npg

ORIGINAL ARTICLE

RNA secondary structures located in the interchromosomal region of human ACAT1 chimeric mRNA are required to produce the 56-kDa isoform

Jia Chen[1,*], Xiao-Nan Zhao[1,*], Li Yang[1], Guang-Jing Hu[1], Ming Lu[1], Ying Xiong[1], Xin-Ying Yang[1], Catherine CY Chang[2], Bao-Liang Song[1], Ta-Yuan Chang[2], Bo-Liang Li[1]

[1]State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China; [2]Department of Biochemistry, Dartmouth Medical School, Hanover, NH 03756, USA

We have previously reported that the human ACAT1 gene produces a chimeric mRNA through the interchromosomal processing of two discontinuous RNAs transcribed from chromosomes 1 and 7. The chimeric mRNA uses $AUG_{1397-1399}$ and $GGC_{1274-1276}$ as translation initiation codons to produce normal 50-kDa ACAT1 and a novel enzymatically active 56-kDa isoform, respectively, with the latter being authentically present in human cells, including human monocyte-derived macrophages. In this work, we report that RNA secondary structures located in the vicinity of the $GGC_{1274-1276}$ codon are required for production of the 56-kDa isoform. The effects of the three predicted stem-loops (nt 1255-1268, 1286-1342 and 1355-1384) were tested individually by transfecting expression plasmids into cells that contained the wild-type, deleted or mutant stem-loop sequences linked to a partial ACAT1 AUG open reading frame (ORF) or to the ORFs of other genes. The expression patterns were monitored by western blot analyses. We found that the upstream stem-loop$_{1255-1268}$ from chromosome 7 and downstream stem-loop$_{1286-1342}$ from chromosome 1 were needed for production of the 56-kDa isoform, whereas the last stem-loop$_{1355-1384}$ from chromosome 1 was dispensable. The results of experiments using both monocistronic and bicistronic vectors with a stable hairpin showed that translation initiation from the $GGC_{1274-1276}$ codon was mediated by an internal ribosome entry site (IRES). Further experiments revealed that translation initiation from the $GGC_{1274-1276}$ codon requires the upstream AU-constituted RNA secondary structure and the downstream GC-rich structure. This mechanistic work provides further support for the biological significance of the chimeric nature of the human ACAT1 transcript.

Keywords: human ACAT1 isoform, chimeric human ACAT1 mRNA, interchromosomal region, RNA secondary structure, internal ribosome entry site

Introduction

Acyl-coenzyme A:cholesterol acyltransferase (ACAT) is a key enzyme in cellular cholesterol metabolism. It catalyzes the formation of cholesteryl ester from cholesterol and long-chain fatty acyl-coenzyme A in various cell types [1], and therefore controls the dynamic interconversion between cellular free cholesterol and cholesterol ester. Two ACAT

genes (encoding ACAT1 and ACAT2) have been identified: ACAT1, an allosteric tetrameric enzyme [2], is expressed ubiquitously in all human tissues examined, whereas ACAT2 is mainly expressed in the embryonic liver and in the intestine [3-5]. Under pathological conditions, ACAT1 plays a central role in the accumulation of cholesterol ester in macrophages during the early stages of atherosclerotic disease [6-8]. Surprisingly, the genomic DNA that encodes the human ACAT1 cDNA K1 [9] is located on two different chromosomes (1 and 7): exons 1-16 are located on chromosome 1, while the optional long exon Xa (1279 bp) is located on chromosome 7 [10]. Northern blot analyses revealed the presence of four ACAT1 transcripts (7.0, 4.3, 3.6 and 2.8 kb) in almost all of the human tissues and cells examined. These mRNAs share the sequence located on

chromosome 1, but only the 4.3-kb mRNA contains the long exon Xa, which is located on chromosome 7. Thus, this transcript is produced from two different chromosomes by a novel RNA recombination event that presumably involves interchromosomal trans-splicing [10]. The human ACAT1 cDNA K1 is derived from this mRNA.

RNA secondary structures are involved in the regulation of protein expression in multiple ways. For example, in the cap-dependent ribosome scanning model, RNA secondary structures located at the 5′-untranslated region (5′-UTR) of mRNA inhibit translation, while those downstream of the initiation codon can increase translational efficiency [11]. On the other hand, in the case of cap-independent internal ribosome entry, RNA secondary structures located in the long 5′-UTR can facilitate the internal translation initiation by recruiting various translation initiation factors, especially when the translation is initiated from non-AUG codons [12, 13]. Additionally, at the RNA splicing level, it has been reported that RNA secondary structures can recruit several positive (B52, SRp55, and NOVA-1) and negative (hnRNP A1) regulatory protein factors [14, 15]. We had previously shown that, for the ACAT1 mRNAs without the optional exon Xa, an upstream stem-loop structure can enhance the selection of the adjacent downstream AUG to produce the 50-kDa human ACAT1 protein [16]. For the chimeric 4.3-kb ACAT1 mRNA, we have noted that various predicted RNA secondary structures exist within the optional long 5′-UTR, but whether they are related to the production of the 56-kDa isoform has not been tested.

Previously, the chimeric 4.3-kb ACAT1 mRNA was expressed in mutant Chinese hamster ovary (CHO) cells lacking endogenous ACAT1. It was shown that this mRNA produces a 50-kDa ACAT1 protein from the initiation codon $AUG_{1397-1399}$, as well as a novel ACAT1 isoform with an apparent molecular weight of 56 kDa [17]. The 56-kDa isoform is enzymatically active in the absence of the 50-kDa isoform, although its normalized activity is only 30% of the activity of the smaller form [17]. When the 56- and 50-kDa ACAT1 isoforms were co-expressed in the same cell, the normalized ACAT1 activity was about 50% of the activity of the 50-kDa ACAT1 protein alone [17]. Deletion analysis demonstrated that the partial sequences of the human ACAT1 gene present on chromosomes 7 and 1 are both required to produce the 56-kDa isoform. We also showed that the partial sequence (nt 1243-1786) of human ACAT1 mRNA is sufficient to produce a 25-kDa N-terminal truncation of the 56-kDa ACAT1 isoform. Further mutation and mass spectrometry analysis showed that $GGC_{1274-1276}$ was used as the initiation codon for production of the 56-kDa ACAT1 isoform [17]. Importantly, the native 56-kDa ACAT1 isoform can be endogenously expressed in all examined human THP-1 macrophages

and human monocyte-derived macrophages [17]. Furthermore, limited proteolysis, by the V8 protease, of the native 56-kDa isoform from macrophages and the recombinant version from transfected cells confirmed that the native 56-kDa ACAT1 isoform is the same as that expressed in transfected cells [17]. Given that the optional long 5′-UTR in the chimeric human ACAT1 mRNA will prevent scanning of the 40S ribosome complex and that translation of the native 56-kDa isoform is initiated from a non-AUG codon, we hypothesize that the translation initiation from the $GGC_{1274-1276}$ codon may be mediated by some special RNA secondary structures.

In the current work, we analyzed whether the three predicted stem-loops (nt 1255-1268, 1286-1342 and 1355-1384) located either upstream or downstream of the $GGC_{1274-1276}$ codon could affect the production of proteins initiated from this codon. Our results revealed that both the upstream stem-loop$_{1255-1268}$ and downstream stem-loop$_{1286-1342}$ from two different chromosomes are required for production of the 56-kDa human ACAT1 isoform.
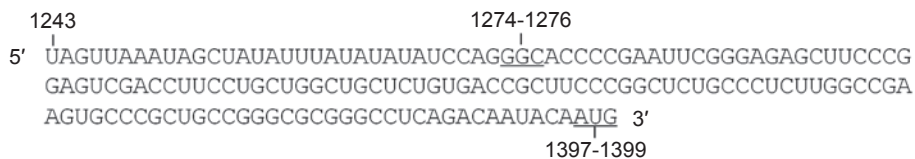
Results

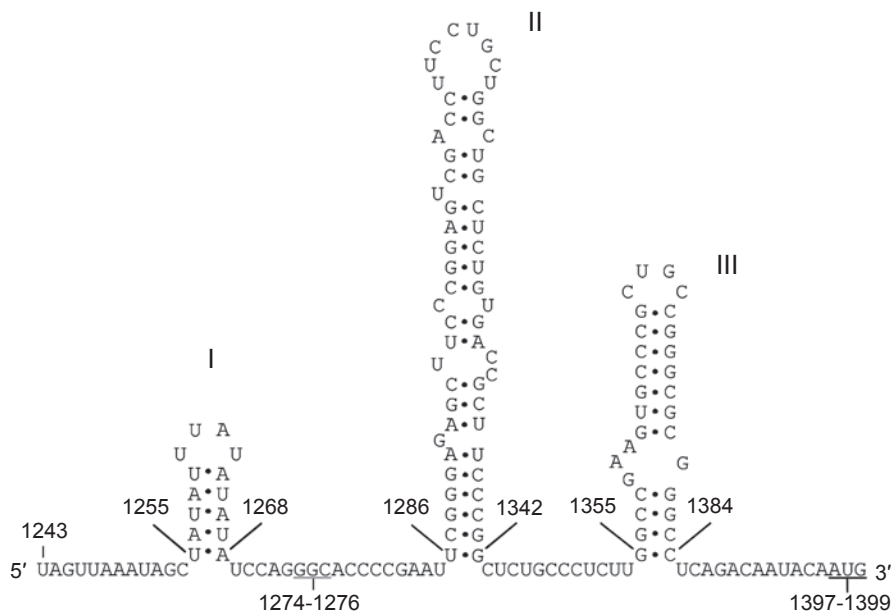*The predicted stem-loops from two different chromosomes are required for the production of protein initiated from the $GGC_{1274-1276}$ codon*

Our previous study indicated that a native 56-kDa isoform can be produced from the chimeric human ACAT1 mRNA by using $GGC_{1274-1276}$ as the initiation codon. The minimal region required for this non-AUG codon-initiated translation has been narrowed down to nt 1243-1786 [17]. The in-frame codon $GGC_{1274-1276}$ is located within the sequence (nt 1243-1396) that is upstream of the AUG-ORF (open reading frame) used to produce the normal 50-kDa ACAT1 isoform (Figure 1A). In the regions upstream and downstream of the $GGC_{1274-1276}$ codon, there are three predicted stem-loops. Stem-loop$_{1255-1268}$ (stem-loop I) is transcribed from chromosome 7, while most of stem-loop$_{1286-1342}$ (stem-loop II) and the entirety of stem-loop$_{1355-1384}$ (stem-loop III) are transcribed from chromosome 1 (Figure 1B). Stem-loop I is composed of only A and U nucleotides. Stem-loop II, downstream of the $GGC_{1274-1276}$ codon, is more complex than the other two, and contains a GC-rich stem with multiple bulges. Finally, stem-loop III, which is downstream of stem-loop II and adjacent to the AUG-ORF, also has a high GC content, and contains only one bulge in the stem. Previous work showed that stem-loop III could modulate the selection of a downstream AUG for translation initiation to produce the normal 50-kDa human ACAT1 protein [16].

