bind directly to the motif with high specificity. To examine the role of these *trans* factors genome-wide, we profiled the transcriptomes of cells depleted for hnRNP A1 and A2/B1 using RNAi. Our results confirm that motif-dependent repression requires these *trans* factors and indicate that hundreds of transcripts harboring the motif are subject to regulation by hnRNP A1 and A2/B1.

Like many RNA binding proteins, both hnRNP A1 and A2/B1 possess a variety of different cellular roles, including mRNA processing, export, stability and translation.^{8–12} Our work has established a new role for these proteins. The two proteins are paralogs and, together with hnRNP A0 and A3, belong to the A/B family of heterogeneous nuclear ribonucleoproteins. The A/B family represents the most abundant cellular hnRNP proteins, with ~3×10⁷ copies of hnRNP A2/B1 in HeLa cells.^{12,13} Furthermore, mutations in hnRNP A1 and A2/B1 are strongly associated with human diseases, including cancers and neurological disorders,^{14,15} and abnormal regulation of these proteins is observed in several cancers, including lung cancer.¹⁶

The impact of regulatory pathways can be gauged in several ways; informative metrics include the quantitative outcome of regulation and the number and types of targets impacted. The UAAC/GUUAU motif mediates approximately 2-fold repression, which, although subtle, is comparable to that typically mediated by individual miRNA target sites.¹⁷ The motif is found in the 3'UTRs of ~7% of human genes and is frequently deeply conserved, with conserved instances of the motif occurring more than twice as frequently as expected by chance. Thus, in terms of prevalence and conservation, this motif is comparable to binding sites for some of the most deeply

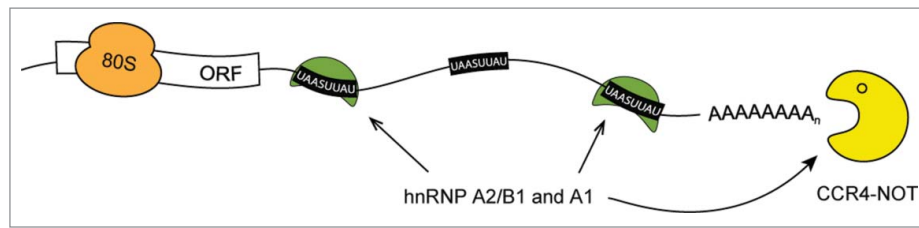


Figure 1. Model of UAASUUU-mediated mRNA deadenylation. hnRNP A2/B1 and A1 bind to UAASUUU (black rectangles, S = G or C) preferentially at the 5' and 3' edges of 3'UTRs (first and last ~300 nt) and recruit the CCR4-NOT complex to mRNAs to promote deadenylation.

conserved miRNAs, sites that are themselves among the most prevalent and strongly conserved sequences within 3'UTRs.⁴ It is worth noting that transcripts encoding regulatory and signaling proteins are enriched in the motif, which implies that the pathway we have found contributes to large and complex regulatory networks. We also found effective instances of the motif in 3'UTRs of multiple core RNA decay enzymes, including the 5' exonuclease XRN1, the decapping complex (DCP1a) and the CCR4-NOT deadenylase complex itself (CNOT1/4/6). Together, these results imply a broad regulatory impact for the pathway we have discovered, including auto-regulation of mRNA decay.

The sensitivity of the functional impact of the UAAC/GUUU motif to its location within the 3'UTR is striking. Indeed, analysis of transcriptome profiling to discover functional sites suggests that only motifs located within the terminal portions of 3'UTRs (first and last ~300 nucleotides) are competent to attenuate transcript abundance. This finding is reminiscent of miRNA target sites, which tend to be more effective when located toward the ends of 3'UTRs rather than in the center, although the degree of position-specificity appears significantly stronger for the UAAC/GUUU motif than for miRNA target sites.^{17,18} For both miRNA target sites and the UAAC/GUUU motif, the physical or biochemical mechanisms underlying their position-dependent effects remain unclear – investigating and understanding these mechanisms represent important opportunities for insights into the biology of 3'UTRs.

Several possible models could explain the positional sensitivity of 3'UTR regulatory elements. First, the central regions of 3'UTRs might be more likely to form occlusive secondary structures, thereby limiting access of *trans* factors to regulatory elements. Second, proximity of the motifs to the poly(A) tail and proteins that interact with the 3' terminal mRNP might potentiate certain regulatory pathways. For example, it is easy to envision that *trans* factors found close to the 3' terminus of a 3'UTR could have more effective interactions with PABP, deadenylases, or even the translation initiation machinery, when compared with such interactions with same *trans* factor bound elsewhere in a 3'UTR. These models are not mutually exclusive, nor may a single explanation apply to all sequence motifs. Nevertheless, it is likely that mechanistic insights into the phenomenon will be valuable in building toward a more comprehensive understanding of 3'UTR mediated gene regulation.

