Biochemical and Biophysical Research Communications 380 (2009) 508–513

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Non-AUG translational initiation of a short CAPC transcript generating protein isoform

Suresh Anaganti [a,1,2], Johanna K. Hansen [a,1], Duc Ha [a,3], Yoonsoo Hahn [a,4], Oleg Chertov [b], Ira Pastan [a], Tapan K. Bera [a,*]

[a] *Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, 37 Convent Drive, Room 5110, Bethesda, MD 20892-4264, USA*
[b] *Protein Chemistry Laboratory, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA*

ARTICLE INFO

Article history:
Received 13 January 2009
Available online 23 January 2009

Keywords:
LRRC26
Prostate cancer
Breast cancer
Alternate transcript
Non-AUG initiation

ABSTRACT

CAPC (also known as LRRC26) is a new gene with restricted expression in normal tissues, and with expression in many cancers and cancer cell lines. We have identified and characterized a short-transcript of *CAPC* (*S-CAPC*). The nucleotide sequence analysis of *CAPC* mRNA showed that the transcription for *S-CAPC* starts at position +610 on the *L-CAPC* transcript. Interestingly, no translation initiation codon 'AUG' is present in this transcript. To determine if a non-AUG start site is utilized, the *S-CAPC* sequence was cloned into an expression vector with C-terminal myc and histidine tags, and transfected into 293T cells. Western blot and MALDI-TOF MS analysis on purified *S-CAPC* gave two distinct peaks at approximately 7.5 kDa. N-terminal amino acid sequencing of the purified 7.5 kDa protein product indicated that translation starts at the codon for cysteine on the *S-CAPC* transcript generating a 7.5 kDa CAPC protein products translated from a non-AUG initiation site.

Breast and prostate cancers are the most common forms of cancer among women and men, respectively, in the United States. It is estimated that during their lifetime, 1 in 8 women will develop breast cancer, and 1 in 5 men will develop prostate cancer [1]. Despite these statistics there are currently no curative therapies for either of these cancers after they have metastasized from the point of origin.

To identify immunotherapeutic targets and diagnostic markers for these cancers, we had generated and analyzed a cDNA library from membrane-associated polyribosomal RNA derived from breast and prostate cancer cell lines [2,3]. From this library we identified and reported a gene termed *CAPC* [3]. The *CAPC* gene (also known as LLRC26) is located on the chromosome 9, and contains a 1005 bp translational open reading frame (ORF) with 2 exons. This ORF begins near the 5′-end of exon 1 and extends through approximately two-thirds of exon 2, encoding a 334 amino acid protein [3]. Analysis of *CAPC* expression by RT-PCR in cancers and normal tissues showed a high level of *CAPC* expression in breast, prostate, colon and pancreatic cancers and low or no expression in normal tissues with the exception of the prostate and salivary gland. The high *CAPC* expression in cancers suggests that the *CAPC* transcript could have a regulatory role in carcinogenesis. When we further characterized *CAPC* expression we observed that the *CAPC* gene-specific probe consistently hybridized with two transcripts, one having the expected length of ~1.2 kb and a second shorter one of ~600 bp. In this study we cloned and characterized the *short-CAPC* (*S-CAPC*) transcript from cancer cell lines. We found that long and short *CAPC* transcripts have differential expression in cancer cell lines, and that *S-CAPC* is translated into a 7.5 kDa protein using a non-AUG start codon.

Materials and methods

Cells and reagents. LNCaP, MCF7 and 293T cells were grown and maintained as recommended by the supplier (ATCC, Manassas, VA) in medium supplemented with 10% fetal bovine serum (FBS; Quality Biologicals) and 100 U penicillin/streptomycin.

RNA isolation and Northern blotting. Northern blot hybridization with 2 μg of mRNA was performed as described [4]. Oligo probes specific to the *CAPC* sense strand RNA (5′-CCG CCA GAC CCG AAC CCC GAT CCC GAC CCC CAC GGC TGT GCC TCG CCC GCG-3′) and to the antisense strand (5′-CGC GGG CGA GGC ACA GCC GTG GGG GTC GGG ATC GGG GTT CGG GTC TGG CGG-3′) were synthesized (Lofstrand Laboratories, Gaithersburg, MD). A 430-bp cDNA probe specific for *L-CAPC* is previously described [3]. Three

---

[*] Corresponding author. Fax: +1 301 402 1344.
E-mail address: tkbera@helix.nih.gov (T.K. Bera).
[1] These two authors contributed equally to this work.
[2] Present address: Molecular Carcinogenesis Cluster International Agency for Research on Cancer (WHO), 150 Cours Albert Thomas, 69372, Lyon Cedex, France.
[3] Present address: University of Colorado School of Medicine, 4200 E. Ninth Avenue, Denver, CO 80262, USA.
[4] Present address: Department of Life Science, College of Natural Science, Chung-Ang University, Seoul 156-756, Korea.

