Prediction of kinase-specific phosphorylation sites with sequence features by a log-odds ratio approach

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · FEBRUARY 2007

Impact Factor: 2.63 · DOI: 10.1002/prot.21563 · Source: PubMed

3 AUTHORS:

Tingting Li
Peking University Health Science Center
26 PUBLICATIONS **363** CITATIONS
SEE PROFILE

Fei Li
Zhejiang University
62 PUBLICATIONS **1,300** CITATIONS
SEE PROFILE

Xuegong Zhang
Tsinghua University
129 PUBLICATIONS **3,141** CITATIONS
SEE PROFILE

Alteration of Protein Subcellular Location and Domain Formation by Alternative Translational Initiation

Jun Cai,* Ying Huang, Fei Li, and Yanda Li
Institute of Bioinformatics, MOE Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University, Beijing, China

ABSTRACT **Alternative translation is an important cellular mechanism contributing to the generation of proteins and the diversity of protein functions. Instead of studying individual cases, we systematically analyzed the alteration of protein subcellular location and domain formation by alternative translational initiation in eukaryotes. The results revealed that 85.7% of alternative translation events generated biological diversity, attributed to different subcellular localizations and distinct domain contents in alternative isoforms. Analysis of isoelectric point values revealed that most N-terminal truncated isoforms significantly lowered their isoelectric point values targeted at different subcellular localizations, whereas they had conserved domain contents the same as the full-length isoforms. Furthermore, Fisher's exact test indicated that the two ways—targeting at different cellular compartments and changing domain contents—were negatively associated. The N-term truncated isoforms should have only one way to diversify their functions distinct from the full-length ones. The peculiar consequence of subcellular relocation as well as change of domain contents reflected the very high level of biological complexity as alternative usage of initiation codons. Proteins 2006;62:793–799.** © 2005 Wiley-Liss, Inc.

Key words: alternative translation initiation events; nucleotide context; subcellular relocalization; domain contents

INTRODUCTION

The mechanisms to generate protein diversity by alternative gene expression pathways allow an organism to increase its level of complexity. These mechanisms include alternative use of promoters, splice sites, and translation initiation codons. Specifically, the use of alternative translation initiation codons in a single mRNA, by expressing several proteins from a single gene, contributes to the generation of protein diversity. Wherein, the genes produce two or more versions of the encoded proteins, and the shorter version, initiated from a downstream in-frame start codon, lacks the N-terminal amino acids of the full-length isoform version.[1]

Since the phenomenon of alternative translation initiation was first discovered, a small, yet growing, number of mRNAs initiating translation from alternative start codons have been discovered.[2–10] Possible mechanisms enabling initiation events at alternative start codons in these mRNAs were reviewed.[11,12] According to the ribosome-scanning model for translation, protein synthesis in eukaryotes involves a process of 40S ribosomal subunit attachment to the 5′ end of the mRNA, followed by ATP-dependent movement down the mRNA until an initiation codon is reached.[13] Usually, but not always, this initiation site is the first Met codon. If the first initiation site in an mRNA is used inefficiently, the 40S ribosomal subunits will move through the site by a "leaky scanning" process and may initiate translation at an alternative downstream position.[14] This leaky scanning mechanism produces protein isoforms with alternative amino initiation sites as well as the mechanism of internal entry of the 40S ribosomal subunit.[1] Moreover, non-ATG codons such as CTG (Leu), ACG (Thr), and GTG (Val) can also serve as alternative initiation sites efficiently as well as ATG (Met).[1,13,15–18] In these cases, however, the initial amino acid that is incorporated appears to be methionine, apparently because these non-ATG codons can still form weak base pairs with the initiator Met-tRNAMet.[18,19]

The first evidence that alternative initiation of translation might have a crucial role in cell fate came from studies of chromosomal relocations in tumor-derived cell lines.[21,22] Various studies began to pave the way for this new field in gene expression: the biological significance of the use of alternative initiation. Experimental studies of individual genes revealed cases of functional differences (attributed to insertions, deletions, or substitutions of functional protein domains) between alternative initiated isoforms and also showed that selection of alternative translation initiation codons might be used to control the cellular localization of the isoforms, as demonstrated for the FGF2,

Abbreviations: ATIEs, alternative translational initiation events; MW, molecular weight; pI, isoelectric point; NLS, nuclear localization signal.

