Functional premature polyadenylation signals and aberrant splicing within a recombinant protein coding sequence limit expression

Sujeewa D. Wijesuriya, Robyn L. Cotter, Arnold H. Horwitz *

XOMA Corporation, 2910 Seventh Street, Berkeley, CA 94710, United States

A R T I C L E   I N F O

Article history:
Received 6 May 2013
and in revised form 19 July 2013
Available online 28 August 2013

Keywords:
Expression
Optimization
RNA
Polyadenylation
Aberrant
Splicing

A B S T R A C T

Recombinant glycoproteins can be produced at high levels in permanently transfected mammalian cells using expression vectors with strong viral promoters. CHO-K1 cell lines developed to produce the recombinant complement activator blocking protein, CAB-2 (a fusion of membrane co-factor protein, MCP, and decay accelerating factor, DAF), showed unexpectedly low expression. Northern blot analysis revealed that in addition to the expected 2300 base CAB-2 mRNA species, these cell lines expressed 790 and 1500 base mRNA species accounting for ~50% and ~10% of the total CAB-2 mRNA, respectively. RT-PCR studies established that the 1500 base species resulted from aberrant splicing from within the DAF region of the CAB-2 coding sequence to a site within the 3' untranslated region. 3' RACE analysis confirmed that the 790 base species resulted from premature polyadenylation at an AATAAA site within the MCP coding region of CAB-2. Another prematurely polyadenylated species, not observed on Northern blots, was observed in the DAF region by 3' RACE. Analysis of human tissues and cell lines revealed that these internal polyadenylation signals in native MCP and DAF coding regions also generated prematurely polyadenylated mRNAs. Genetic modification of these functional RNA processing elements within the CAB-2 gene eliminated the aberrant mRNA species and significantly increased recombinant CAB-2 expression. These results illustrate that protein expression can be limited by aberrant mRNA processing and demonstrate the importance of identifying and eliminating these mRNA processing signals from within coding DNA to maximize recombinant protein expression.

© 2013 The Authors. Published by Elsevier Inc. Open access under CC BY license.

Introduction

A common strategy employed within the biotechnology industry is to develop expression systems containing genetic elements proven to yield high – level productivity and to employ these systems for production of numerous products. Productivity, however, can vary from protein to protein, often leading to significant development challenges for poorly-expressed proteins. In some cases, the cause of poor expression has been determined [1,2], while in others, the causes remain unresolved [3,4].

We have utilized a CHO-K1 production system in conjunction with a sequential transfection strategy using expression vectors containing the CMV promoter and the neo gene encoding G418 resistance for the first transfection and the hisD gene encoding histidinol resistance for the second transfection to stably express the soluble chimeric complement activator blocker protein, CAB-2. CAB-2 consists of portions of two human regulators of complement activation (RCAs), membrane cofactor protein (MCP) and decay accelerating factor (DAF) [5]. Native MCP and DAF are both plasma membrane-associated molecules that contain four short consensus repeat (SCR) motifs, a Serine–Threonine–Proline (STP) rich region, a transmembrane domain and a cytoplasmic tail. CAB-2 retains the signal sequence, 4 SCR units and the first 2 amino acids of the STP region of MCP fused to the 4 SCR units and STP region of DAF (Fig. 1). CAB-2 possesses in vitro complement inhibition activities greater than the individual activities of soluble forms of MCP or DAF alone and acts by blocking both the C3 and C5 convertases which are essential components of the complement activation pathway thus preventing complement – mediated tissue damage in vivo [5,6]. Complement activation is believed to contribute to harmful inflammatory responses which can occur in cardiopulmonary bypass procedures resulting in significant complications [6]. CAB-2 was

1 Abbreviations used: RCAs, regulators of complement activation; MCP, membrane cofactor protein; DAF, decay accelerating factor; SCR, short consensus repeat; STP, Serine–Threonine–Proline; AP, Adapter; AUAP, Abridged Universal Adapter Primer.
developed as a therapeutic to reduce the incidence of these complications by inhibiting complement activation.

In our CHO-K1 production system, sequential transfection to increase gene copy number typically can yield up to a twofold increase in expression from the first to second transfection. For CAB-2, the top-producing G418- and histidinol- resistant sequential transfectants expressed only modestly higher levels (~20–30%) than the initial top G418-resistant transfectants. The lower than expected expression levels suggested that transcriptional, translational or secretional factors could be limiting protein expression. To determine if transcriptional factors were a cause for the relatively poor expression, we performed a detailed analysis of CAB-2 mRNA from CHO-K1 cells transfected with expression vectors containing CAB-2 cDNA. We found that functional polyadenylation sequences within the MCP and DAF coding regions of the CAB-2 cDNA as well as aberrant splicing originating from a cryptic splice donor site within DAF produced truncated RNA species which comprised a significant proportion of the total CAB-2-specific mRNA. We also show that modification of the CAB-2 cDNA to eliminate sequences within the MCP and DAF coding regions of the CAB-2 cDNA produced CAB-2 mRNA from CHO-K1 cells transfected with expression vectors containing two copies of the CAB-2 gene and the selectable marker (hisD) encoding resistance to histidinol.