0006-291X/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.bbrc.2009.01.089

different probes with sequences hybridizing the 5′-end (Probe-I), the exon–exon junction region (Probe-II) and the 3′-end (Probe-III) of *CAPC* transcript were generated by PCR amplification of each fragment from the pcDNA-CAPC-myc plasmid [3] using primers 5′-ATG CGG GGC CCT TCC TGG TCG CGG-3′ and 5′-CGC ACA TGC ACC GAG TGC AGC CCG-3′ for Probe-I, 5′-TCG CGC CGG GGC TGC TGG GCC GCC TGC-3′ and 5′-TGC GCG CAA TGG CTA AAG GCG GC-3′ for Probe-II as well as 5′-GGG CCG GCC TCC TTC CTC GTC AGC-3′ and 5′-TCA GGC TTG GGC GGC AGC GGC GGG-3′ for Probe-III. All probes were labeled with ^{32}P by random primer extension (Lofstrand) to a specific activity of 1 μCi/ng (1 Ci = 37 GBq).

Rapid amplification of cDNA ends (RACE) PCR analysis and cloning. An adaptor ligated first-strand 5′-RACE-Ready cDNA was synthesized from 1 μg of LNCaP and MCF7 mRNA by using the SMART™ RACE cDNA amplification kit (CLONTECH, Mountain View, CA) according to the manufacturer's instructions. The cDNA was used as template for performing nested 5′-RACE PCR. An adaptor specific primer and the *CAPC* gene-specific primer-I (5′-CGC GGG CGA GGC ACA GCC GTG GGG GTC G-3′) complimentary to 3′-end of full-length *CAPC* cDNA were used for first set of 5′-RACE PCR. A second PCR was performed using the nested adaptor specific primer and the *CAPC* gene-specific primer-II (5′-CCC AGA GCC CAG CGC CAG GCA GGA AGC C-3′) and the product was cloned into the pCR®4-TOPO® vector (Invitrogen) and sequenced.

The cloning of the full-length *CAPC* transcript into CAPC-pcDNA-3-myc (pcDNA-3-myc) has been described [3]. The short *CAPC* transcript was amplified from CAPC-pcDNA-3-myc using primers 5′-CTC GAG CTC AGC CTG CAG GAC AAC GAG-3′ and 5′-AAG CTT GGC TTG GGC GGC AGC GGC-3′. The PCR fragment was cloned into XhoI/HindIII sites of pcDNA3.1A$^-$ (myc-His) mammalian expression vector (Invitrogen) fused in frame at the C-terminus with a myc epitope and a histidine tag.

Immunoblotting for S-CAPC. For Western blot experiments, 293T cells were transfected with plasmid carrying desired construct and 48 h after transfection cells were disrupted in lysis buffer, resolved by 18% SDS–PAGE and transferred onto poly (vinylidene difluoride) membrane. Blots were incubated sequentially with 5% BSA in Tris-buffered saline containing 0.05% Tween 20, and monoclonal anti-c-myc antibody (9E10; Santa Cruz Biotechnology Inc., Santa Cruz, CA) for 1 h at room temperature. The membrane was then treated with goat anti-mouse IgG-HRP secondary antibody (Santa Cruz Biotechnology Inc.) and the immunoreactive bands were visualized using the enhanced chemiluminescence method. Each membrane was stripped and reprobed with anti-actin antibody to normalize for differences in protein loading.

Immobilized metal affinity chromatography (IMAC). Cells were disrupted as described in Immunoblotting for S-CAPC. Supernatants were dialyzed against IMAC A buffer (50 mM $NaPO_4H_2$ (pH 7.5), 10 mM Imidazole, 500 mM NaCl) overnight at 4 °C and passed over a 0.2 μm filter. Short CAPC protein was purified using IMAC on Ni Sepharose™ High Performance resin (GE Healthcare, Piscataway, NJ) in a 2 ml column. Following binding in IMAC A buffer, proteins were eluted with a linear gradient of 0–100% IMAC B buffer (50 mM $NaPO_4H_2$ (pH 7.5), 350 mM Imidazole, 500 mM NaCl) over 40 min. Short CAPC containing fractions were determined by Western blot analysis. All positive fractions were pooled and concentrated in a 5K MWCO Amicon Ultra-15 Centrifugal Filter Unit (Millipore, Billerica, MA). Protein concentrations were estimated with a modified Bradford assay (Bio-Rad, Hercules, CA).

HPLC purification and N-terminal sequence analysis of S-CAPC. Two hundred milliliters of IMAC purified short CAPC protein was reduced by adding 10 ml of *b*-mercaptoethanol by incubating for 10 min at 56 °C. The solution was acidified by adding 5 ml of trifluoroacetic acid (TFA) and injected onto a C18 column (2.1 × 250 mm; VYDAC®, Deerfield, IL) connected to a Waters HPLC system consisting of Waters 600S Controller, 626 Pump and 996 Photodiode Array UV Detector. The column was equilibrated in 98% of buffer A (0.1% TFA in water) and 2% buffer B (0.1% TFA in acetonitrile). Proteins were eluted with a linear gradient of buffer B (from 2% to 90% over 60 min) at a flow rate of 0.2 ml/min. The column eluant was monitored at 206 nm. Thirty-six peak fractions (~0.2–0.3 ml each) were collected. From each fraction a 20 ml aliquot was taken, mixed with 2 ml of a BSA solution (1 mg/ml in water) and lyophilized. The samples were separated on 18% PAGE and immunoblotted with anti-Myc to detect short CAPC expression. Fifty percent (100 ml) of the fraction containing S-CAPC was analyzed using an automated protein sequencer Procise 494 cLC (Applied Biosystems, Foster City, CA).