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: National Natural Science (China); Grant numbers: 60171038/F010605, 60234020/F03; Grant sponsor: Tsinghua University.

*Correspondence to: Jun Cai, Institute of Bioinformatics, MOE Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University, Beijing, 100084, China. E-mail: caijun99@mails.tsinghua.edu.cn

Received 14 April 2005; Revised 5 August 2005; Accepted 2 September 2005

Published online 9 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20785

hck, and VEGF genes.[8,20–26] Recent analysis on subcellular localizations in *Arabidopsis thaliana* genes indicated translational polymorphism as a potential source of plant proteins variety.[27]

Actually, the process of alternative initiation of translation raised several interesting questions that were partially solved during the 20 y that followed its discovery. These questions include what the biological significance of such a process is and how the isoforms bear distinct functions.[1,11,12] Therefore, it is important and interesting to study alternative translation initiation events (ATIEs) in eukaryotes. Instead of studying individual cases, we collected almost all the known ATIEs in this study. Then we systematically analyzed the characteristics of ATIEs and the impact on biological diversification from the viewpoint of the alteration of domain contents and subcellular localizations.

MATERIALS AND METHODS

ATIEs Materials

We strictly extracted the known ATIEs in eukaryotes from two ways: ATIE records from public protein databases, such as SWISS-PROT database, released in February of 2003[28]; and literature reporting ATIEs. The genes in these ATIEs cover many species including *Homo sapiens*, *Mus musculus*, *Bos taurus*, and *Saccharomyces cerevisiae*. For genes with the same name but from different species, only one of them was included. With the above stringent criteria, 140 ATIEs (301 alternatively translated variants with complete amino acid sequences available) in total form the major original material of this study. Of them, 75 ATIEs belong to *H. sapiens*. A list of the ATIE materials is presented as supplementary information.

Cellular Compartments of Isoforms

Proteins can target at different subcellular compartments in cells. These compartments include nucleus, mitochondria, cytoplasm, cytoskeleton, vesicle of secretory system, endoplasmic reticulum (ER), Golgi, vacuole, plasma membrane, peroxisome, and extracellular space. The cellular compartments of most isoforms in our study were extracted from SWISS-PROT database annotations and the literature. The subcellular localizations of the remainder were evaluated by widely accepted PSORT software, which was implemented by Web server http://psort.nibb.ac.jp/.[29,30] The PSORT is a computer program for the prediction of protein localization sites in cells. It receives the information of an amino acid sequence as inputs. Then, it analyzes the input sequence by applying the stored rules for various sequence features. Finally, it reports the possibility for the input isoform to be localized at each candidate site with additional information. In this version of PSORT, we can analyze the subcellular localizations for *S. cerevisiae* (or plant) sequences as well as animal sequences. *S. cerevisiae* has a candidate localization site, vacuole, instead of lysosome in animals and plants have a chloroplast as an extra candidate.

Scanning Domains in Protein Isoforms

We searched for the domain content of a protein isoform from the feature annotation of the protein record in the protein sequence database of the NCBI Entrez system (http://www.ncbi.nlm.nih.gov). Other potential domains of this isoform sequence were additionally analyzed by family matches software Pfam 17.0 at http://pfam.wustl.edu.[31] Pfam version 17.0 (March, 2005) contains alignments and models of domain structures for 7868 protein families, based on the Swissprot 46.0 and SP-TrEMBL 29.0 protein sequence databases.