Cell culture

CHO-K1 cells were maintained in EX-CELL® 301 medium (SAFC Biosciences, Lenexa, KS). Transfectants were selected and propagated in Ex-Cell 301 or EX-CELL® 302 medium (SAFC Biosciences, Lenexa, KS) containing 0.8 g/L G418 (Invitrogen, Carlsbad, CA) or 0.4 g/L G418 plus 8 mM histidinol (Sigma Chemicals, St Louis, MO). Culture supernatants from CAB-2-expressing clones were evaluated by ELISA at the 96-well and 24-well screening stages. The top-producing clones derived from single colonies were further evaluated in shake flask tests performed in selective agent-free media at a starting cell density of $2 \times 10^5$ cells/ml. Extinct culture supernatants were assayed on Day 14 by ELISA.

Elisa

ELISA plates were pre-coated with a murine monoclonal IgG, against MCP. After blocking with PBS containing 1% BSA and 0.5% TWEEN® 20, plates were incubated with cell culture supernatants. This was followed by a two-step incubation first with a polyclonal rabbit anti-CAB-2 antibody followed by a mouse anti-rabbit antibody conjugated to horse radish peroxidase. The plates were developed using the TMB detection system (KPL, Gaithersburg MD) and read on a Biotek ELISA plate reader at 450 nm.

Northern blotting

Total RNA was isolated from cell lines using TRI-reagent (Molecular Research, Cincinnati, OH). RNA from CAB-2–producing clones (20 µg each) was run on an agarose-formaldehyde gel, blotted and hybridized with a $^{32P}$–labeled cDNA probe (Fig. 1) using the NorthernMax® Kit (Ambion Inc, Austin, TX). The integrity and uniformity of the RNA was confirmed by stripping blots and re-hybridizing with a probe for glyceraldehyde phosphate dehydrogenase (GAPDH) as an internal expression control.
RT-PCR

Total RNA, isolated as described above, was treated with DNase I to remove any contaminating DNA. Random hexamer (1 μg) and oligo dt (0.5 μg) primers were used for first-strand cDNA synthesis using SuperScript II reverse transcriptase (Invitrogen Corp, Carlsbad, CA) and 2 μg of each DNase-I–treated RNA. PCR was performed using 2 μg of first-strand cDNA and primers (P1, P2 and P3) in the 5’ and 3’ untranslated regions flanking the CAB-2 cDNA (see Fig. 1).

3' RACE

3’ RACE was performed using a 3’ RACE kit (Invitrogen Corp, Carlsbad, CA). Total RNA isolated from CAB-2 – transfected cells, cell lines and various tissues was reverse-transcribed into cDNA using the Adapter (AP) primer. PCR was performed using the Abridged Universal Adapter Primer (AUAP) and gene-specific forward primers (P4, P5, P6 and P7) located within MCP and DAF (see Fig. 1) to amplify any prematurely polyadenylated transcripts. The resulting PCR products were verified by DNA sequence analysis.

Site-directed mutagenesis

Changes in the CAB-2 gene to correct the two functional, internal AATAAA sites and the splice donor site were made using inverse PCR and DpnI digestion [7]. Complementary primers overlapping the site of change were used to PCR amplify the whole plasmid in a mixture containing 500 mM dNTPs, 2 units of Pfu polymerase, 250 ng each of sense and antisense primers and 200 ng of CAB-2 plasmid DNA. PCR involved 18 cycles with an extention time of 2.5 min for each Kb of DNA. PCR products were treated with DpnI (which only digests the adenine-methylated plasmid DNA) and transformed into Escherichia coli DH5α cells. Transformants were evaluated by restriction enzyme digestion for incorporation of the changes and confirmed by DNA sequence analysis.