MALDI-TOF MS. One milliliter aliquot of fraction containing S-CAPC was mixed with 0.7 ml of matrix (20 mg/ml of sinapinic acid in 50% acetonitrile/water, 0.05% TFA) directly on the target plate and analyzed using an Applied Biosystems Voyager-DE Pro MALDI-TOF MS. The accelerating voltage was 25 kV, guide wire 0.2% and grid voltage 92%. The instrument was operated in linear mode under positive ion conditions. A nitrogen laser was used at 337 nm with 150 laser shots averaged per spectrum. Calibration was performed using instrument default settings (mass accuracy 0.1%) and data analysis was carried out using Data Explorer software resident on the instrument.

Results

Identification and cloning of S-CAPC transcript from cancer cell line

We reported earlier that the *CAPC* gene is expressed in prostate as a major transcript of 1.2 kb in size [3]. In addition, a weak band representing a smaller transcript of 0.6 kb in size was detected in prostate tissue. We performed Northern blot analysis to investigate and verify the expression pattern of *CAPC* transcripts in normal tissue and cancer cell lines. A membrane with 2 μg of mRNA derived from three different normal tissues (prostate, brain and salivary glad) and two different cancer cell lines (MCF7 and LnCaP) were probed with a radiolabeled *CAPC* fragment. As shown in Fig. 1A, the major *CAPC* transcript in prostate and in salivary gland is 1.3 kb in size. There is a weak but specific band of 0.6 kb in size present in both lanes indicating that these transcripts share some sequence similarity. The 1.3 and 0.6 kb transcripts are also expressed in LnCap and MCF cell lines. The intensity of the 1.3 kb transcript decreases in both cancer cell lines, while the intensity of the 0.6 kb transcript is stronger in the LnCap cell line and equivalent in the MCF7 cell line when compared to normal tissues. There is no detectable band in normal brain RNA indicating the *CAPC* signal is specific (Fig. 1A).

To determine if the two *CAPC* transcripts share similar exons, we designed three distinct probes spanning the full-length *CAPC* sequence, designated as Probe-I, Probe-II and Probe-III (Fig. 1C). A Northern blot analysis was performed on a membrane containing 2 μg of mRNA derived from prostate, brain and salivary gland tissue, as well as cancer cell lines, MCF7 and LnCaP. The results are shown in Fig. 1D–F. As shown in Fig. 1D, Probe-I, specific to the 5′-end of *CAPC*, failed to detect the short-transcript but detected the *L-CAPC* transcript. Probe-II, specific for the exon–exon junction region, and Probe-III, specific for the transcript 3′-end, recognized both *CAPC* transcripts (Fig. 1E and F). These results indicate that the two transcripts have sequence similarity at the 3′ region and exon–exon junction region, but differ in their sequence at the 5′ region.

To investigate the possibility that the *S-CAPC* transcript might be an antisense RNA to the *CAPC* gene, we performed Northern blot analysis with sense and antisense oligo probes designed from the 3′-end of the *CAPC* transcript. As shown in Fig. 1B, both the *S-CAPC*