RESULTS

Nucleotide Context of the Translational Starting Sites

Traditionally, Kozak's[32–34] theory indicates that initiation only proceeds efficiently if the starting codon occurs within the Kozak consensus context [GCC(A/G)CC***G] in which the most important features are the G at downstream position +2 adjacent to the starting codon and the A/G at upstream position −4. Herein, the translational starting codon consisting of three nucleotide acids is denoted as *** and the nucleotide position with an underline is marked as position 0. Because the first initiation site and the alternative in-frame starting codon in ATIEs are both efficient to encode protein isoforms, we list the nucleotide context around these translational starting sites. The nucleotide context logos ranging from upstream nucleotide position −13 to downstream position +13 adjacent to the starting codon are shown in Figure 1. The nucleotide contexts of 140 first starting codons encoding full-length isoforms and 161 downstream in-frame alternative starting codons initiating N-term truncated isoforms are considered in subplot A and subplot B, respectively. For a case-control study, we randomly selected 1,000 nonhomological genes of *H. sapiens* that are not subject to ATIEs. The nucleotide contexts of initiation sites, in-frame ATG sequences, and out-of-frame ATG sequences in these 1,000 genes are also calculated in subplots C, D and E, respectively. Both the initiation sites that are not subject to ATIEs and the natural alternative initiation codons (shown as subplots A, B, and C in Fig. 1) retain nucleotide context features that distinguish them from in-frame or out-of-frame ATG sequences that are not subject to ATIEs (shown as subplots D and E in Fig. 1). However, Kozak consensus context [GCC(A/G)CC***G] is less conserved surrounding the first or alternative initiation sites that are subject to ATIEs than the sites that are not subject to ATIEs. There are many exceptional nucleotide contexts. Indeed, many alternative initiation codons are located within segments not conforming properly to the Kozak consensus sequences. Especially, the feature G at downstream position +2 occurs in the contexts of 56 first starting codons and 80 downstream in-frame alternative starting codons. There are 44 first starting codons and 42 internal alternative starting codons without the feature A/G at upstream position −4.

Statistics on Molecular Weights (MWs) and Isoelectric Point Values

The MWs and isoelectric point (pI) values are two important protein features. We calculated these features of the isoforms in ATIEs to quantify the effect of domain