In order to test the effect of predicted interchromosomal stem-loops on the production of protein initiated from the

A

1243   1274-1276
| |
5′ UAGUUAAAUAGCUAUAUUUAUAUAUAUCCAGGGCACCCCGAAUUCGGGAGAGCUUCCCG
GAGUCGACCUUCCUGCUGGCUGCUCUGUGACCGCUUCCCGGCUCUGCCCCUCUUGGCCGA
AGUGCCCGCUGCCGGGCGCGGGCCUCAGACAAUACAAUG 3′
                                      |
                                 1397-1399

B

C

D

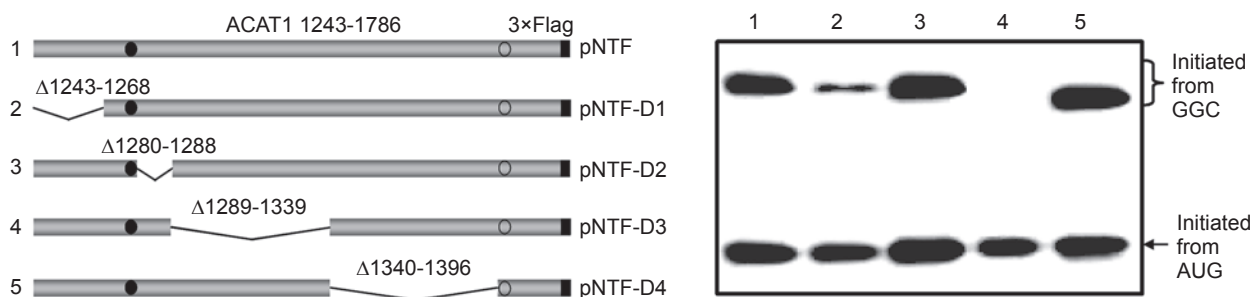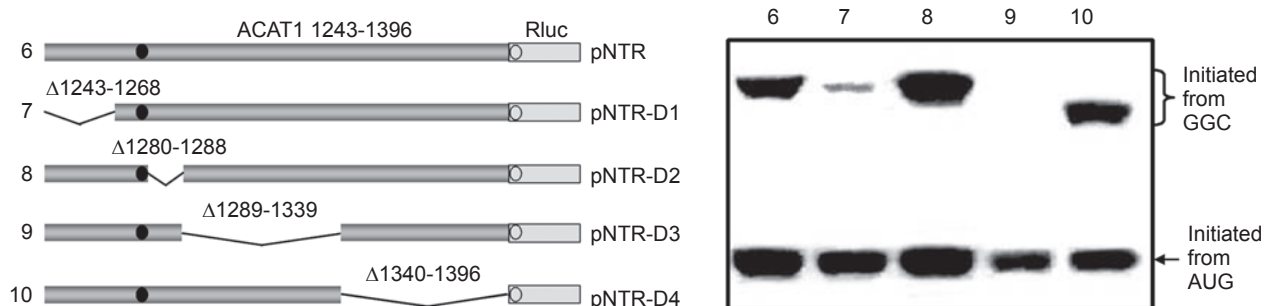$GGC_{1274-1276}$ codon, a plasmid was constructed to contain nt 1243-1786 of human ACAT1 cDNA K1 (ACAT1 1243-1786), which is the minimal sequence identified for the proteins translated from the $GGC_{1274-1276}$ codon [17], with a 3×Flag sequence at the 3′-end. Following this, sequential deletions (Δ1243-1268, Δ1280-1288, Δ1289-1339 and Δ1340-1369) were performed (Figure 1C, left panel). After transfection with these expression plasmids, the production of both proteins, initiated from either the in-frame $GGC_{1274-1276}$ codon or the $AUG_{1397-1399}$ codon, was observed. As shown in the right panel of Figure 1C, deletion of stem-loop I or II dramatically reduced the production of protein initiated from the $GGC_{1274-1276}$ codon, but had little effect on the production of protein initiated from the $AUG_{1397-1399}$ codon (No. 2 and 4). Deletion of stem-loop III had no significant effect on the production of protein initiated from the $GGC_{1274-1276}$ codon (Figure 1C, right panel, No. 5). The same results were obtained when an anti-Flag antibody (M2) was employed for western blot analysis (data not shown).

To study the effect of the partial ACAT1 AUG-ORF sequence (nt 1397-1786) on translation initiation from the in-frame $GGC_{1274-1276}$ and $AUG_{1397-1399}$ codons, we replaced this sequence with the AUG-ORF for Renilla luciferase (Rluc) in the expression plasmids (Figure 1D, left panel). After transfection, western blotting with anti-Rluc antibody was performed to examine the protein expression from these plasmids. All the results in the right panel of Figure 1D, as with those in the right panel of Figure 1C, indicate that the partial ACAT1 AUG-ORF sequence has no effect on the production of protein initiated from either the $GGC_{1274-1276}$ codon or the $AUG_{1397-1399}$ codon. Additionally, owing to the deletions in the coding region of the protein initiated from the $GGC_{1274-1276}$ codon, proteins shorter than the controls in the right panels of Figure 1C and 1D were observed.

These results indicate that both stem-loop I (nt 1255-1268, from chromosome 7) and stem-loop II (nt 1286-1342, from chromosome 1), but not stem-loop III (nt 1355-1384, from chromosome 1), are required for the production of proteins initiated from the $GGC_{1274-1276}$ codon. In conclusion, the optimal production of proteins from the $GGC_{1274-1276}$ codon requires two distinct upstream and downstream RNA secondary structures, which are transcribed from two different chromosomes.

The translation initiation from the $GGC_{1274-1276}$ codon is mediated by an internal ribosome entry site

Considering the long 5′-UTR in the chimeric human ACAT1 mRNA and the translation initiation from a non-AUG codon needed for production of the 56-kDa isoform, we proposed that the translation initiation from the $GGC_{1274-1276}$ codon, which requires the upstream stem-loop I and downstream stem-loop II, might be mediated by an internal ribosome entry site (IRES). So far, two strategies for constructing monocistronic or bicistronic vectors involving a stable hairpin have been used for identification of different IRESs [18-20]. We used both of these strategies to test whether translation initiation from the $GGC_{1274-1276}$ codon is mediated by an IRES. Under the first protocol, nt 1243-1786 of human ACAT1 cDNA K1 (ACAT1 1243-1786) with a 3×Flag sequence at the 3′-end was inserted into the vectors without or with a 5′-stable hairpin (ΔG = −57 kcal/mol) that can impair cap-dependent ribosome scanning to generate monocistronic plasmids (Figure 2A, No. 1 and 11). Two negative control plasmids (Figure 2A, No. 12 and 13) were then constructed by deleting the neighboring sequence of the $GGC_{1274-1276}$ codon (Δ1243-1396) containing the predicted stem-loops within the previously constructed monocistronic plasmids with or without a 5′-stable hairpin. The plasmid phNTF (No. 11) was expected

Figure 1 Predicted stem-loops in the vicinity of the $GGC_{1274-1276}$ codon are required for the production of ACAT1 isoforms. **(A)** RNA sequence of the $GGC_{1274-1276}$ codon-containing vicinity (nt 1243-1396). The initiation codons $GGC_{1274-1276}$ for the 56-kDa and $AUG_{1397-1399}$ for the 50-kDa human ACAT1 isoforms are underlined. **(B)** Predicted RNA secondary structures in the vicinity (nt 1243-1396) of the $GGC_{1274-1276}$ codon. The three successive stem-loops are, respectively, labeled as I, II and III. **(C)** Schematic representation of the partial ACAT1 mRNA sequence (nt 1243-1786) and its truncated forms in the left panel. The deleted regions (Δ1243-1268, Δ1280-1288, Δ1289-1339 and Δ1340-1369) are marked on the top of each bar. Gray bar, ACAT1 mRNA sequence (ACAT1 1243-1786); black bar, 3×Flag coding sequence (3×Flag); filled circle, $GGC_{1274-1276}$ initiation codon; hollow circle, $AUG_{1397-1399}$ initiation codon. The expression plasmids depicted on the left were transiently transfected into AC29; lysates were prepared and immunoblotting was carried out with anti-ACAT1 antibodies (DM10). In the right panel, the curly bracket and the arrow indicate the positions of ACAT1-NT-Flag proteins initiated from $GGC_{1274-1276}$ and $AUG_{1397-1399}$, respectively. The experiments were repeated three times with similar results. **(D)** Schematic representation of the replacement of partial ACAT1 AUG-ORF with Rluc AUG-ORF in the left panel. Gray bar, vicinity of the $GGC_{1274-1276}$ codon (ACAT1 1243-1396); hatched bar, Rluc AUG-ORF (Rluc); others representing the same in **(C)**. The expression plasmids depicted on the left were transiently transfected into AC29; lysates were prepared and immunoblotting was carried out with an anti-Rluc antibody. In the right panel, the curly bracket and the arrow indicate the positions of the fused ACAT1-Rluc protein initiated from the $GGC_{1274-1276}$ and the Rluc protein initiated from $AUG_{1397-1399}$, respectively. The experiments were repeated three times with similar results.

to express the protein initiated from the GGC$_{1274-1276}$ codon only if there was an IRES in the vicinity of this codon (nt 1243-1396). Under the second protocol, three bicistronic vectors as negative controls (pRnF, phRnF and pRhnF) lacking any intercistronic sequence were constructed so that they contained the AUG-ORF for Rluc as the first cistron and the AUG-ORF for Firefly luciferase (Fluc) as the second cistron without (Figure 2C, No. 15) or with (Figure 2C, No. 17 and 19) the stable hairpin ($\Delta G = -57$ kcal/mol) at the position upstream or downstream of the first Rluc cistron (see "Materials and Methods"). Next,

using these bicistronic vectors, three bicistronic plasmids (pRAF, phRAF and pRhAF) were further constructed to constitute the fused second cistron by linking the in-frame GGC$_{1274-1276}$ codon-containing region (nt 1243-1396) to the 5′-end of Fluc AUG-ORF (Figure 2C, No. 14, 16 and 18), and used to detect IRESs in the nearby sequences of the GGC$_{1274-1276}$ codon. If the fused protein (ACAT1-Fluc) initiated from the GGC$_{1274-1276}$ codon of the fused second cistron could be detected with these three bicistronic plasmids, both ribosomal read-through over the first Rluc cistron and re-initiation of the fused second cistron can be ruled out,
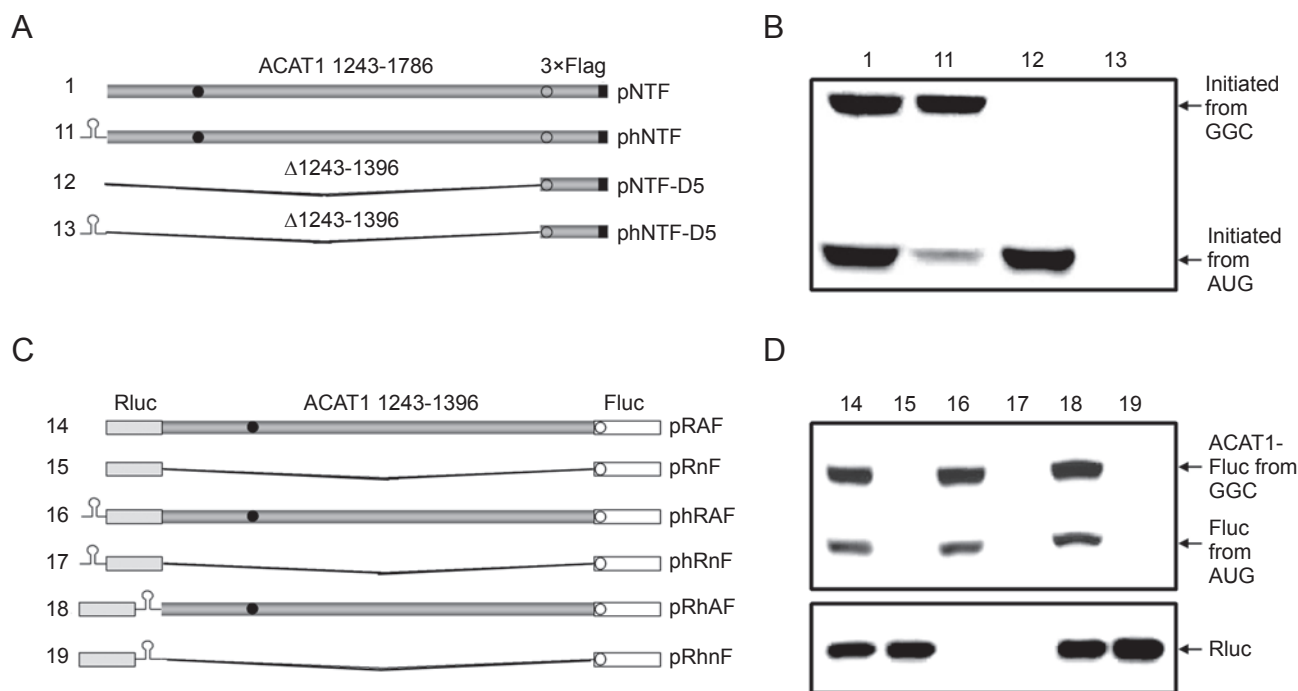