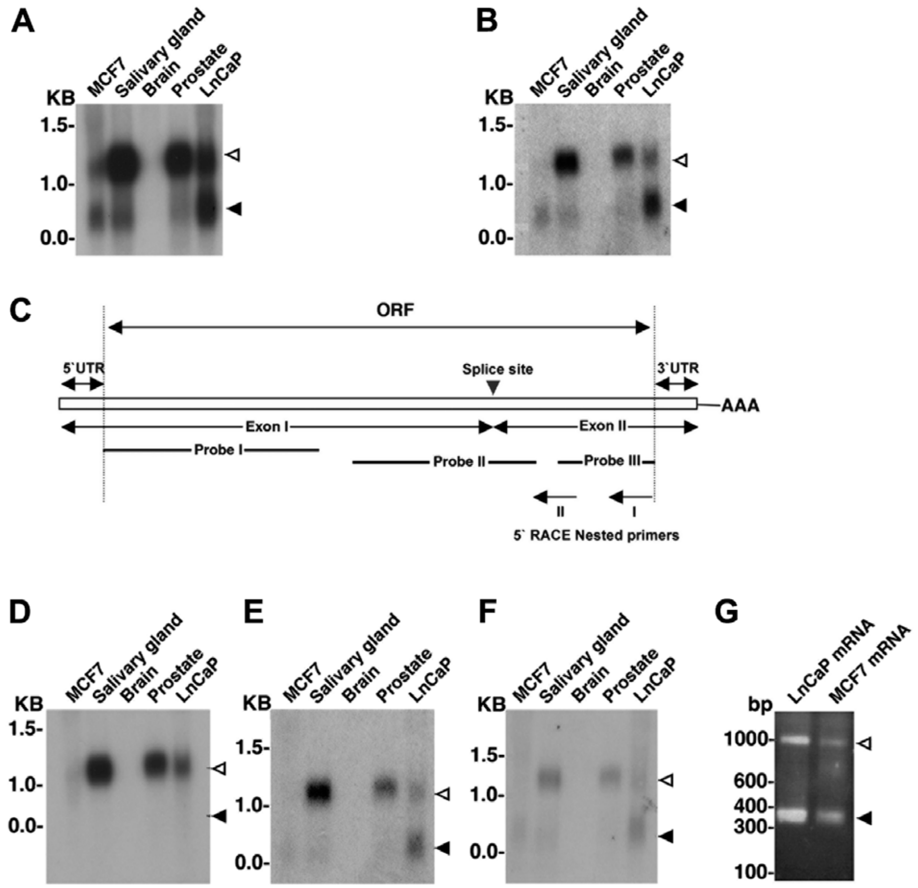


Fig. 1. Northern blot analysis of CAPC transcripts and mapping of the S-CAPC transcript 5′-end. (A) Each lane contains 2 mg of mRNA from: 1, MCF7; 2, salivary gland; 3, brain; 4, prostate; and 5, LnCaP. The membrane was probed with a 430-bp ^{32}P-labeled CAPC fragment. The migration of molecular weight marker is indicated on the left of the first lane. The ~1.3 kb transcript corresponds to the predicted L-CAPC cDNA (white arrow) and the ~0.6 kb transcript corresponds to the shorter isoform of CAPC (black arrow). (B) The blot containing 2 mg of mRNA from: 1, MCF7; 2, salivary gland; 3, brain; 4, prostate; and 5, LnCaP was probed with an antisense oligo probe with sequence corresponding to 3′-end of CAPC. (C) The CAPC gene structure is schematized, showing UTR regions and two exons including the intro-exon splice site. The position of three different probes used to perform Northern blot are indicated as well as the two primers used to perform 5′-nested RACE PCR. (D) The blot containing 2 mg of mRNA from: 1, MCF7; 2, salivary gland; 3, brain; 4, prostate; and 5, LnCaP was probed with Probe-I. The same blot was probed with (E) Probe-II and (F) Probe-III. (G) 5′-Nested RACE PCR cDNA products of the 5′ terminal sequences of CAPC transcripts. The relative migrating positions of the cDNA products on the 1.5% agarose gel are indicated. The migration of molecular weight markers is shown on the left.

and *L-CAPC* transcripts are detected in the blot probed with antisense oligo probe (Fig. 1B). There was no signal detected on the blot probed with sense oligo (data not shown) indicating that both transcripts are sense strands.

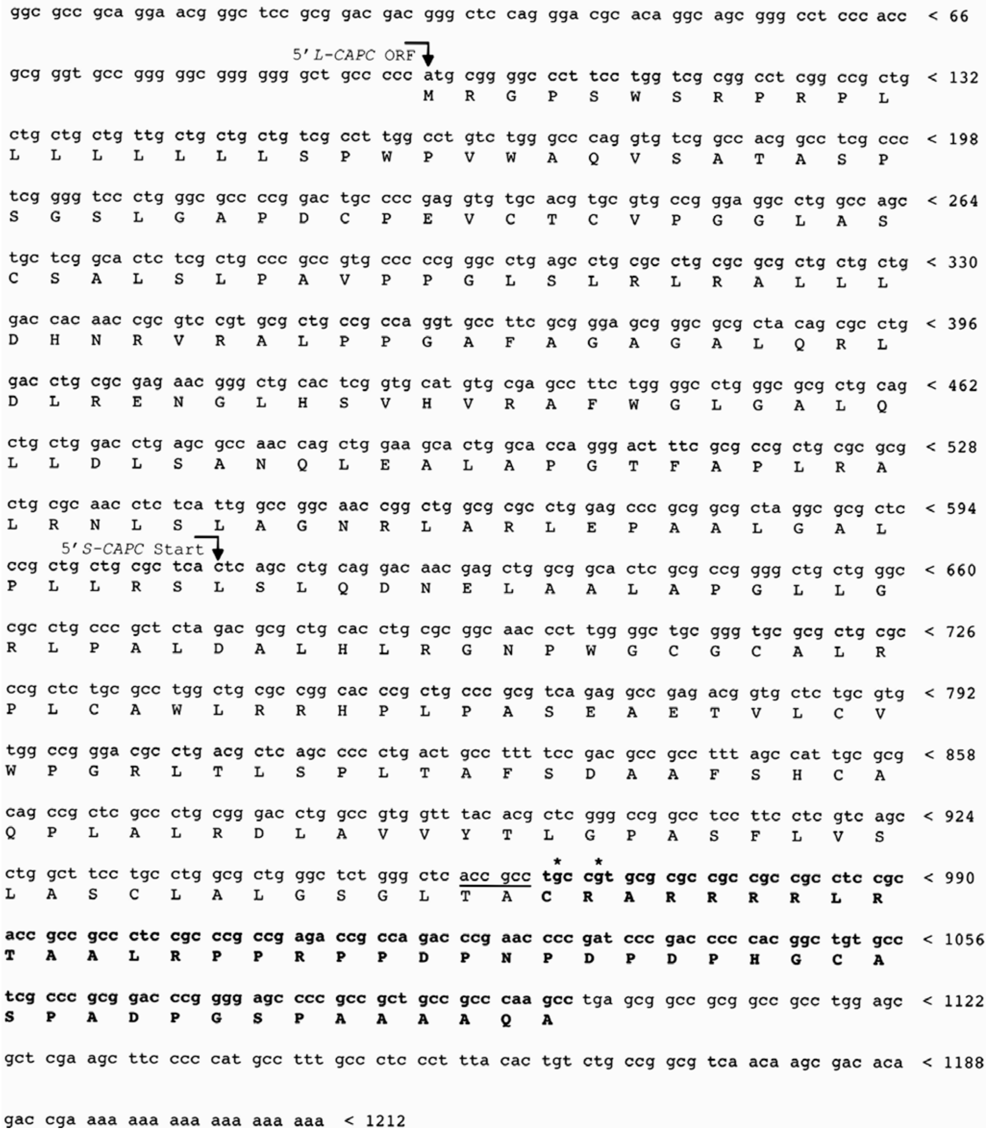
We then determined the 5′ sequence of the *S-CAPC* transcript using 5′-RACE with primers I and II, which are designed to bind the 3′-end of *L-CAPC* (Fig. 1C). 5′-RACE was performed on MCF7 and LnCaP mRNA and two distinct bands corresponding to the *CAPC* isoforms were observed (Fig. 1G). The shorter band (~340 bp) was eluted from the gel and cloned into a pCR®4-TOPO® vector and several independent clones were sequenced. There was complete sequence alignment between *S-CAPC* and *L-CAPC* transcripts with the *S-CAPC* starting at base +513 on the *L-CAPC* ORF (Fig. 2). The sequence has been submitted into the GenBank with an Accession No. EU588721.