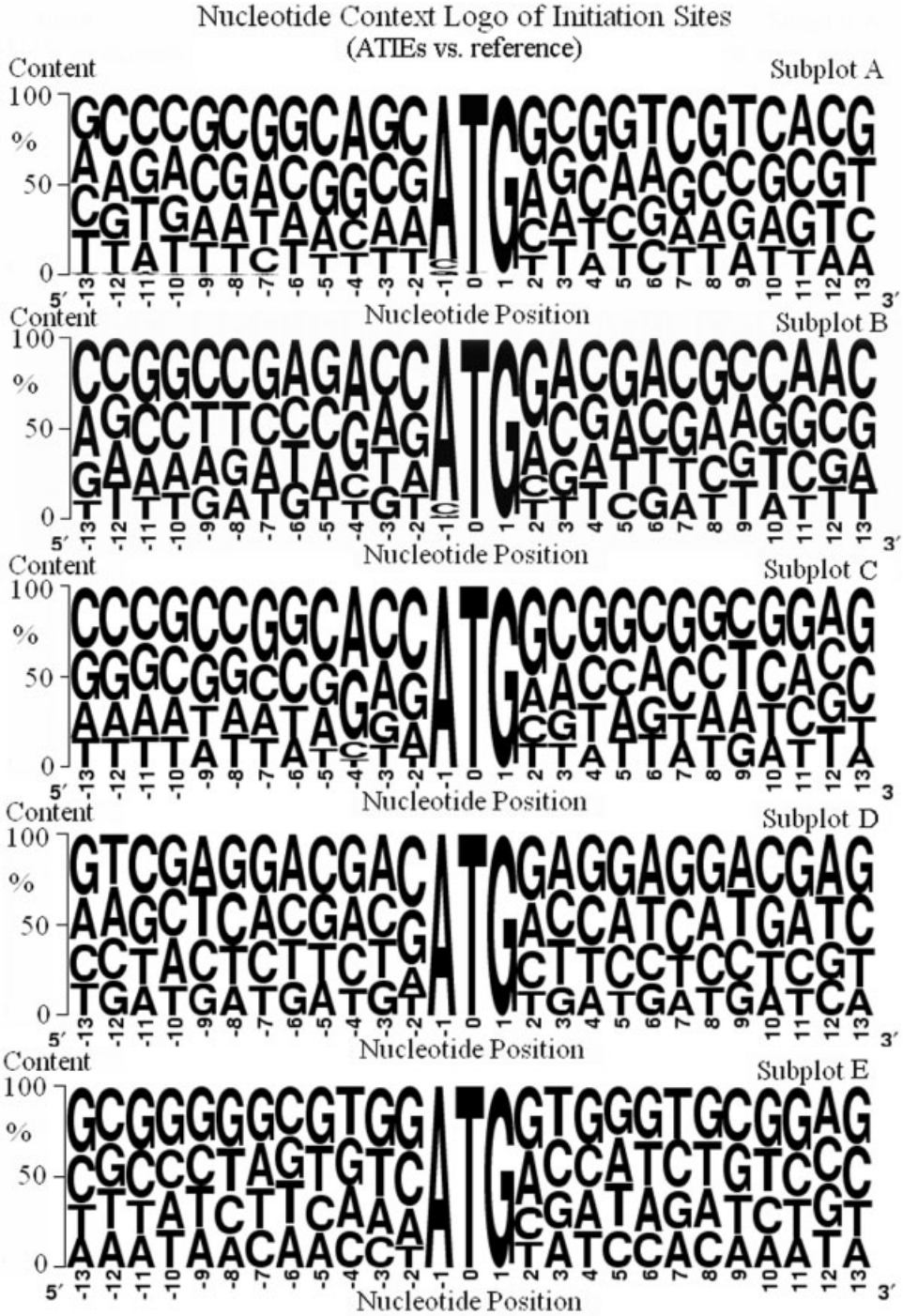

Fig. 1. The nucleotide context logos. The figure shows the context logos of first initiation sites (subplot A) and the downstream in-frame alternative initiation sites (subplot B) in ATIEs. Also, for a case-control study, the logos of initiation sites, in-frame ATG sequences, and out-of-frame ATG sequences that are not subject to ATIEs are listed in subplots C, D, and E, respectively. The nucleotide context ranges from upstream nucleotide position −13 to downstream position +13 adjacent to the ATG site.

contents and subcellular compartments on biological diversity.[35] Statistical plots on theoretical pI values and MWs of N-term truncated isoforms versus full-length isoforms are shown in Figure 2. According to the alteration of domain content or subcellular localization, four variants may be further considered: (I) alternative isoforms both possess different domain contents and target at different compartments; (II) alternative isoforms only change their domain contents; (III) isoforms with conserved domain contents target at different subcellular localizations; (IV) there is no change of domain contents or subcellular localizations for alternative isoforms. These four variants are represented