Figure 2 Translation initiation from the GGC$_{1274-1276}$ codon is mediated by an IRES. **(A)** Schematic representation of the partial ACAT1 mRNA sequence (nt 1243-1786) and its truncated form (nt 1397-1786) without or with a 5′-stable hairpin. The stable hairpin ($\Delta G = -57$ kcal/mol) is located at the position upstream of the partial ACAT1 mRNA sequences, and its truncated form in the plasmids phNTF and phNTF-D5. The vicinity of the GGC$_{1274-1276}$ codon was deleted (Δ1243-1396) in the negative control plasmids pNTF-D5 and phNTF-D5. Gray bar, ACAT1 mRNA sequence (ACAT1 1243-1786); black bar, 3×Flag coding sequence (3×Flag); filled circle, GGC$_{1274-1276}$ initiation codon; hollow circle, AUG$_{1397-1399}$ initiation codon. **(B)** The expression plasmids depicted in **(A)** were transiently transfected into AC29; lysates were prepared and immunoblotting was carried out with an anti-ACAT1 antibody (DM10). Arrows indicate the positions of ACAT1-NT-Flag proteins initiated from GGC$_{1274-1276}$ and AUG$_{1397-1399}$, respectively. The experiments were repeated three times with similar results. **(C)** Schematic representation of the two cistrons without or with the stable hairpin. The plasmids pRnF, phRnF and pRhnF as negative controls contain the first Rluc and second Fluc cistrons. The plasmids pRAF, phRAF and pRhAF contain the first Rluc cistron and the fused second cistron by linking the GGC$_{1274-1276}$ codon-containing vicinity (nt 1243-1396) to the 5′-end of Fluc AUG-ORF. The stable hairpin depicted in **(A)** is located at the position upstream or downstream of the first Rluc cistron. Gray bar, vicinity of the GGC$_{1274-1276}$ codon (ACAT1 1243-1396); hatched bar, Rluc AUG-ORF (Rluc); white bar, Fluc AUG-ORF (Fluc); others representing the same in **(A)**. **(D)** The expression plasmids depicted in **(C)** were transiently transfected into AC29; lysates were prepared and immunoblotting was carried out with anti-Fluc antibody and anti-Rluc antibody, respectively. The immunoblotting result with anti-Fluc antibody is shown in the top panel and arrows indicate the positions of the fused ACAT1-Fluc protein initiated from the GGC$_{1274-1276}$ and Fluc protein initiated from AUG$_{1397-1399}$. The immunoblotting result with anti-Rluc antibody is shown in the bottom panel and an arrow indicates the position of Rluc protein. The experiments were repeated three times with similar results.

further implying that there is an IRES in the vicinity of the GGC$_{1274-1276}$ codon (nt 1243-1396).

After transfection with these expression plasmids, western blotting was performed to detect the expressed proteins. The results of the experiments using the monocistronic plasmids (Figure 2A) showed that the amount of protein initiated from the GGC$_{1274-1276}$ codon was apparently unaffected by the presence or absence of the 5′-stable hairpin (Figure 2B, No. 1 and 11), while no target protein could be observed in the negative controls lacking the sequence neighboring this codon (Figure 2B, No. 12 and 13). It was clear that there was an IRES in the vicinity of the GGC$_{1274-1276}$ codon (nt 1243-1396) that mediated translation from this initiation codon in a cap-independent manner (Figure 2B, No. 1 and 11). Further results obtained with three bicistronic plasmids demonstrated that the levels of the fused proteins (ACAT1-Fluc) translated from the GGC$_{1274-1276}$ codon of the fused second cistron were similar in the absence (Figure 2D, top panel, No. 14) or presence of the stable hairpin at the position upstream or downstream of the first Rluc cistron (Figure 2D, top panel, No. 16 and 18). In contrast, the 5′-stable hairpin potently impaired the translation of the first cistron (Rluc) in a cap-dependent manner (Figure 2D, bottom panel, No. 16 and 17). The second cistron (Fluc) in the control plasmids, which lacked an intercistronic sequence, could not be translated (Figure 2D, top panel, No. 15, 17 and 19). These results conclusively indicate that translation from the GGC$_{1274-1276}$ codon is mediated by an IRES in a cap-independent manner. In addition, protein initiated from the downstream in-frame AUG$_{1397-1399}$ codon was also observed in both the monocistron with the 5′-stable hairpin (Figure 2B, No. 11) and the fused second cistron (Figure 2D, top panel, No. 14, 16 and 18) only when the vicinity (nt 1243-1396) of the GGC$_{1274-1276}$ codon was present.

Although the foregoing data provide strong evidence that translation from the GGC$_{1274-1276}$ codon is mediated by an IRES, two possibilities cannot yet be excluded through the above experiments. The first is that cryptic promoter activity may result in transcription of RNA from the vicinity (nt 1243-1396) of the GGC$_{1274-1276}$ codon, which could then be translated into protein. The other is that the presence of splicing sites may affect protein expression. To rule out the first possibility, transfections were performed by using the monocistronic and bicistronic plasmids with or without the eukaryotic CMV promoter (Figure 3A). Western blotting showed that protein expression could not be detected in the absence of the eukaryotic CMV promoter (Figure 3B, No. 20 and 22). This result clearly shows that there is no cryptic promoter in the vicinity (nt 1243-1396) of the GGC$_{1274-1276}$ codon. To rule out the second possibility, the integrity of the bicistronic mRNAs was examined in

cells transfected with the bicistronic plasmids (Figure 3C). Reverse transcription-PCR (RT-PCR) analysis of the total RNA from pRAF (No. 14)-transfected cells showed that only the full-length bicistronic mRNAs (Figure 3D, lanes b and d) were present *in vivo*, similar to the results from the cells transfected with the control plasmid pRHIF (No. 23) containing the HCV IRES (Figure 3D, lanes f and h). These results demonstrate that there is no splicing site in the vicinity (nt 1243-1396) of the GGC$_{1274-1276}$ codon.

Collectively, these data show that translation initiation from the GGC$_{1274-1276}$ codon is mediated by an IRES. The above-described sequential deletions in the vicinity (nt 1243-1396) of the GGC$_{1274-1276}$ codon (shown in Figure 1C and 1D) were then performed to generate monocistronic plasmids with the 5′-stable hairpin and the bicistronic plasmids containing the first (Rluc) and second (Fluc) cistrons (Figure 4A and 4C). Western blotting revealed that, as predicted, the deletion of stem-loop I (Δ1243-1268) or II (Δ1289-1339) clearly reduced the production of protein from the GGC$_{1274-1276}$ initiation codon (Figure 4B, No. 24 and 26; Figure 4D, No. 28 and 30). We have thus confirmed that the translation initiated at the GGC$_{1274-1276}$ codon mediated by an IRES requires both upstream stem-loop I and downstream stem-loop II from two different chromosomes.

AU constitution of upstream RNA secondary structure from chromosome 7 is important for translation initiation from the GGC$_{1274-1276}$ codon