Western blot analysis of S-CAPC expression

The *S-CAPC* transcript lacks any translational initiation codon (AUG) in all three reading frames. To determine if *S-CAPC* encodes a protein initiating from a non-AUG codon, we cloned the 492 bp *S-CAPC* transcript into a mammalian expression vector with an in-frame C-terminal myc and histidine tags for Western blot analysis and protein purification, respectively (Fig. 3A). We also made a *S-CAPC* construct containing a stop codon (TGA) between the *CAPC* coding region and C-terminal myc and his tag coding regions (Fig. 3A). This construct served as an indirect negative control, confirming the translational initiation reading frame by checking the expression of the myc epitope. We examined expression of these constructs in 293T cells by transient transfection. Supernatant proteins were resolved by 18% SDS–PAGE and analyzed for *CAPC* protein products by Western blot using an anti-myc antibody (Fig. 3B). As shown in Fig. 3B, we detected a 15 kDa protein from cells transfected with *S-CAPC* without a STOP codon between *CAPC* and myc and a 37 kDa protein from cells transfected with *L-CAPC*-myc. As expected, we did not detect either CAPC proteins in supernatants from 293T cells transfected with vector alone or with construct II containing the STOP codon (Fig. 3B). Blots were stripped and reprobed for b-actin as a loading control (Fig. 3B, bottom panel).