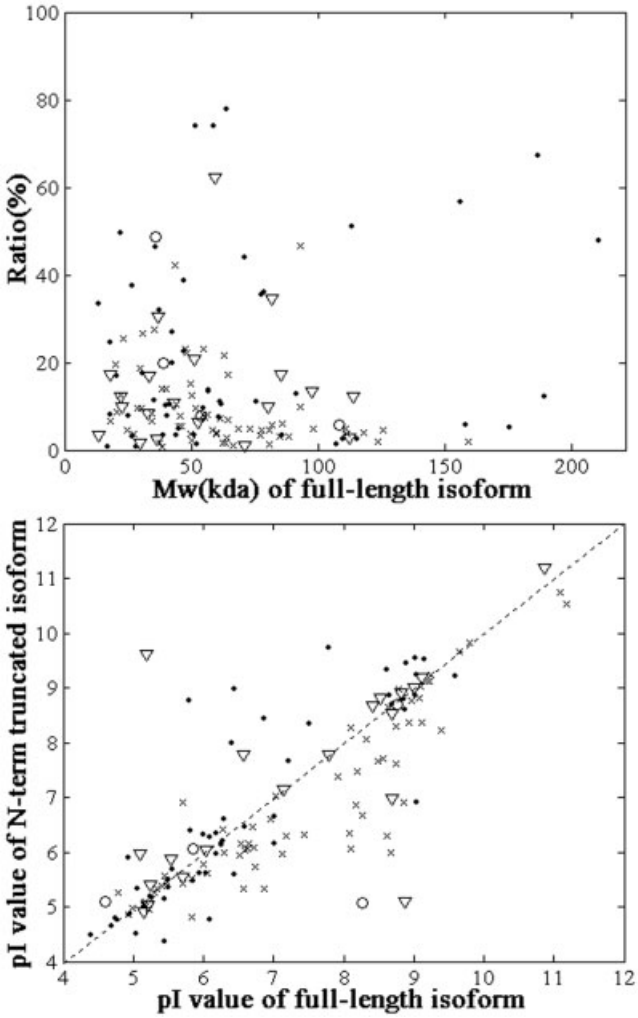

Fig. 2. Statistics on pI values and MWs of alternative isoforms in ATIEs. The variable of the *y* axis in the upper plot denotes the ratio of MW of the truncated N-term amino acids fragment versus MW of full-length isoform in ATIEs. Its value is limited in the range of 1–79%. According to the alteration of domain content or subcellular localization, the ATIEs are classified into four variants: (I) alternative isoforms both possess different domain contents and target at different compartments; (II) alternative isoforms only change their domain contents; (III) isoforms with conserved domain contents target at different subcellular localizations; (IV) there is no change of domain contents or subcellular localizations for alternative isoforms. These four variants are represented as the symbols ○, ●, ×, and ▽, respectively. kda, kilodalton.

as the symbols ○, ●, ×, and ▽ in Figure 2, respectively. The MWs of N-term truncated amino acids fragments range between 1 and 79% ratios of the MWs of full-length isoforms. The theoretical pI values of N-term truncated isoforms are also distinguished with those of the full-length ones. The N-term truncated isoform increases or lowers the pI to a new pH at which the protein has an equal number of positive and negative charges. And most N-term truncated ones significantly lower their pI values target at different subcellular localizations, whereas they have conserved domain contents the same as the full-length isoforms. As illustrated in the work of Schwartz et al.,[36] whole proteome pI values correlate with subcellular localizations of proteins. An explanation is that when the pI value changes, alteration of the localization in ATIEs occurs. As a result of lowering pI values, uncharged groups or amino acids containing additional positively charged moieties are truncated from the full-length protein molecule resulting in an increased net negative charge of the alternatively translated protein.

TABLE I. Subcellular Localizations and Domain Contents in ATIEs

ATIEs[a]	Affecting protein domain[b]	With conserved domains	Total ATIEs
Targeting at different subcellular localizations	4 ATIEs[c]	65 ATIEs	69
Targeting at the same cellular compartments	51 ATIEs	20 ATIEs	71
Total	55 ATIEs	85 ATIEs	140

[a]Fisher's test on the negative association between rows and columns: *P* value <0.0001.[36]
[b]Lack of domain structures implies that the alternative isoforms in ATIEs function in distinct ways in cell fate.
[c]These four ATIEs are CTNND1, RNPT, BAG1, and NRI12.

Alteration of Different Cellular Compartments and Domain Contents

Then, we systematically analyzed the impact of ATIEs on biological diversification from the viewpoint of the alteration of domain contents and subcellular localizations. With the alteration of domain content or subcellular localization, the ATIEs are divided into four categories and these four variants were calculated as a 2×2 contingency table (Table I).