To characterize the effects of the upstream RNA secondary structure transcribed from chromosome 7 on translation initiation from the GGC$_{1274-1276}$ codon, various mutations were introduced into stem-loop I (Figure 5A). Initially, twelve A-U interconversions (nt 1255-1259, nt 1261-1263 and 1264-1268, pNTF-M1) that do not disrupt the secondary structure were introduced into stem-loop I (Figure 5C, No. 32). The resultant production of proteins initiated from both the GGC$_{1274-1276}$ and AUG$_{1397-1399}$ codons was similar to the wild-type levels (Figure 5B, No. 1 and 32). When three U to A substitutions were introduced (pNTF-M2) to disrupt step-loop I (Figure 5C, No. 33), the production of protein initiated from the GGC$_{1274-1276}$ codon was dramatically reduced, while the production of protein initiated from the AUG$_{1397-1399}$ codon did not show any obvious changes (Figure 5B, No. 33). Interestingly, two additional A to U substitutions (pNTF-M3) that reconstitute this stem-loop (Figure 5C, No. 34) can completely rescue the production of protein initiated from the GGC$_{1274-1276}$ codon (Figure 5B, No. 34). However, disruption of stem-loop I by mutating UAUU to CGCG (pNTF-M4) reduced the production of protein initiated from the GGC$_{1274-1276}$ codon (Figure 5B, No. 35), and this production could not be rescued by re-
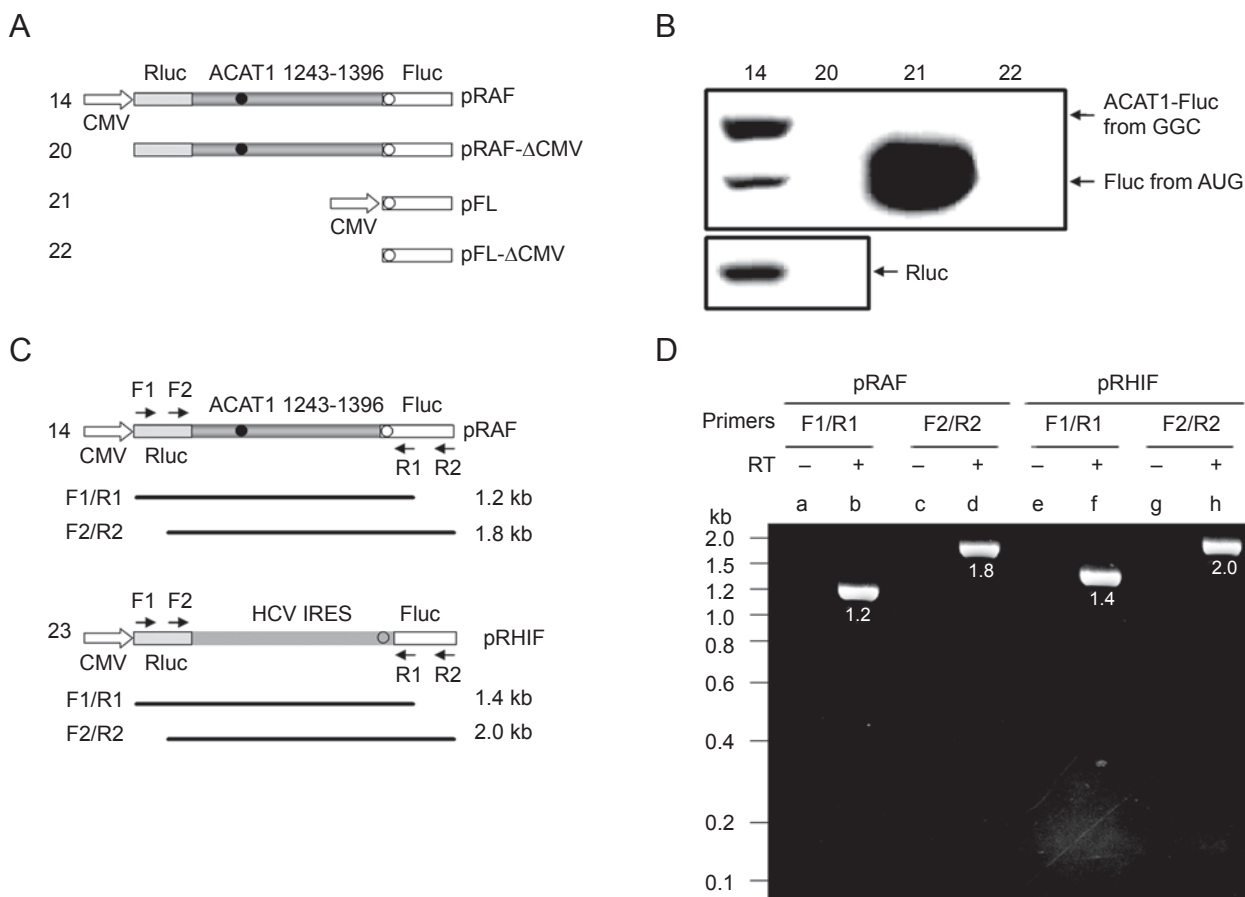
Figure 3 The vicinity of the $GGC_{1274-1276}$ codon in the bicistronic plasmid does not show cryptic promoter activity or contain an aberrant splicing site. **(A)** Schematic representation of the bicistron and the monocistron with or without the eukaryotic CMV promoter. The bicistronic plasmids pRAF and pRAF-ΔCMV contain the first Rluc cistron and the fused second cistron depicted in Figure 2C. The monocistronic plasmids pFL and pFL-ΔCMV only contain a Fluc cistron. The eukaryotic CMV promoter is deleted in the plasmids pRAF-ΔCMV and pFL-ΔCMV, respectively. Gray bar, vicinity of the $GGC_{1274-1276}$ codon (ACAT1 1243-1396); hatched bar, Rluc AUG-ORF (Rluc); white bar, Fluc AUG-ORF (Fluc); filled circle, $GGC_{1274-1276}$ initiation codon; hollow circle, $AUG_{1397-1399}$ initiation codon. **(B)** The expression plasmids depicted in **(A)** were transiently transfected into AC29; lysates were prepared and immunoblotting was carried out with anti-Fluc antibody and anti-Rluc antibody, respectively. The immunoblotting result with anti-Fluc antibody is shown in the top panel and arrows indicate the positions of the fused ACAT1-Fluc protein initiated from the $GGC_{1274-1276}$ and Fluc protein initiated from $AUG_{1397-1399}$. The immunoblotting result with anti-Rluc antibody is shown in the bottom panel and an arrow indicates the position of Rluc protein. The experiments were repeated three times with similar results. **(C)** Schematic representation of the location of four PCR primers at the bicistronic plasmids pRAF and pRHIF. The primers F1 and F2 are complementary to the 5′- and 3′-ends of Rluc AUG-ORF, respectively, and the primers R1 and R2 are complementary to the 5′- and 3′-ends of Fluc AUG-ORF, respectively. The calculated sizes of PCR products from using primer sets F1/R1 and F2/R2 are shown. Spots bar, region of HCV IRES; hollow circle, AUG initiation codon; others representing the same in **(A)**. **(D)** RT-PCR analysis of total RNA from the cells transfected with the plasmids pRAF and pRHIF was performed with two sets of primers, F1/R1 and F2/R2. Lanes b (1.2 kb), d (1.8 kb), f (1.4 kb) and h (2.0 kb) show the amplified products with the size depicted on left from the pRAF- or pRHIF-transfected cells when the F1/R1 and F2/R2 primer sets were used. Lanes a, c, e and g show reverse transcription-negative controls for each sample. The experiments were repeated three times with similar results.

building the stem-loop with a GC-rich stem (Figure 5B, No. 36).

According to the above results, we proposed that the stability of upstream stem-loop I (Figure 5C, No. 36) was related to the production of protein initiated from the $GGC_{1274-1276}$ codon (Figure 5B, No. 36). In further experiments, the A-U base pairs in the stem were progressively changed to G-C base pairs (Figure 6A) to increase the
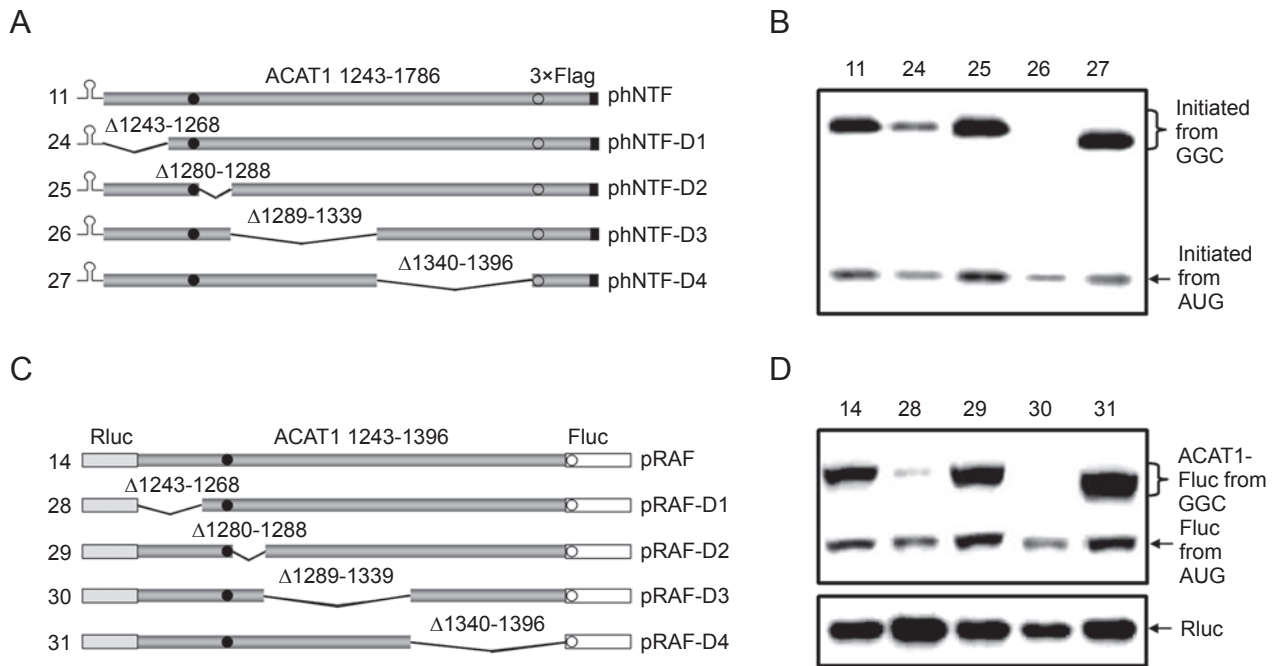
Figure 4 The translation initiation from the $GGC_{1274-1276}$ codon mediated by an IRES requires both stem-loop I and II. **(A)** Schematic representation of the partial ACAT1 mRNA sequence (nt 1243-1786) and its truncated forms with the 5′-stable hairpin. The deleted regions (Δ1243-1268, Δ1280-1288, Δ1289-1339 and Δ1340-1369) are marked on top and the 5′-stable hairpin (ΔG = –57 kcal/mol) is located at the position upstream of the 5′-end. Gray bar, ACAT1 mRNA sequence (ACAT1 1243-1786); black bar, 3×Flag coding sequence (3×Flag); filled circle, $GGC_{1274-1276}$ initiation codon; hollow circle, $AUG_{1397-1399}$ initiation codon. **(B)** The expression plasmids depicted in **(A)** were transiently transfected into AC29, lysates were prepared and immunoblotting was carried out with anti-ACAT1 antibodies (DM10). The curly bracket and the arrow indicate the positions of ACAT1-NT-Flag proteins initiated from $GGC_{1274-1276}$ and $AUG_{1397-1399}$, respectively. The experiments were repeated twice with similar results. **(C)** Schematic representation of two cistrons and the truncated forms of the fused second cistron depicted in Figure 2C. The deleted regions (Δ1243-1268, Δ1280-1288, Δ1289-1339 and Δ1340-1369) are marked on top. Gray bar, vicinity of the $GGC_{1274-1276}$ codon (ACAT1 1243-1396); hatched bar, Rluc AUG-ORF (Rluc); white bar, Fluc AUG-ORF (Fluc); others representing the same in **(A)**. **(D)** The expression plasmids depicted in **(C)** were transiently transfected into AC29, lysates were prepared and immunoblotting was carried out with anti-Fluc antibody and anti-Rluc antibody, respectively. The immunoblotting result with anti-Fluc antibody is shown in the top panel, and the curly bracket and the arrow indicate the positions of the fused ACAT1-Fluc protein initiated from the $GGC_{1274-1276}$ and Fluc protein initiated from $AUG_{1397-1399}$, respectively. The immunoblotting result with anti-Rluc antibody is shown in the bottom panel and an arrow indicates the position of the Rluc protein. The experiments were repeated twice with similar results.

stability of this stem-loop (Figure 6C). Western blotting demonstrated that the production of protein initiated from the $GGC_{1274-1276}$ codon gradually decreased as the GC content in the stem-loop increased (Figure 6B, No. 1 and 37-39), indicating that the production of protein initiated from the $GGC_{1274-1276}$ codon is inversely correlated with the predicted stability of the upstream RNA secondary structures (Figure 6D).

Taken together, these results demonstrate that the upstream AU-constituted RNA secondary structure from chromosome 7 is important for the translation initiation of the ACAT1 56-kDa isoform from the $GGC_{1274-1276}$ codon.