Purification of S-CAPC with IMAC and HPLC

Because the *S-CAPC* transcript did not contain an AUG start site, we purified the expressed protein to determine the first amino acid. S-CAPC was expressed with a C-terminal histidine tag for IMAC purification. We had previously shown with Western blot analysis that S-CAPC was soluble and could be detected in the cell supernatants after SDS–PAGE separation. Forty plates of 293T cells



```
ggc gcc gca gga acg ggc tcc gcg gac gac ggg ctc cag gga cgc aca ggc agc ggg cct ccc acc  < 66

                         5' L-CAPC ORF ↓
gcg ggt gcc ggg ggc ggg ggg gct gcc ccc atg cgg ggc cct tcc tgg tcg cgg cct cgg ccg ctg  < 132
                                        M   R   G   P   S   W   S   R   P   R   P   L

ctg ctg ctg ttg ctg ctg ctg tcg cct tgg cct gtc tgg gcc cag gtg tcg gcc acg gcc tcg ccc  < 198
L   L   L   L   L   L   L   S   P   W   P   V   W   A   Q   V   S   A   T   A   S   P

tcg ggg tcc ctg ggc gcc ccg gac tgc ccc gag gtg tgc acg tgc gtg ccg gga ggc ctg gcc agc  < 264
S   G   S   L   G   A   P   D   C   P   E   V   C   T   C   V   P   G   G   L   A   S

tgc tcg gca ctc tcg ctg ccc gcc gtg ccc ccg ggc ctg agc ctg cgc ctg cgc gcg ctg ctg ctg  < 330
C   S   A   L   S   L   P   A   V   P   P   G   L   S   L   R   L   R   A   L   L   L

gac cac aac cgc gtc cgt gcg ctg ccg cca ggt gcc ttc gcg gga gcg ggc gcg cta cag cgc ctg  < 396
D   H   N   R   V   R   A   L   P   P   G   A   F   A   G   A   G   A   L   Q   R   L

gac ctg cgc gag aac ggg ctg cac tcg gtg cat gtg cga gcc ttc tgg ggc ctg ggc gcg ctg cag  < 462
D   L   R   E   N   G   L   H   S   V   H   V   R   A   F   W   G   L   G   A   L   Q

ctg ctg gac ctg agc gcc aac cag ctg gaa gca ctg gca cca ggg act ttc gcg ccg ctg cgc gcg  < 528
L   L   D   L   S   A   N   Q   L   E   A   L   A   P   G   T   F   A   P   L   R   A

ctg cgc aac ctc tca ttg gcc ggc aac cgg ctg gcg cgc ctg gag ccc gcg gcg cta ggc gcg ctc  < 594
L   R   N   L   S   L   A   G   N   R   L   A   R   L   E   P   A   A   L   G   A   L
   5' S-CAPC Start ↓
ccg ctg ctg cgc tca ctc agc ctg cag gac aac gag ctg gcg gca ctc gcg ccg ggg ctg ctg ggc  < 660
P   L   L   R   S   L   S   L   Q   D   N   E   L   A   A   L   A   P   G   L   L   G

cgc ctg ccc gct cta gac gcg ctg cac ctg cgc ggc aac cct tgg ggc tgc ggg tgc gcg ctg cgc  < 726
R   L   P   A   L   D   A   L   H   L   R   G   N   P   W   G   C   G   C   A   L   R

ccg ctc tgc gcc tgg ctg cgc cgg cac ccg ctg ccc gcg tca gag gcc gag acg gtg ctc tgc gtg  < 792
P   L   C   A   W   L   R   R   H   P   L   P   A   S   E   A   E   T   V   L   C   V

tgg ccg gga cgc ctg acg ctc agc ccc ctg act gcc ttt tcc gac gcc gcc ttt agc cat tgc gcg  < 858
W   P   G   R   L   T   L   S   P   L   T   A   F   S   D   A   A   F   S   H   C   A

cag ccg ctc gcc ctg cgg gac ctg gcc gtg gtt tac acg ctc ggg ccg gcc tcc ttc ctc gtc agc  < 924
Q   P   L   A   L   R   D   L   A   V   V   Y   T   L   G   P   A   S   F   L   V   S
                                                     *   *
ctg gct tcc tgc ctg gcg ctg ggc tct ggg ctc acc gcc tgc cgt gcg cgc cgc cgc cgc ctc cgc  < 990
L   A   S   C   L   A   L   G   S   G   L   T   A   C   R   A   R   R   R   R   L   R

acc gcc gcc ctc cgc ccg ccg aga ccg cca gac ccg aac ccc gat ccc gac ccc cac ggc tgt gcc  < 1056
T   A   A   L   R   P   P   R   P   P   D   P   N   P   D   P   D   P   H   G   C   A

tcg ccc gcg gac ccg ggg agc ccc gcc gct gcc gcc caa gcc tga gcg gcc gcg gcc gcc tgg agc  < 1122
S   P   A   D   P   G   S   P   A   A   A   A   Q   A

gct cga agc ttc ccc cat gcc ttt gcc ctc cct tta cac tgt ctg ccg gcg tca aca agc gac aca  < 1188


gac cga aaa aaa aaa aaa aaa aaa  < 1212
```

Fig. 2. Sequence alignment between L-CAPC and S-CAPC transcripts. The L-CAPC transcript base number is shown on the right. The S-CAPC transcript starts at codon +610 on the L-CAPC transcript, which corresponds to codon +513 on the L-CAPC ORF. The determined nucleotide and protein sequence of S-CAPC are in bold font and possible initiation sites (∗) as well as the putative Kozak sequence (ACC GCC; underlined) are indicated.

were transfected and harvested, and S-CAPC was purified from lysates by IMAC. The eluted fractions were pooled and concentrated before S-CAPC was purified by reverse-phase chromatography. After both IMAC and HPLC purification of S-CAPC, we performed a Western blot with anti-myc on all collected fractions to verify the presence of the protein. Fig. 4A shows the reverse-phase chromatogram. S-CAPC was predominantly in fraction 16, which was used for N-terminal sequencing and mass spectrometry analysis (Fig. 4A and B).

N-terminal protein sequencing and MALDI-TOF MS analysis of S-CAPC start site

The N-terminal sequence of the protein in fraction 16 is XRARRRRLRTAALRP. Since the first residue could not be determined, it was unclear if initiation was at the cysteine that corresponds to "X" in the CAPC sequence (Fig. 2). MALDI-TOF MS analysis of fraction 16 identified major peaks at m/z 7622 and m/z 7779. The peak at m/z 7622 (± 8 mass units with mass accuracy of 0.1%) was in good agreement to the CAPC predicted mass of 7598.41 kDa calculated from the protein sequence: CRARRRRLRT AALRPPRPPD PNPDPDPHGC ASPADPGSPA AAAQAKLGPE QKLISEEDLN SAVDHHHHHH (Fig. 4C). Carboxyamidomethylation of fraction 16 with iodoacetamide (data not shown) indicated the presence of only one free sulfhydryl group while there are two cysteine residues present according to the sequence. This result suggests that the difference between experimental and calculated CAPC masses may be attributed to post-translational modification of one of the cysteine residues. The second peak at m/z 7779 has a mass difference of approximately 157 mass units from the m/z 7622 peak and may correspond to a different form of CAPC with an additional post-translational modifications (Fig. 4C). This second polypeptide also has only one free sulfhydryl group, which can be modified by iodoacetamide. We conclude from these data that the N-terminus of S-CAPC is a cysteine residue, which may be post-translationally modified, and that the S-CAPC protein is close to 7.5 kDa in size and not 15 kDa as observed by SDS–PAGE.