The enhanced efficacy of the UAAC/GUUU motif when located toward the 5' terminus of a 3'UTR is more difficult to rationalize. The most prevalent internal mRNA modification of higher eukaryotes is m⁶A (N6-methyladenosine). This

modification has been implicated in affecting diverse aspects of post-transcriptional gene regulation, including mRNA translation, decay, splicing, alternative polyadenylation, and heat shock response.^{19–24} Transcriptome-wide maps indicate that m⁶A is found in at least ~50% of mammalian mRNAs. More importantly, m⁶A is preferentially found within the 5' end of 3'UTRs, suggesting a regulatory role restricted to a certain region within 3'UTRs. Importantly, analyzing CLIP-seq data,⁹ we found that hnRNP A2/B1 and A1 preferentially bind elements at the 5' end of 3'UTRs, indicating a potential connection to m⁶A sites (Fig. 2). Consistent with this model, hnRNP A2/B1 was recently identified as an m⁶A-binding protein.²⁵ We are currently exploring whether the abundance of m⁶A modifications close to the stop codon results in more effective recruitment (or activity) of hnRNP A2/B1 to UAAC/GUUU motifs located near the 5' end of 3'UTRs.

Irrespective of the mechanistic basis for the position effect of the UAAC/GUUU motif and other regulatory elements in 3'UTRs, there are clear implications when the prevalence of alternative 3'UTR isoforms are considered. Alternative splicing and, even more so, alternative polyadenylation can alter the length and composition of 3'UTRs for different transcripts of the same gene.^{26,27} Thus, a single gene may gain or lose *cis*-regulatory sites and consequential post-transcriptional gene regulation due to alternative mRNA processing. In addition, because the pathway we have found is so sensitive to the position of the motif within the 3'UTR, we suspect that alternative 3'UTR processing may have even greater impacts upon transcripts harboring the motif than for other *cis*-regulatory sites in 3'UTRs. That is, depending on the location of a motif, and the

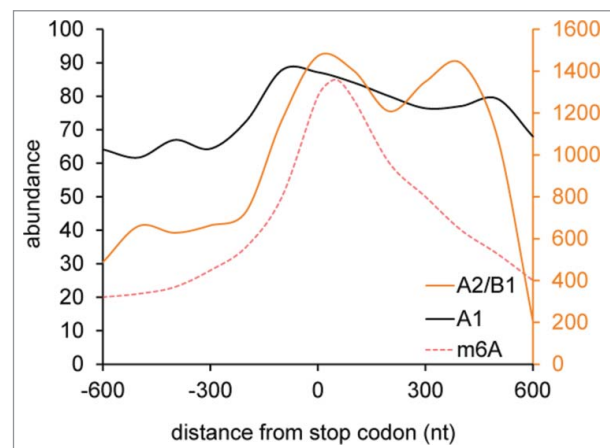


Figure 2. Abundance of genome-wide binding sites for hnRNP A2/B1 and A1 (region –600 to 600 nt from the stop codon) within mRNAs analyzed using public CLIP-seq datasets,⁹ m⁶A abundance according to Ke et al.²⁰

specific 3'UTR isoform(s) produced, alternative transcripts of the same gene will gain or lose sensitivity to the pathway we describe even in conditions where different isoforms all contain the motif. Moreover, even though hnRNP A1 and A2/B1 are expressed ubiquitously,¹² dynamic regulation dependent on the motif will still occur as a result of alternative 3'UTR processing.

The 5' and 3' untranslated regions of human mRNAs contain hundreds of different cis-regulatory elements, creating an extraordinarily complex collection of post-transcriptional events with the potential to significantly impact gene regulation. Despite the remarkable progress in understanding post-transcriptional biology, it seems likely, in addition to position-specific effects of sequence motifs, that many novel layers affecting 3'UTR regulation remain to be discovered. Although the last decade has seen remarkable progress toward understanding the roles of 3'UTRs in gene regulation, the abundance of new approaches and techniques suited to the study of gene regulation will likely propel many additional insights into 3'UTR biology in the near future.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