Results

To maximize CAB-2 expression, CHO-K1 cells first were transfected with a vector containing the CAB-2 gene and the neo gene encoding resistance to G418. The top G418-resistant CAB-2-producing clones then were re-transfected with a vector containing the CAB-2 gene and the neo gene encoding resistance to G418. The top G418-resistant CAB-2-producing clones expressed only 20–30% higher levels than the G418 onward transfection, the top G418- and histidinol-resistant CAB-2 producing transfectants which would result from polyadenylation at the AAATAAA site in MCP (see Fig. 1). DNA sequence analysis of this PCR product revealed that it terminates shortly after the internal MCP AAATAAA sequence. Polyadenylation at this site would account for the abundant ~790 base transcript observed by Northern blot analysis (see Fig. 2). The distances between the premature polyA signal identified in MCP and the first and second premature polyA signals identified in DAF and the native poly A signal are 1000, 1300, and 1600 base pairs, respectively. The short PCR extension time used for amplification to the MCP poly site would not be expected to amplify the larger products terminated at the DAF or native poly A sites. 3’ RACE using primer P5, located 342 bp upstream of the AAATAAA site in DAF SCR4, should capture RNA polyadenylated at the two internal DAF AAATAAA sites and at the AAATAAA site in the 3’ UTR of the vector. The results (Fig. 3B) demonstrated that only the AAATAAA sequence located in DAF SCR4 was used, resulting in a truncated ~380 bp PCR product. As expected, polyadenylation at the AAATAAA site in the 3’ UTR yielded a ~890 bp PCR product. These results were confirmed by DNA sequence analysis. It should be noted that the aberrantly spliced mRNA (described below) would not be detected in Fig. 3B since the forward primer, P5 (see Fig. 1), is downstream of the splice donor site.

We next investigated whether premature polyadenylation was unique to the chimeric CAB-2 protein or whether it also occurs in tissues and cells that express native MCP and DAF proteins. 3’ RACE using primer P6 located in the MCP signal sequence identified products of a size indicating use of the MCP internal AAATAAA site in salivary gland, peripheral blood leukocytes (PBL), placenta and the human T–cell line, HSB2, (Fig. 3C). DNA sequence analysis of the PCR products from the placenta and HSB2 cells confirmed them to be products from prematurely – truncated MCP mRNA. Similarly, using primer P7, 3’ RACE PCR products of a size (~340 bp) suggesting use of the internal AAATAAA site in DAF SCR4 were faintly observed in the salivary gland and PBL (Fig. 3D). DNA sequence analysis confirmed that only the PCR product from the salivary gland resulted from prematurely polyadenylated mRNA. PCR products (~800 bp) corresponding to correct polyadenylation at the AAATAAA site in the 3’ UTR for the DAF gene also were observed in both tissues (Fig. 3D).

RT-PCR was used to identify the source of the 1500 base RNA species observed by Northern blot analysis. Primers P1 and P2
(located immediately upstream and downstream of the transcription unit) amplified an expected cDNA of 2000 base pairs from the full length 2300 base mRNA (Fig. 4A, panel a). However, when reverse primer P3, which lies further downstream in the 3' UTR of the vector, was used in combination with P1, a predominant product of 1500 bases was observed (Fig. 4A, panel b). DNA sequencing confirmed that this product resulted from aberrant splicing from a non-canonical donor site in SCR2 of DAF to an acceptor site in the 3' UTR of the vector (Fig. 4B). No such aberrant splicing was detected in RNA isolated from any natural source expressing DAF (data not shown), suggesting that this process is unique to the CAB-2 fusion protein expressed using our expression vector system in CHO-K1 cells.

To eliminate the truncated mRNA species, the two functional internal poly(A) sequences and the non-canonical splice donor sequence were modified by replacing two nucleotides at each site without changing the native protein coding sequence. The functional MCP and DAF AATAAA sites were changed to AAGATC and AACAAG, respectively. The splice donor site encoding amino acids 390–391 was changed from GAACCT to GAGCCC. To evaluate the benefit of these modifications, plasmids with changes at each individual site and at multiple sites were transiently expressed in CHO-K1 cells. Expression levels were compared in pools of transfectants developed using the unmodified and modified CAB-2 plasmids. ELISA results of culture supernatants from extinct pools revealed that those from transfectants generated using the modified vector contained, on average, 5-fold higher CAB-2 titers than those from transfectants generated with the unmodified vector (Fig. 5A). Stable clones subsequently were generated by transfection with the modified vector followed by limited dilution cloning. RNA isolated from individual stable clones and analyzed by 3' RACE (Fig. 5B and C) and RT-PCR (Fig. 5D) confirmed the elimination of the truncated RNA species.