GC-richness of the downstream RNA secondary structure transcribed from chromosome 1 is essential for translation initiation from the $GGC_{1274-1276}$ codon

As the earlier results (Figure 1C and 1D) demonstrated that the production of protein initiated at the $GGC_{1274-1276}$ codon was significantly decreased when stem-loop II was entirely deleted, we performed sub-deletions (Δ1289-1306, Δ1307-1324 and Δ1325-1339) of this stem-loop (pNTF-D6 to -D8 in Figure 7A). Western blotting (Figure 7B) indicated that none of these sub-deletions inhibited the production of protein initiated from the $GGC_{1274-1276}$ codon (No. 40-42) to the same extent as the complete deletion (No. 4). By RNA secondary structure prediction, we found that the sequence (nt 1274-1354) with the complete deletion was unlikely to be able to form a stable secondary
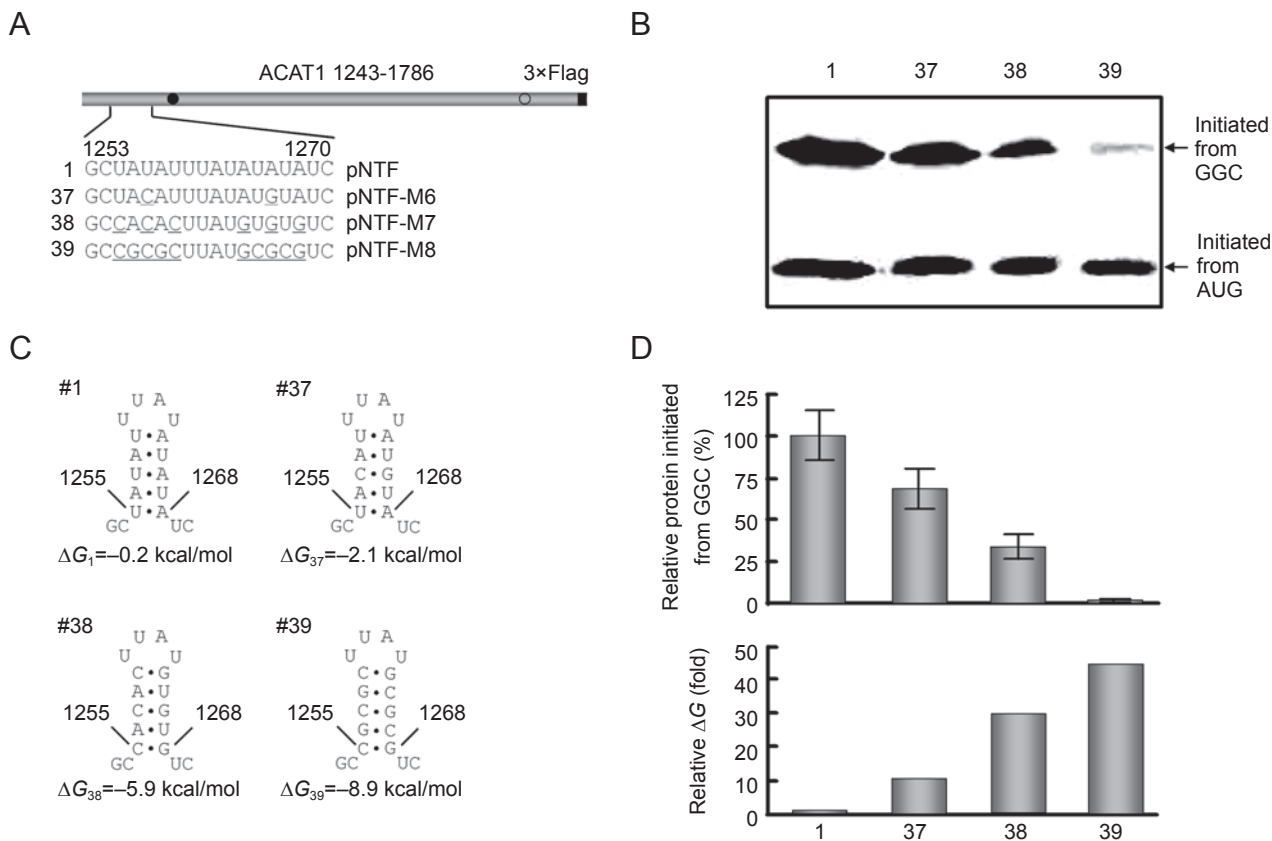
A

B

C

Figure 5 AU constitution of the upstream stem-loop I is important for producing proteins initiated from the $GGC_{1274-1276}$ codon. **(A)** Schematic representation of the partial ACAT1 mRNA sequence (nt 1243-1786) and mutations of stem-loop I. The detailed sequence of nt 1253-1270 is listed without or with the mutant nucleotides (underlined). Gray bar, partial ACAT1 mRNA sequence (ACAT1 1243-1786); black bar, 3×Flag coding sequence (3×Flag); filled circle, $GGC_{1274-1276}$ initiation codon; hollow circle, $AUG_{1397-1399}$ initiation codon. **(B)** The expression plasmids depicted in **(A)** were transiently transfected into AC29; lysates were prepared and immunoblotting was carried out with an anti-ACAT1 antibody (DM10). Arrows indicate the positions of ACAT1-NT-Flag proteins initiated from $GGC_{1274-1276}$ and $AUG_{1397-1399}$, respectively. The experiments were repeated twice with similar results. **(C)** Predicted RNA secondary structures of nt 1253-1270 without or with mutations of stem-loop I depicted in **(A)**. The folding Gibbs free energy (ΔG) of predicted RNA secondary structure is listed beneath each structure. These experiments were repeated three times with similar results.

structure (Figure 7C, No. 4), while the three sub-deletions could still form RNA secondary structures with relatively strong stabilities (compared with the complete deletion) and high GC contents (Figure 7C, No. 40-42). We therefore proposed that a downstream GC-rich RNA secondary structure could be essential for translation initiation from the $GGC_{1274-1276}$ codon.

To validate this hypothesis, we replaced most of the GC nucleotides with AU nucleotides in the downstream RNA secondary structure formed after sub-deletions of the stem-loop II region (Figure 8A). The mutated nucleotides in the mRNA produced from the constructed plasmids pNTF-M9, -M11 and -M13 disrupted the formation of secondary structures, but in the pNTF-M10, -M12 and -M14 plasmids the mutated nucleotides maintained these

secondary structures via AU-richness rather than GC-richness (Figure 8B). Western blotting showed that the production of protein initiated from the $GGC_{1274-1276}$ codon was completely eliminated (Figure 8C, No. 43-44, No. 45-46 and No. 47-48) when most of the GC nucleotides were replaced by AU nucleotides, independently of whether the downstream RNA secondary structures were disrupted or maintained (Figure 8B). It should be noted that the production of protein initiated from the $AUG_{1397-1399}$ codon was also altered when the AU nucleotides were introduced (Figure 8C, No. 43-48). Thus, to determine whether this effect is caused by the sequence changes in the RNA, a RT-quantitative PCR (qPCR) experiment was performed. The results showed no significant alterations among the related mRNAs (Figure 8D).

A

ACAT1 1243-1786   3×Flag

```
          1253            1270
   1   GCUAUAUUUAUAUAUC   pNTF
  37   GCUACAUUUAUAUGUAUC pNTF-M6
  38   GCCACACUUAUGUGUGUC pNTF-M7
  39   GCCGCGCUUAUGCGCGUC pNTF-M8
```

B

C

```
#1      U A                   #37     U A
      U     U                       U     U
       U • A                         U • A
       A • U                         A • U
1255   U • A   1268         1255    C • G   1268
       A • U                         A • U
       U • A                         U • A
      GC     UC                     GC     UC
```

$\Delta G_1 = -0.2$ kcal/mol $\Delta G_{37} = -2.1$ kcal/mol

```
#38     U A                   #39     U A
      U     U                       U     U
       C • G                         C • G
       A • U                         G • C
1255   C • G   1268         1255    C • G   1268
       A • U                         G • C
       C • G                         C • G
      GC     UC                     GC     UC
```

$\Delta G_{38} = -5.9$ kcal/mol $\Delta G_{39} = -8.9$ kcal/mol

D

Figure 6 Stability of the upstream RNA secondary structure is inversely correlated with production of protein initiated from the GGC$_{1274-1276}$ codon. **(A)** Schematic representation of the partial ACAT1 mRNA sequence (nt 1243-1786) and the replacement of AU nucleotides by GC nucleotides in stem-loop I. The detailed sequence of nt 1253-1270 is listed and the replaced nucleotides are underlined. Gray bar, partial ACAT1 mRNA sequence (ACAT1 1243-1786); black bar, 3×Flag coding sequence (3×Flag); filled circle, GGC$_{1274-1276}$ initiation codon; hollow circle, AUG$_{1397-1399}$ initiation codon. **(B)** The expression plasmids depicted in **(A)** were transiently transfected into AC29, lysates were prepared and immunoblotting was carried out with an anti-ACAT1 antibody (DM10). Arrows indicate the positions of ACAT1-NT-Flag proteins initiated from GGC$_{1274-1276}$ and AUG$_{1397-1399}$, respectively. The experiments were repeated three times with similar results. **(C)** Predicted RNA secondary structures of nt 1253-1270 without or with mutations of stem-loop I depicted in **(A)**. The folding Gibbs free energy (ΔG) of predicted RNA secondary structure is listed beneath each structure. **(D)** Relative production of ACAT1-NT-Flag protein initiated from the GGC$_{1274-1276}$ codon (up panel) and folding ΔG of RNA secondary structures (down panel). The intensity of products from the GGC$_{1274-1276}$ codon in **(B)** was quantified by using the UVP Labwork software (UVP Inc.) for densitometric analysis and normalized to the value for wild type. The data represent the means ± SD from three independent experiments. The relative folding ΔG of RNA secondary structures in **(C)** is shown by setting the wild type value at 1.0.

In summary, in addition to the upstream AU-constituted RNA secondary structure from chromosome 7, the downstream GC-rich RNA secondary structure from chromosome 1 is also essential for proper translation initiation from the GGC$_{1274-1276}$ codon.

Discussion

There are three stem-loops in the vicinity of the GGC$_{1274-1276}$ codon, which is located in the interchromosomal region of the chimeric human ACAT1 mRNA (Figure 1B). We find