References

- Kuersten S, Goodwin EB. The power of the 3' UTR: translational control and development. *Nat Rev Genet* 2003; 4:626-37; PMID:12897774; <http://dx.doi.org/10.1038/nrg1125>
- Matoulkova E, Michalova E, Vojtesek B, Hrstka R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol* 2012; 9:563-76; PMID:22614827; <http://dx.doi.org/10.4161/rna.20231>
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005; 15:1034-50; PMID:16024819; <http://dx.doi.org/10.1101/gr.3715005>
- Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 2005; 434:338-45; PMID:15735639; <http://dx.doi.org/10.1038/nature03441>
- Baltz AG, Munschauer M, Schwanhauser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* 2012; 46:674-90; PMID:22681889; <http://dx.doi.org/10.1016/j.molcel.2012.05.021>
- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 2012; 149:1393-406; PMID:22658674; <http://dx.doi.org/10.1016/j.cell.2012.04.031>
- Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 2007; 8:113-26; PMID:17245413; <http://dx.doi.org/10.1038/nrm2104>
- Goodarzi H, Najafabadi HS, Oikonomou P, Greco TM, Fish L, Salavati R, Cristea IM, Tavazoie S. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* 2012; 485:264-68; PMID:22495308; <http://dx.doi.org/10.1038/nature11013>
- Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S, et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* 2012; 1:167-78; PMID:22574288; <http://dx.doi.org/10.1016/j.celrep.2012.02.001>
- Pinol-Roma S, Dreyfuss G. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* 1992; 355:730-32; PMID:1371331; <http://dx.doi.org/10.1038/355730a0>
- Svitkin YV, Ovchinnikov LP, Dreyfuss G, Sonenberg N. General RNA binding proteins render translation cap dependent. *EMBO J* 1996; 15:7147-55; PMID:9003790
- He Y, Smith R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci* 2009; 66:1239-56; PMID:19099192; <http://dx.doi.org/10.1007/s00018-008-8532-1>
- Singh G, Pratt G, Yeo GW, Moore MJ. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem* 2015; 84:325-54; PMID:25784054; <http://dx.doi.org/10.1146/annurev-biochem-080111-092106>
- Honda H, Hamasaki H, Wakamiya T, Koyama S, Suzuki SO, Fujii N, Iwaki T. Loss of hnRNPA1 in ALS spinal cord motor neurons with TDP-43-positive inclusions. *Neuropathology* 2015; 35:37-43; PMID:25338872; <http://dx.doi.org/10.1111/neup.12153>
- Liu X, Zhou Y, Lou Y, Zhong H. Knockdown of HNRNPA1 inhibits lung adenocarcinoma cell proliferation through cell cycle arrest at G0/G1 phase. *Gene* 2016; 576:791-7; PMID:26581508; <http://dx.doi.org/10.1016/j.gene.2015.11.009>
- Boukakis G, Patrinoiu-Georgoula M, Lekaraku M, Valavanis C, Guialis A. Deregulated expression of hnRNP A/B proteins in human non-small cell lung cancer: parallel assessment of protein and mRNA levels in paired tumour/non-tumour tissues. *BMC Cancer* 2010; 10:434; PMID:20716340; <http://dx.doi.org/10.1186/1471-2407-10-434>
- Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; 27:91-105; PMID:17612493; <http://dx.doi.org/10.1016/j.molcel.2007.06.017>
- Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* 2007; 8:69; PMID:17331257; <http://dx.doi.org/10.1186/1471-2105-8-69>
- Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 2012; 485:201-06; PMID:22575960; <http://dx.doi.org/10.1038/nature11112>
- Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, Haripal B, Zucker-Scharff I, Moore MJ, Park CY, et al. A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev* 2015; 29:2037-53; PMID:26404942; <http://dx.doi.org/10.1101/gad.269415.115>
- Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, Pestova TV, Qian SB, Jaffrey SR. 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* 2015; 163:999-1010; PMID:26593424; <http://dx.doi.org/10.1016/j.cell.2015.10.012>
- Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 2014; 505:117-20; PMID:24284625; <http://dx.doi.org/10.1038/nature12730>
- Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* 2015; 161:1388-99; PMID:26046440; <http://dx.doi.org/10.1016/j.cell.2015.05.014>
- Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian SB. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* 2015; 526:591-4; PMID:26458103; <http://dx.doi.org/10.1038/nature15377>
- Alarcon CR, Goodarzi H, Lee H, Liu X, Tavazoie S, Tavazoie SF. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell* 2015; 162:1299-308; PMID:26321680; <http://dx.doi.org/10.1016/j.cell.2015.08.011>
- Bicknell AA, Cenik C, Chua HN, Roth FP, Moore MJ. Introns in UTRs: why we should stop ignoring them. *Bioessays* 2012; 34:1025-34; PMID:23108796; <http://dx.doi.org/10.1002/bies.201200073>
- Mayr C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol* 2015; 26:227-37; PMID: 26597575; <http://dx.doi.org/10.1016/j.tcb.2015.10.012>