**Discussion**

Sequential transfection of CHO-K1 cells with vectors expressing recombinant proteins typically yields an increase in expression that is proportional to gene copy number. However, in the case of CAB-2, expression increased only modestly following a sequential transfection. We conclude that this reduced expression was due primarily to a non-productive truncated RNA species arising from a polyadenylation signal element within the MCP coding sequence and secondarily to an alternatively spliced isoform arising from a cryptic splice donor site within the DAF coding sequence. This conclusion is based on the observation that (1) the prematurely polyadenylated mRNA comprised the majority of the aberrant mRNA species and (2) elimination of these aberrantly
processed RNAs resulted in significantly increased CAB-2 expression.

In a previous study of MCP, Milland et al.,[8] found that replacing AT-rich sequences with GC bases, including those within the same AATAAA sequence we describe, significantly increased MCP expression in both transiently and permanently transfected cells. They concluded that the reduced expression they observed was due to translational effects of AU rich mRNA sequences. We cannot rule out translational effects as an additional cause for our lower than expected CAB-2 expression since eliminating the internal polyadenylation and cryptic splice donor sequences yielded a 5-fold increase in expression which was greater than expected as the amount of full-length CAB-2 mRNA increased from ~40% to ~100% (~2.5-fold). However, since we did not significantly reduce the number of AT base pairs within the MCP region of CAB-2 (only 2 bp were substituted), it is unlikely that the observed expression improvements are due strictly to reduced AT content.

Of the three AATAAA sequences within the CAB-2 coding region, the one within MCP is preferentially used, accounting for ~50% of CAB-2 mRNA. This preference may be due to a GU – rich sequence (UAUUUGAGUAUYUU) located 32 bp downstream of this AATAAA that may act as a downstream regulatory element enhancing 3’ end formation [9–11]. No such U – rich or GU – rich sequences are located downstream of either of the two internal AATAAA sequences in DAF. In addition, the actual nucleotide sequences at the site of 3’ end cleavage can influence the efficiency of processing [12] and that could be the case for the preferential use of the poly A site within MCP versus the sites within DAF. Upstream enhancer sequences also are known to modulate/enhance 3’ end processing and have been found in viral genes and in some human genes and although a consensus has not emerged, they are often U-rich [9–13]. Interestingly, all three regions within CAB-2 which contain the AATAAA sequences also contain sequences resembling upstream enhancer elements. A GUUUUGUGU sequence is present 14 bp upstream of the AATAAA in MCP while UUUUGUAAAU and UUUUCAUU sequences lie 25 bp and 13 bp upstream of the AATAAA sites in DAF SCR4 and STP regions, respectively. The presence of GU-rich sequences both upstream and downstream of the poly site in MCP also might be responsible for its preferential usage.

Use of the internal AATAAA sites for polyadenylation does not appear to be unique to expression of the hybrid CAB-2 cDNA in CHO-K1 cells since prematurely polyadenylated RNAs were observed for MCP and DAF expressed in salivary gland and MCP expressed in PBL, placenta and the T-cell line, HSB2. While the use of alternative poly(A) signals is known to be a molecular mechanism of regulation in genes containing one or more sites in the
3' UTR [10,14], the occurrence of these sites within coding DNAs in human genes is rare [15]. Therefore, our demonstration of functional polyadenylation signals within the coding exons of CAB-2 and in human tissues and cell lines expressing the native proteins is unusual and their function (if any) is unknown.

Alternative splicing is known to be used for differential expression of genes in different tissues, during different phases of cell growth and for alternative cellular localization of proteins [16]. Although mRNA heterogeneity of native MCP and DAF is well documented with several isoforms resulting from alternative splicing [17–19], the biological significance of these structural variations is still under investigation. However, there is some evidence to suggest that the various isoforms differ in their complement inhibition profiles [18] as well as in their cellular localization [19]. While this propensity for heterogeneity suggested the possibility of multiple mRNA forms, the specific non-productive RNA species we observed, especially the abundant prematurely polyadenylated RNA, have not been noted before. In addition, the alternative splicing we observed with CAB-2 could not have been predicted as all splice variants described to date arise from alternative splicing downstream of the exons encoded in CAB-2. In addition, several splice site recognition programs were unable to identify the sites utilized in the CHO-K1 cell environment. It also should be noted that the expression vector used in this study in which the cryptic acceptor site was identified, has been used successfully for expression of numerous other recombinant proteins with no evidence of cryptic splicing.

In conclusion, this study demonstrates the importance of screening recombinant gene sequences not only for mRNA structure effects but also for potential internal RNA processing signals, such as premature polyadenylation signals and known canonical splice sequences, that could adversely affect recombinant protein expression. This type of screening is especially important when creating fusion genes in which DNA sequences not naturally found together are joined, bringing putative genetic elements promoting alternative processing into close proximity. By identifying and eliminating such internal RNA processing signals, recombinant protein expression can be maximized.

References