that two of these RNA secondary structures are required for the production of proteins initiated from the GGC$_{1274-1276}$ codon. The upstream stem-loop I from chromosome 7 most effectively promotes production of protein from the GGC$_{1274-1276}$ codon when it is constituted with AU nucleotides, although this structure can tolerate some changes to the nucleotide sequence that do not disrupt the structure (Figures 5 and 6). The downstream GC-rich stem-loop II from chromosome 1 is the largest and the most complicated of the three predicted stem-loops. Experimental results demonstrated that production of protein initiated
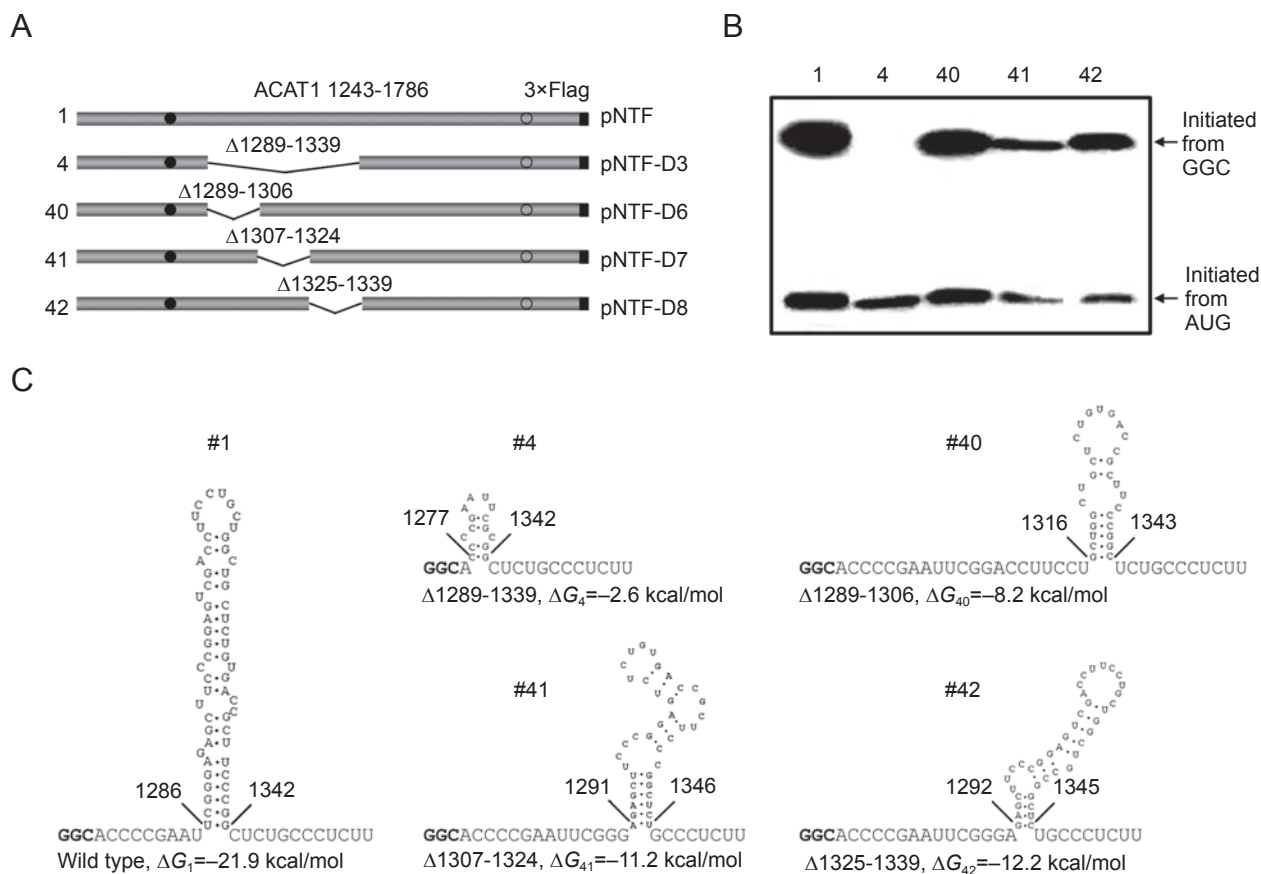
Figure 7 The downstream RNA secondary structure is required for producing proteins initiated from the $GGC_{1274-1276}$ codon. **(A)** Schematic representation of the partial ACAT1 mRNA sequence (nt 1243-1786) and the whole- or sub-deletions of stem-loop II. The deleted regions (Δ1289-1339, Δ1289-1306, Δ1307-1324 and Δ1325-1339) are marked on top. Gray bar, partial ACAT1 mRNA sequence (ACAT1 1243-1786); black bar, 3×Flag coding sequence (3×Flag); filled circle, $GGC_{1274-1276}$ initiation codon; hollow circle, $AUG_{1397-1399}$ initiation codon. **(B)** The expression plasmids depicted in **(A)** were transiently transfected into AC29, lysates were prepared and immunoblotting was carried out with an anti-ACAT1 antibody (DM10). Arrows indicate the positions of ACAT1-NT-Flag proteins initiated from $GGC_{1274-1276}$ and $AUG_{1397-1399}$, respectively. The experiments were repeated twice with similar results. **(C)** Predicted RNA secondary structures of nt 1274-1354 without or with the whole- or sub-deletions depicted in **(A)**. The folding Gibbs free energies (ΔG) of each predicted RNA secondary structure are listed beneath each structure.

from the $GGC_{1274-1276}$ codon requires a downstream RNA secondary structure with GC-richness. This downstream secondary structure can be formed after certain sub-deletions are made, even if the resultant stabilities are relatively weaker than the wild type (Figure 7). Further results showed that the GC-richness of this downstream RNA secondary structure is indispensable (Figure 8). Detailed experiments revealed that translation initiation from the $GGC_{1274-1276}$ codon is mediated by an IRES (Figures 2 and 3). Our results suggest that this IRES-mediated translation initiation requires the upstream RNA secondary structure with AU-constitution and the downstream GC-rich structure (Figure 4). In addition, there was a decrease in the amount of protein initiated from the downstream in-frame $AUG_{1397-1399}$ codon in a cap-independent manner from both

the monocistron with 5′-stable hairpin (Figure 2B, No. 11) and the fused second cistron (Figure 2D, No. 14, 16 and 18), but only when the upstream vicinity (nt 1243-1396) of the $GGC_{1274-1276}$ codon is present. This result implies that there might be another IRES in the nt 1243-1396 region that can mediate the translation initiated from the $AUG_{1397-1399}$ codon. This mechanistic work provides further evidence that the chimeric human ACAT1 mRNA is expressed from two different chromosomes.

Our previous studies showed that the 56-kDa isoform was enzymatically active and that its activity is about 30% that of the 50-kDa ACAT1 [17]. When the 56-kDa isoform and 50-kDa ACAT1 were co-expressed in the same cell, the normalized ACAT1 activity was obviously lower than that of the 50-kDa ACAT1 protein alone [17]. Considering

that the 50-kDa ACAT1 protein forms homotetramers in intact cells and *in vitro* [2], it is possible that the 56-kDa isoform might serve as an endogenous inhibitor of the 50-kDa ACAT1 by forming hetero-oligomers that reduce allosteric enzyme activity in cases where the high activity of ACAT should be down-regulated. On the other hand, in some cell-stressed conditions that inhibit cap-dependent translation, it can be postulated that the dynamic interconversion between cellular free cholesterol and cholesterol ester could be maintained by the 56-kDa isoform, which is initiated from the $GGC_{1274-1276}$ codon by an IRES in a cap-independent manner. Therefore, this 56-kDa ACAT1 isoform could be produced under various conditions including cellular stress and permit a more elaborate regulation of the activity of this allosteric tetrameric enzyme for cellular cholesterol homeostasis.

Other investigators have shown that most of the RNA secondary structures located in the long 5′-UTR could function as IRESs [21, 22], and, in a few cases, the sequence downstream of the translation initiation codon is required for IRES activity [23]. In the present context, both the upstream and downstream RNA secondary structures that are transcribed from two different chromosomes are required for IRES-mediated translation initiation from the $GGC_{1274-1276}$ codon. It has been reported that RNA secondary structures perform their IRES functions by recruiting various proteins, including translation initiation factors. For example, the EMCV IRES needs to recruit eIF4A and eIF4G to initiate translation [24, 25]. Poliovirus and rhinovirus IRESs both bind to PTB and to the poly(rC) binding protein 2 (PCBP2) [26, 27]. In addition, RNA secondary structures are also involved in regulating gene expression at multiple levels, such as splicing, polyadenylation and editing [28-30]. We note that the upstream RNA secondary structure is AU-constituted and that the downstream structure is GC-rich, suggesting that the characteristics of these structures may be needed to recruit and bind certain proteins in order to produce the 56-kDa human ACAT1 isoform from the $GGC_{1274-1276}$ codon. In our future studies, we will attempt to identify the proteins bound to the upstream and downstream RNA secondary structures from two different chromosomes, and to further examine the mechanism by which they regulate translation initiation.

Materials and methods

Materials

Cell culture reagents and T4 DNA ligase were purchased from Invitrogen (Carlsbad, USA). The agarose and all restriction enzymes were from Promega (Madison, USA). Anti-rabbit and anti-mouse antibodies (IgGs) conjugated with HRPs were from Pierce (Rockford, USA). Taq DNA polymerase and dNTPs were from Sino-American Biotech (Shanghai, China). All the oligonucleotides were synthesized with an automated DNA synthesizer at the Institute of Biochemistry and Cell Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences.

Cell culture and transfection

AC29 cells, the mutant CHO cell line lacking the endogenous ACAT1 [31], were maintained in a basal Ham's F12 medium, supplemented with 10% fetal bovine serum (FBS) and antibiotics, in a humid atmosphere of 5% CO_2 and 95% air at 37 °C. The constructed expression plasmids were individually transfected into AC29 cells using FuGENE 6^{TM} transfection reagent (Roche) according to the manufacturer's instructions, and the transfected cells were cultured for another 48 h before harvesting.

Expression plasmids

The sequence encoding 3×Flag was amplified from the p3×FLAG-CMV-14 expression vector (Sigma) by PCR with the primer set (3FlagF, 5′-AGT GAA CCG TCA GAA TTA AGC-3′/3FlagR, 5′-AAA GGG CCC ATC ACT ACT TGT CAT CGT C-3′). The amplified fragment was then purified, digested with *Xba*I and *Apa*I, and inserted into the *Xba*I and *Apa*I sites of pcDNA3 (Invitrogen) to generate the vector p3Flag that can express the 3×Flag tag fused to the C-terminus of the target protein encoded by a gene which is inserted into it.

The partial human ACAT1 cDNA K1 sequence (1243-1786 bp) containing the neighboring sequence of the $GGC_{1274-1276}$ codon followed by the partial AUG-ORF sequence, which can encode a 25-kDa N-terminal fragment of the 56-kDa isoform (ACAT1-NT25) from the $GGC_{1274-1276}$ codon and a 17-kDa N-terminal fragment of the 50-kDa isoform (ACAT1-NT17) from $AUG_{1397-1399}$ codon, was amplified by PCR with the primer set (1243F, 5′-AAA GGT ACC TAG TTA AAT AGC TAT ATT TAT-3′/1786R, 5′-AAA TCT AGA ATC TAA GAG AGA GCG CCT-3′) and inserted into the *Kpn*I and *Xba*I sites of p3Flag to generate the expression plasmid pNTF for production of 26-kDa ACAT1 N-terminal fragment with 3×Flag (ACAT1-NT-Flag26) from the $GGC_{1274-1276}$ codon and 18-kDa ACAT1 N-terminal fragment with 3×Flag (ACAT1-NT-Flag18) from $AUG_{1397-1399}$ codon. The sequential deletions in the $GGC_{1274-1276}$ codon-containing vicinity (1243-1396 bp) upstream of the $AUG_{1397-1399}$ codon were then obtained using the constructed plasmid pNTF. The 5′-region deletion (Δ1243-1268) was obtained by the above one-step PCR with the forward primer D1F (Supplementary information, Table S1) and reverse primer 1786R. The other three deletions (Δ1280-1288, Δ1289-1339 and Δ1340-1396) were obtained by a two-step PCR method described by Higuchi *et al.* [32] with a common external primer set (1243F/1786R, shown above) and individual internal primer sets (D2F/D2R, D3F/D3R and D4F/D4R, shown in Supplementary information, Table S1) with partial complementary sequences. All the amplified fragments with the sequential deletions of the vicinity of $GGC_{1274-1276}$ codon upstream of the $AUG_{1397-1399}$ codon were respectively inserted into the *Kpn*I and *Xba*I sites of p3Flag to generate expression plasmids pNTF-D1 to -D4.