Discussion

In this study we have identified and characterized a new, shorter form of the *CAPC* transcript, termed *S-CAPC*. We initially identi-

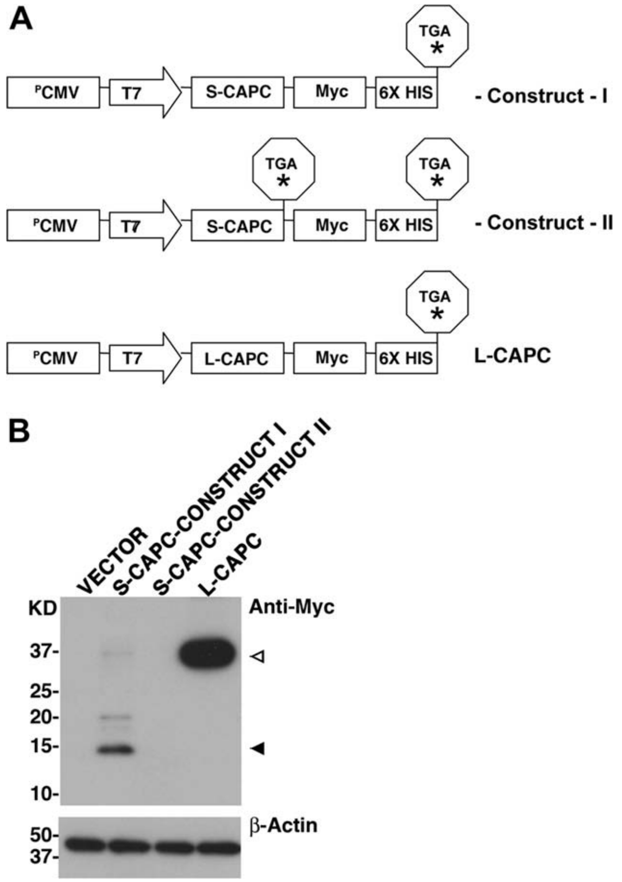


Fig. 3. S-CAPC constructs and protein expression. (A) Schematic representation of three different constructs. Myc and 6× histidine tags are shown. The stop codons 'TGA' are represented with (∗). (B) Western blot analysis of S-CAPC expression: 50 mg of lysate from Vector (Lane 1), S-CAPC-Construct I (black arrow; Lane 2), S-CAPC-Construct II (Lane 3), and L-CAPC (white arrow; Lane 4) was resolved on 18% SDS–PAGE. The blot was probed with anti-actin as a loading control.

fied the *S-CAPC* transcript in cancer cell lines while probing for the *L-CAPC* transcript. Northern blot analysis using three probes spanning the full-length *CAPC* ORF showed that *S-* and *L-CAPC* shared homology over the exon–exon junction region and at the 3′ region, but differed at their 5′ regions (Fig. 1C–F). In addition, sequence analysis by 5′-RACE showed that *S-CAPC* has 100% sequence identity with *L-CAPC*, and that the transcript starts at position +513 on the *L-CAPC* ORF (Figs. 1G and 2). These results show that *S-CAPC* does not have homology to the 5′ region of *L-CAPC* and therefore indicate that *S-CAPC* is not an alternative splicing product. Instead *S-CAPC* is most likely a new transcript that initiates from a different promoter.

After determining the *S-CAPC* transcript sequence, we transfected cDNAs corresponding to the putative *S-CAPC* ORF into mammalian cells to examine protein expression. Western blot analysis of S-CAPC indicated that the protein was 15 kDa in size (Fig. 3B), which is smaller than expected according to the length of the transcript. In addition, sequence analysis showed that the *S-CAPC* transcript has a non-standard start codon (Fig. 2), which impeded identification of the translation initiation site. To identify the start site, we determined both the N-terminal sequence and the molecular weight of purified S-CAPC. The N-terminal amino acid sequence analysis suggested that S-CAPC translation was initiated at the cysteine codon, UGC, and MALDI-TOF analysis also confirmed that S-CAPC is 7.5 kDa. The N-terminal sequencing result was unexpected as it has only been shown with mutant tRNAs that initiation occurs with a non-methionine residue at a non-AUG codon [5,6]. For the most part studies have shown that initiation can occur at a non-cognate codon as the result of methionine tRNA base pairing to a codon complimentary at only two bases. Such identified codons include CUG, ACG and UUG [6,7], and more recently GUG in the fungi, *Mucor circinelloides* [8]. From the amino acid sequencing data we cannot rule out the possibility that X in our N-terminal sequences is a methionine and not a cysteine residue. Iodoacetamide treatment of CAPC confirmed the presence of only one sulfhydryl group in S-CAPC, further supporting this notion. However, in the case of S-CAPC, the UGC codon is mismatched to AUG at all three bases and therefore would be a unique non-cognate codon for methionine initiation. Instead, we cannot rule out the possibility that the initiation starts at some upstream codon and the S-CAPC protein product is proteolytically cleaved to generate the XRARRRRLRTAALRP poly-peptide that we identified by mass spectrometry. Future work will be needed to determine if X is indeed a modified cysteine or if UGC is a noncognate codon for methionine.