Furthermore, two reference models, MOD1 and MOD2, were constructed to consider the randomness of alteration of domain structures and subcellular locations in ATIEs. The genes in model MOD1 are subject to ATIEs. We constructed this reference model MOD1 according to the following procedures: (a) truncate the N-term sequence fragment upstream of the randomly selected amino acid Met in each full-length isoform of the ATIEs. The ratio of the MW of the randomly truncated fragment versus the MW of the full-length isoform cannot be out of the range from 1 to 79% (refer to the statistics on MWs of the alternative variants in Fig. 2). (b) Cellular compartments and domain contents are scanned for these N-term truncated isoforms. (c) Recalculate the counts of four previous variants caused by the alteration of domain content or subcellular localization among full-length isoform and N-term truncated isoform pairs. We repeated the above steps 10 times, then a 2×2 contingency table with average counts of the four variants was formed (Table II). Later, a second random reference model MOD2 was constructed: (a) 140 nonhomologous non-ATIE genes are randomly selected from the SWISS-PROT database; (b) only the first and the second possible ATG codons in these genes are used for hypothetic translation initiation; (c) the same techniques are used to analyze cellular compartments and domain contents of these protein isoforms; (d) repeat the above steps 50 times and recount the ATIEs undergoing the alteration of domain content or subcellular

TABLE II. Subcellular Localizations and Domain Contents in Random Reference Set MOD1

Random events in reference model[a]	Affecting protein domain	With conserved domains	Total events
Targeting at different subcellular localizations	70 events[b]	24 events	94
Targeting at the same cellular compartments	35 events	11 events	46
Total	105 events	35 events	140

[a]The Fisher's exact test indicates the association between rows and columns is considered to be not statistically significant.
[b]Reference sets are randomly generated 10 times via strict procedures, and average counts are given.

TABLE III. Subcellular Localizations and Domain Contents in the Random Reference Set MOD2

Random events in reference model[a]	Affecting protein domain	With conserved domains	Total events
Targeting at different subcellular localizations	34 events[b]	10 events	44
Targeting at the same cellular compartments	23 events	73 events	96
Total	57 events	83 events	140

[a]The Fisher's exact test indicates that there is positive association between the variables.
[b]Reference sets are randomly generated 50 times via strict procedures, and average counts are given.

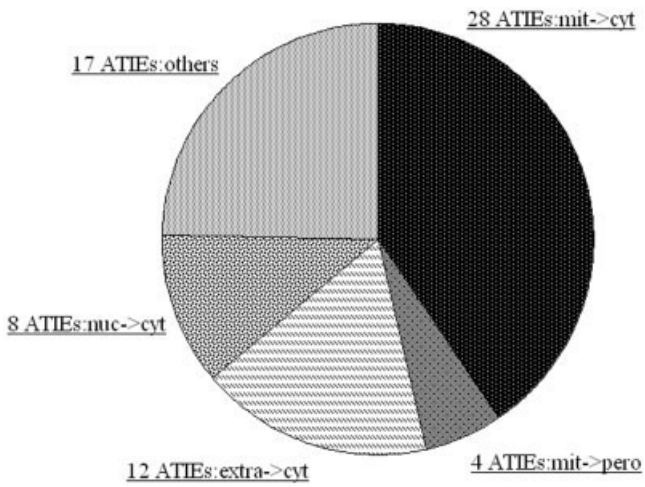

Fig. 3. Modes of subcellular relocations for protein isoforms in ATIEs. If the longer isoform in an ATIE targets at localization site A in the cell, and the shorter localizes at subcellular site B (different from A), we define that subcellular relocation from site A to B occurs in this ATIE (denoted as 1 ATIE: A→B). Signal A or B can be substituted by subcellular sites: cytoskeleton, cytoplasm, nucleus, mitochondria, vesicle of secretory system, ER, Golgi, vacuole, plasma membrane, peroxisome, and extracellular space including cell wall. We shortened these subcellular sites as cytos, cyt, nuc, mit, ves, er, gol, vac, pla, per, and ext, respectively.

localization. Thus, a 2×2 contingency table with average counts of model MOD2 was also formed (Table III).