By using pNTF and pNTF-D1 to -D4, the partial ACAT1 AUG-ORF was replaced with the AUG-ORF encoding Rluc (316 amino acids with apparent molecular mass of 36 kDa) to obtain expression plasmids pNTR and pNTR-D1 to -D4. The coding sequence for Rluc was amplified from pRL-CMV (Promega) by PCR with the primer set (RlucF1, 5′-AAA GGC GGC CGC GAC TTC GAA AGT TTA TGA TCC AGA ACA AAG G-3′/RlucR1, 5′-AAA GTC TAG ATT

A

ACAT1 1243-1786   3×Flag

Δ1289-1306   1316   1343
40 **GGC**ACCCCGAAUUCG⌄GACCUUCCUGCUGGCUGCUCUGUGACCGCUUCCCGGCUCUGCCCUCUU   pNTF-D6
43 **GGC**ACCCCGAAUUCG⌄GACCUUCC<u>UAUCUA</u>UUCU<u>A</u>U<u>UAUCAAUA</u>UUC<u>A</u>C<u>UUA</u>UCUGCCCUCUU   pNTF-M9
44 **GGC**ACCCCGAAUUCG⌄GACCUUCC<u>UAUCUA</u>UUCU<u>A</u>UUAUA<u>A</u>CG<u>A</u>UUC<u>A</u>G<u>AUA</u>UCUGCCCUCUU   pNTF-M10
         1291   Δ1307-1324   1346
41 **GGC**ACCCCGAAUUCGGGGAGAGCUUCCCGGAGUC⌄UCUGUGACCGCUUCCCGGCUCUGCCCUCUU   pNTF-D7
45 **GGC**ACCCCGAAUUCGGGGAAAAAAUCAUAAACUG⌄UAUUGUCAAUAUUCCCAUAUAUGCCCUCUU   pNTF-M11
46 **GGC**ACCCCGAAUUCGGGGAU<u>AUA</u>UUC<u>A</u>CGGA<u>C</u>UG⌄UAUUUC<u>AGAUA</u>UUCCC<u>AUAUA</u>UGCCCUCUU   pNTF-M12
         1292   Δ1325-1339   1345
42 **GGC**ACCCCGAAUUCGGGAGAGCUUCCCGGAGUCGACCUUCCUGCUGGCUGG⌄CGGCUCUGCCCUCUU   pNTF-D8
47 **GGC**ACCCCGAAUUCGGGAUUGAUUUUCAGAUUAAUUUAACUUAAAAUAUUUC⌄CUUAUAUGCCCUCUU   pNTF-M16
48 **GGC**ACCCCGAAUUCGGGAGAC<u>U</u>UU<u>A</u>AGAG<u>A</u>UUCUU<u>AA</u>UU<u>AA</u>U<u>A</u>U<u>A</u>G<u>A</u>AU<u>A</u>C⌄UC<u>A</u>GUCUGCCCUCUU   pNTF-M14

B

C

D

GTT CAT TTT TGA GAA CTC GCT CAA CGA A-3′) and inserted into the *Not*I and *Xba*I sites of pcDNA3 to generate vector pRL. The upstream sequences of AUG$_{1397-1399}$ in pNTF, pNTF-D2 and pNTF-D3 were amplified with the forward (1243F shown above) and reverse (1396R, 5′-TCT GCG GCC GCC ATT GTA TTG TCT GAG GCC CGC GCC C-3′) primers. The upstream sequences of AUG$_{1397-1399}$ in pNTF-D1 and pNTF-D4 were respectively amplified with the above primer sets 1269F/1396R and 1243F/1396R-2 (5′-AAG CGG CCG CCA TGG AAG CG-3′). These amplified fragments were individually digested with *Kpn*I and *Not*I and inserted into the *Kpn*I and *Not*I sites of pRL to generate the expression plasmids pNTR, pNTR-D1 to -D4.

The fragment involving a stable hairpin that can block the scanning of the 40S ribosome-factor complex [33] was annealed from two complementary primers (HF1, 5′-AGC TTG GGG CGC GTG GTG GCG GCT GCA GCC GCC ACC ACG CGC CCC GGT AC-3′/HR1, 5′-CGG GGC GCG TGG TGG CGG CTG CAG CCG CCA CCA CGC GCC CCA-3′) and inserted into the *Hind*III and *Kpn*I sites of p3Flag to generate the stable hairpin-containing vector ph3Flag. The partial human ACAT1 cDNA K1 sequence (1243-1786 bp) from pNTF was inserted into the *Kpn*I and *Xba*I sites of ph3Flag to generate the expression plasmid phNTF. The partial ACAT1 AUG-ORF sequence (1396-1786 bp) was amplified with the forward (D5F, shown in Supplementary information, Table S1) and reverse (1786R, shown above) primers and inserted into the *Kpn*I and *Xba*I sites of p3Flag and ph3Flag, respectively, to generate expression plasmids pNTF-D5 and phNTF-D5.

The *Nhe*I site was introduced into pcDNA3 with the primer set (pcDNA3-Nhe-F, 5′-CCC GAA AAG TGC CAC CTG ACG TCG ACG GAT-3′/ pcDNA3-Nhe-R, 5′-AAA AAG CTT ACG CTA GCG GGT CTC CTA TAG TGA GTC GTA TTA ATT TC-3′) to generate the vector pcDNA3-NheI. The AUG-ORF encoding Firefly luciferase (Fluc) was amplified with the primer set (FlucF1, 5′-AAA TCT AGA ATG GAA GAC GCC AAA AAC ATA-3′/FlucR, 5′-AAA GGG CCC TTA CAC GGC GAT CTT TCC-3′) and inserted into the *Xba*I and *Apa*I sites of pcDNA3-NheI to generate the plasmid pFL. The AUG-ORF for Rluc was amplified with the primer set (RlucF2, 5′-AAA AAG CTT ATG ACT TCG AAA GTT TAT GAT-3′/RlucR2, 5′-AAA GCG GCC GCT TAT TGT TCA TTT TTG AGA ACT-3′) and inserted into the *Hind*III and *Not*I sites of the plasmid pFL to generate the

bicistronic plasmid pRnF containing the first Rluc and second Fluc cistrons as the negative control. The vicinity (1243-1396 bp) of the GGC$_{1274-1276}$ codon was amplified with the primer set (1243-*Not*I-F, 5′-AAG CGG CCG CTA GTT AAA TAG CTA TAT TTA T-3′/1396-*Xba*I-R, 5′-AAA TCT AGA CAT GGA AGC GGT CAC AGA-3′) and inserted into the *Not*I and *Xba*I sites of pRnF to generate the bicistronic plasmid pRAF containing the first Rluc cistron and the fused second cistron by linking the in-frame GGC$_{1274-1276}$ codon-containing region (nt 1243-1396) to the 5′-end of Fluc AUG-ORF. The fragment with an annealed stable hairpin (HF2, 5′-CTA GCG GGG CGC GTG GTG GCG GCT GCA GCC GCC ACC ACG CGC CCC A-3′/HR2, 5′-AGC TTG GGG CGC GTG GTG GCG GCT GCA GCC GCC ACC ACG CGC CCC G-3′) was inserted into the *Nhe*I and *Hind*III sites of pRAF and pRnF to generate the bicistronic plasmids phRAF and phRnF with the 5′-stable hairpin of the first Rluc cistron. Another fragment with the same annealed stable hairpin (HF3, 5′-GGC CGC GGG GCG CGT GGT GGC GGC TGC AGC CGC CAC CAC GCG CCC CA-3′/HR3, 5′-AGC TTG GGG CGC GTG GTG GCG GCT GCA GCC GCC ACC ACG CGC CCC GC-3′) was inserted into the *Hind*III and *Not*I sites of pFL to generate the plasmid phFL, and the Rluc coding sequence amplified with the primer set (RlucF3, 5′-AAA GCT AGC ATG ACT TCG AAA GTT TAT GAT-3′/RlucR3, 5′-AAA AAG CTT TTA TTG TTC ATT TTT GAG AACT-3′) then inserted into the above phFL to generate the bicistronic plasmid pRhnF with the 3′-stable hairpin of the first Rluc cistron. The vicinity (1243-1396 bp) of the GGC$_{1274-1276}$ codon was amplified with the primer set (1243-*Not*I-F, shown above/1396-*Xba*I-R, shown above) and inserted into the *Not*I and *Xba*I sites of pRhnF to generate the bicistronic plasmid pRhAF with the 3′-stable hairpin of the first Rluc cistron and the fused second cistron by linking the in-frame GGC$_{1274-1276}$ codon-containing region (nt 1243-1396) to the 5′-end of Fluc AUG-ORF.

The AUG-ORFs for Rluc and Fluc were amplified with the primer sets (RlucF4, 5′-AAA AGA TCT ATG ACT TCG AAA GTT TAT GAT-3′/RlucR2, shown above; FlucF2, 5′-AAA AGA TCT ATG GAA GAC GCC AAA AAC ATA-3′/FlucR, shown above) and inserted into the *Bgl*II and *Not*I sites of pRAF and pFL to generate the eukaryotic CMV promoter-less plasmids pRAF-ΔCMV and pFL-ΔCMV, respectively. The IRES sequence of HCV was amplified with the primer set (HCVIF, 5′-AAA GCG GCC GCG CCA GCC CCC GAT

Figure 8 GC-richness of the downstream RNA secondary structures is essential for production of proteins initiated from the GGC$_{1274-1276}$ codon. **(A)** Schematic representation of the partial ACAT1 mRNA sequence (nt 1243-1786) and mutations introduced into the formed RNA secondary structures after sub-deletions of stem-loop II. Each sequence after sub-deletion of stem-loop II is listed without or with the mutated nucleotides (underlined), and the deleted regions (Δ1289-1339, Δ1289-1306, Δ1307-1324 and Δ1325-1339) are marked on top. Gray bar, partial ACAT1 mRNA sequence (ACAT1 1243-1786); black bar, 3×Flag coding sequence (3×Flag); filled circle, GGC$_{1274-1276}$ initiation codon; hollow circle, AUG$_{1397-1399}$ initiation codon. **(B)** Predicted RNA secondary structures with introduced mutations depicted in (**A**). The folding Gibbs free energy (ΔG) of predicted RNA secondary structure is listed beneath each structure. **(C)** The expression plasmids depicted in (**A**) were transiently transfected into AC29, lysates were prepared and immunoblotting was carried out with an anti-ACAT1 antibody (DM10). Arrows indicate the positions of ACAT1-NT-Flag proteins initiated from GGC$_{1274-1276}$ and AUG$_{1397-1399}$, respectively. The experiments were repeated three times with similar results. **(D)** RT-qPCR analysis of total RNA from the cells transfected with the plasmids depicted in (**A**) was performed according to procedures described in the Materials and Methods. The ACAT1 mRNA levels were normalized to the GAPDH mRNA levels for each sample and the relative mRNA of the cells transfected with pNTF-D6 was designated as the control (1.0). The data represent mean ± SD from three independent experiments.

TGG GG-3′/HCVIR, 5′-AAA TCT AGA GGT TTT TCT TTG AGG TTT AGG A-3′) from the plasmid pC1b (from Prof. Akio Nomoto) and inserted into the *Not*I and *Xba*I sites of pRAF to generate the eukaryotic CMV promoter-containing plasmid pRHIF.