It has been shown previously that non-AUG initiation of protein synthesis is associated with the expression of regulatory proteins

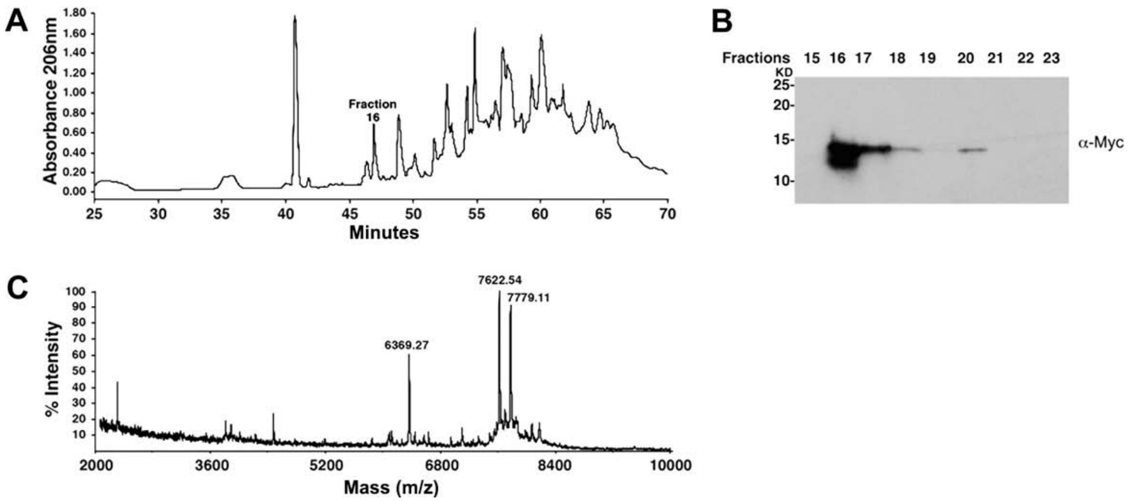


Fig. 4. Purification and determination of S-CAPC molecular weight and translation initiation sites. (A) IMAC-purified CAPC was loaded onto a C18 column (2.1 × 250 mm; Vydac) and eluted with an acetonitrile gradient at 0.2 ml/min, collected fractions (numbered). (B) Western blot analysis of eluted fractions identified CAPC in Fraction 16 (∗). (C) MALDI-TOF MS analysis of Fraction 16 identified two major peaks at *m*/*z* 7622 and *m*/*z* 7779, which corresponds to the mass of CAPC (7598.41) calculated from the cDNA sequence.

[6]. A separate transcript from the CAPC locus and a non-AUG start codon in *S-CAPC* suggests expression of this protein under unique circumstances. Future work will be necessary to better define the initiation site and elucidate the role this protein plays in cancer.

In conclusion, we cloned and characterized the *S-CAPC* transcript from cancer cell lines. The *S-CAPC* transcript is differentially expressed in cancer cell lines form the *L-CAPC* transcript and encodes a protein of 7.5 kDa in size that is translated from a non-AUG translational initiation codon.

Acknowledgments

This research was supported in part by the Intramural Research Program of the NIH, NCI, CCR and with Federal funds from the NCI, NIH, under contract NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

We thank Susan Garfield and Poonam Mannan for technical support; and NIH Fellows Editorial Board for valuable comments.

References

[1] A. Jemal, T. Murray, A. Samuels, A. Ghafoor, E. Ward, M.J. Thun, Cancer statistics, CA Cancer J. Clin. 53 (2003) 5–26.

[2] K.A. Egland, J.J. Vincent, R. Strausberg, B. Lee, I. Pastan, Discovery of the breast cancer gene BASE using a molecular approach to enrich for genes encoding membrane and secreted proteins, Proc. Natl. Acad. Sci. USA 100 (2003) 1099–1104.

[3] K.A. Egland, X.F. Liu, S. Squires, S. Nagata, Y.G. Man, T.K. Bera, M. Onda, J.J. Vincent, R.L. Strausberg, B. Lee, I. Pastan, High expression of a cytokeratin-associated protein in many cancers, Proc. Natl. Acad. Sci. USA 103 (2006) 5929–5934.

[4] M. Essand, G. Vasmatzis, U. Brinkmann, P. Duray, B. Lee, I. Pastan, High expression of a specific T-cell receptor gamma transcript in epithelial cells of the prostate, Proc. Natl. Acad. Sci. USA 96 (1999) 9287–9292.

[5] H.J. Drabkin, U.L. RajBhandary, Initiation of protein synthesis in mammalian cells with codons other than AUG and amino acids other than methionine, Mol. Cell. Biol. 18 (1988) 5140–5147.

[6] C. Touriol, S. Bornes, S. Bonnal, S. Audigier, H. Prats, A.C. Prats, S. Vagner, Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons, Biol. Cell. 95 (2003) 169–178.

[7] D.S. Peabody, Translation initiation at non-AUG triplets in mammalian cells, J. Biol. Chem. 264 (1989) 5031–5035.

[8] L. Murcia-Flores, J.M. Lorca-Pascual, V. Garre, S. Torres-Martinez, R.M. Ruiz-Vazquez, Non-AUG translation initiation of a fungal RING finger repressor involved in photocarotenogenesis, J. Biol. Chem. 282 (2007) 15394–15403.