The partial ACAT1 cDNA K1 sequences (ACAT1 1243-1786) in the plasmids pNTF-D1 to -D4 were inserted into the *Kpn*I and *Xba*I sites of ph3Flag to generate expression plasmids phNTF-D1 to -D4. The sequence (1243-1396 bp) surrounding the $GGC_{1274-1276}$ codon in pNTF-D2 and pNTF-D3 was amplified with the primer set (1243-*Not*I-F, shown above/1396-*Xba*I-R, shown above). The vicinity (1243-1396 bp) of the $GGC_{1274-1276}$ codon in pNTF-D1 and pNTF-D4 was amplified with the primer sets (1269-*Not*I-F, 5′-AAG CGG CCG CTC CAG GGC ACC CCG AAT-3′/1396-*Xba*I-R, shown above) and (1243-*Not*I-F, shown above/1396-*Xba*I-R, 5′-AAA TCT AGA CAT GGA AGC GGT CAC AGA-3′), respectively. These amplified fragments were inserted into the *Not*I and *Xba*I sites of pRAF to generate the expression plasmids pRAF-D1 to -D4.

The sub-deletions (Δ1289-1306, Δ1307-1324, Δ1325-1339) of the stem-loop II region (1289-1339 bp) were performed by two-step PCR with the above common external primer set (1243F/1786R) and individual internal primer sets (D6F/D6R, D7F/D7R and D8F/D8R, shown in supplementary information, Table S1). The amplified fragments were inserted into the *Kpn*I and *Xba*I sites of p3Flag to generate expression plasmids pNTF-D6 to -D8, respectively.

The mutations were introduced into stem-loop I by one-step PCR with individual forward primers M1F to M8F with mutated nucleotides shown in Supplementary information, Table S2, and the common reverse primer 1786R. By using vector p3Flag, the expression plasmids pNTF-M1 to -M8 containing the desired mutations were constructed by the same method described above.

The mutations were introduced into the regions of RNA secondary structures from plasmid pNTF-D6 to -D8 also by two-step PCR with the above common external primer set (1243F/1786R) and individual internal primer sets (M9F/M9R and M10F/M10R for pNTF-D5, M11F/M11R and M12F/M12R for pNTF-D6, M13F/M13R and M14F/M14R for pNTF-D7, shown in supplementary information, Table S3). The amplified fragments were inserted into the *Kpn*I and *Xba*I sites of p3Flag to generate expression plasmids pNTF-M9 to -M14, respectively.

Restriction enzyme digestion and DNA sequencing confirmed all the constructed plasmids.

Preparation of protein samples and western blot analysis

Cells were washed twice, scraped in ice-cold PBS and extracted on ice for 30min in RIPA buffer containing protease inhibitor mixture (Sigma). Cell debris was removed by a spin at 16 000×*g*. Protein concentrations were determined using the BCA protein assay kit (Bio-Rad). For western blot analysis, 50 μg of total protein per lane was resolved by SDS-PAGE. The filters were probed with an ACAT1 antibody DM10 [34], anti-Flag (M2, Sigma), anti-Fluc (ab7358, Abcam) or anti-Rluc (MAB4400, Chemicon) antibody. After incubation with horseradish-conjugated secondary antibody, the signals were developed using ECL Western Blotting detection reagent (Pierce).

Prediction of RNA secondary structures

All the predicted secondary structures with folding Gibbs free energy (Δ*G*) in this study were determined using the mFold program (version 3.2) [35].

RT-PCR and -quantitative PCR (RT-qPCR) analysis

Total RNA isolated by Trizol (Invitrogen) from transfected cells was pretreated with RQ DNase I (Promega) and reverse-trancribed using oligo (dT)$_{18}$.

Two sets of primers were chosen for PCR to check the aberrant splicing site: F1 (5′-AAA GGA AAC GGA TGA TAA CTG GTC-3′)/F2 (5′-TCA AAT CGT TCG TTG AGC GAG T-3′), complementary to the 5′ and 3′ regions of Rluc gene, and R1 (5′-TCT CTT CAT AGC CTT ATG CAG TTG C-3′)/R2 (5′-CTT GCG TCG AGT TTT CCG GTA A-3′), complementary to the 5′ and 3′ regions of Fluc gene, respectively.

The quantification of ACAT1 transcripts was done by qPCR, using Brilliant SYBR Green qPCR Master Mix and a Light Cycler apparatus (Stratagene). The primer sets for GAPDH (glyceraldehyde-3-phosphate dehydrogenase gene) and ACAT1 transcripts were GAPDHF (5′-TGG CTA CAG CAA CAG AGT GG-3′)/GAPDHR (5′-GGG GTT ATT GGA CAG GGA CT-3′) and ACAT1F (5′-CGG GCC TCA GAC AAT ACA AT-3′)/ACAT1R (5′-TCA ATT CCT CTG CCT CTG CT-3′). The human ACAT1 mRNA levels were normalized to the human GAPDH mRNA in each sample.

Other methods

Standard molecular biology techniques were performed according to the methods described by Sambrook *et al*. [36].

References

1 Chang TY, Chang CC, Cheng D. Acyl-coenzyme A:cholesterol acyltransferase. *Annu Rev Biochem* 1997; **66**:613-638.

2 Yu C, Chen J, Lin S, *et al.* Human acyl-CoA:cholesterol acyl-transferase-1 is a homotetrameric enzyme in intact cells and *in vitro*. *J Biol Chem* 1999; **274**:36139-36145.

3 Rudel LL, Lee RG, Cockman TL. Acyl coenzyme A: cholesterol acyltransferase types 1 and 2: structure and function in atherosclerosis. *Curr Opin Lipidol* 2001; **12**:121-127.

4 Chang TY, Chang CC, Lin S, *et al.* Roles of acyl-coenzyme A: cholesterol acyltransferase-1 and -2. *Curr Opin Lipidol* 2001; **12**:289-296.

5 Song BL, Wang CH, Yao XM, *et al.* Human acyl-CoA:cholesterol acyltransferase 2 gene expression in intestinal Caco-2 cells and in hepatocellular carcinoma. *Biochem J* 2006; **394**:617-626.

6 Yang JB, Duan ZJ, Yao W, *et al.* Synergistic transcriptional activation of human Acyl-coenzyme A: cholesterol acyltransferase-1

gene by interferon-gamma and all-trans-retinoic acid THP-1 cells. *J Biol Chem* 2001; **276**:20989-20998.

7 Yang L, Yang JB, Chen J, *et al.* Enhancement of human ACAT1 gene expression to promote the macrophage-derived foam cell formation by dexamethasone. *Cell Res* 2004; **14**:315-323.

8 Chang C, Dong R, Miyazaki A, *et al.* Human acyl-CoA:cholesterol acyltransferase (ACAT) and its potential as a target for pharmaceutical intervention against atherosclerosis. *Acta Biochim Biophys Sin* 2006; **38**:151-156.

9 Chang CC, Huh HY, Cadigan KM, Chang TY. Molecular cloning and functional expression of human acyl-coenzyme A:cholesterol acyltransferase cDNA in mutant Chinese hamster ovary cells. *J Biol Chem* 1993; **268**:20747-20755.

10 Li BL, Li XL, Duan ZJ, *et al.* Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes. *J Biol Chem* 1999; **274**:11060-11071.

11 Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 2005; **361**:13-37.

12 Huez I, Creancier L, Audigier S, *et al.* Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol Cell Biol* 1998; **18**:6178-6190.

13 Sasaki J, Nakashima N. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc Natl Acad Sci USA* 2000; **97**:1512-1515.

14 Burge CB, Tuschl T, Sharp PA. *Splicing of precursors to mRNAs by the spliceosome*. New York: Cold Spring Harbor Laboratory Press, 1999.

15 Krainer AR. *Eukaryotic mRNA processing*. New York: Oxford University Press, 1997.

16 Yang L, Chen J, Chang CC, *et al.* A stable upstream stem-loop structure enhances selection of the first 5′-ORF-AUG as a main start codon for translation initiation of human ACAT1 mRNA. *Acta Biochim Biophys Sin* 2004; **36**:259-268.

17 Yang L, Lee O, Chen J, *et al.* Human acyl-coenzyme A:cholesterol acyltransferase 1 (acat1) sequences located in two different chromosomes (7 and 1) are required to produce a novel ACAT1 isoenzyme with additional sequence at the N terminus. *J Biol Chem* 2004; **279**:46253-46262.

18 Vagner S, Waysbort A, Marenda M, *et al.* Alternative translation initiation of the Moloney murine leukemia virus mRNA controlled by internal ribosome entry involving the p57/PTB splicing factor. *J Biol Chem* 1995; **270**:20376-20383.

19 Yang DQ, Halaby MJ, Zhang Y. The identification of an internal ribosomal entry site in the 5′-untranslated region of p53 mRNA provides a novel mechanism for the regulation of its translation following DNA damage. *Oncogene* 2006; **25**:4613-4619.

20 Dhar D, Roy S, Das S. Translational control of the interferon regulatory factor 2 mRNA by IRES element. *Nucleic Acids Res* 2007; **35**:5409-5421.

21 Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CU. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome bind-ing to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev* 1998; **12**:67-83.

22 Gebauer F, Hentze MW. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* 2004; **5**:827-835.

23 Rijnbrand R, Bredenbeek PJ, Haasnoot PC, *et al.* The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C virus and other flavivirus RNAs. *RNA* 2001; **7**:585-597.

24 Pestova TV, Hellen CU, Shatsky IN. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol* 1996; **16**:6859-6869.

25 Pestova TV, Shatsky IN, Hellen CU. Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol Cell Biol* 1996; **16**:6870-6878.

26 Hellen CU, Witherell GW, Schmid M, *et al.* A cytoplasmic 57-kDa protein that is required for translation of picornavirus RNA by internal ribosomal entry is identical to the nuclear pyrimidine tract-binding protein. *Proc Natl Acad Sci USA* 1993; **90**:7642-7646.

27 Walter BL, Nguyen JH, Ehrenfeld E, Semler BL. Differential utilization of poly(rC) binding protein 2 in translation directed by picornavirus IRES elements. *RNA* 1999; **5**:1570-1585.

28 Buratti E, Baralle FE. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* 2004; **24**:10505-10514.

29 Wu C, Alwine JC. Secondary structure as a functional feature in the downstream region of mammalian polyadenylation signals. *Mol Cell Biol* 2004; **24**:2789-2796.

30 Linnstaedt SD, Kasprzak WK, Shapiro BA, Casey JL. The role of a metastable RNA secondary structure in hepatitis delta virus genotype III RNA editing. *RNA* 2006; **12**:1521-1533.

31 Cadigan KM, Heider JG, Chang TY. Isolation and characterization of Chinese hamster ovary cell mutants deficient in acyl-coenzyme A:cholesterol acyltransferase activity. *J Biol Chem* 1988; **263**:274.

32 Higuchi R, Krummel B, Saiki RK. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* 1988; **16**:7351-7367.

33 Kozak M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* 1989; **9**:5134-5142.

34 Chang CC, Chen J, Thomas MA, *et al.* Regulation and immunolocalization of acyl-coenzyme A: cholesterol acyltransferase in mammalian cells as studied with specific antibodies. *J Biol Chem* 1995; **270**:29532-29540.

35 Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; **31**:3406-3415.

36 Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning*. 2nd Edition. New York: Cold Spring Harbor Laboratory Press, 1989.