According to our results, alternative isoforms in 69 ATIEs, which cover 49.3% of all 140 ATIEs, target at different cellular compartments. Moreover, subcellular relocations between the isoforms in these ATIEs are various. As shown in Figure 3, when the full-length isoforms target at mitochondria, the isoforms initiated from downstream in-frame start codons tend to change their subcellular localizations to cytoplasm or peroxisome because of lack of N′ terminal fragment. This kind of subcellular relocation frequently occurs. Second, some in-frame alternative initiation sites near the C terminal can truncate long-range amino acid fragments so as to result in lack of nuclear localization signals (NLS), which usually locate near the C terminal of sequence. Therefore, the subcellular relocation between nucleus and other compartments in ATIEs occurs. Taken together, different biological roles of alternative products in half of all these ATIEs attribute to the targets at different cellular compartments.

However, domain contents of proteins serve as important structural components to indicate the protein functions as well as subcellular localizations. The synthesis of alternative proteins having different domain contents potentially confers on these isoforms distinct functions.[22,26] In our scanning results, protein isoforms alternatively translated from an mRNA have been broken or deleted domains caused by N-term truncation in 55 ATIEs. They cover 39.3% ratio of all the ATIEs. One or more domain structures are deleted or disrupted because of truncation of N-term fragment caused by alternatively translational initiation. These domains include immunoglobulin cell adhesion molecule domain (IGcam) in tumor suppressor protein DCC, Enolase N-terminal domain (Enolase_N) in α-enolase protein, Alpha-2-macroglobulin family N-terminal domain (A2M_N) in complement C3 precursor.

In total, alternative isoforms in 120 of the 140 ATIEs reside in different cellular compartments or function in distinct domain contents. Detailed information can be found in the supplementary material. For the random model MOD2, although 57 and 44 samples are involved in alterations of domain structure and subcellular location, respectively, 73 (more than half) of the total 140 genes encode alternative isoforms with conserved domains and compartments. Compared with this case-control result of MOD2, the result that ATIEs reside in different cellular compartments or function in distinct domain contents is statistically significant. This good estimate of randomness indicates that the alternative initiation is biologically relevant.

Two Alternative Selected Ways to Diversify Functions of Isoforms

Also shown in Table I, almost all the ATIEs in which the synthesized proteins contain different domain contents encode protein isoforms targeting at the same subcellular sites. And the protein isoforms with the same domain structures in an ATIE are biased to target at various cellular compartments. Fisher's exact test was performed for Table I and both the two-tailed and the left-side P values are <0.0001.[37] This test effect can be explained in

two ways. First, the observation indicates that the negative association between targeting at different subcellular localizations and containing distinct domain structures for the alternatively translated isoforms is considered to be extremely statistically significant. Second, it simply reflects a negative correlation of subcellular relocations and changes of domain contents for the N-term truncated isoforms at randomly selected Met positions along the randomly selected full-length ones.

To discriminate between these, control tests on the two reference models, MOD1 and MOD2, were performed. Fisher' exact test used in Table II indicates that neither two-tailed nor one-side P value is <0.5. That is to say, the association between rows and columns in the Table II of model MOD1 is considered to be not statistically significant. Moreover, we did the Fisher's exact test on Table III derived from reference model MOD2 in order to offset the limitation that the control model MOD1 contained larger N-term deletions that would be possible to affect both compartment targets and domain structures of the truncated proteins. Both the two-tailed and the right-side P values are <0.0001 and this result indicates that there is positive association between the variables. Thus, the first possibility is clearly supported. That is to say, two alternative ways—targeting at different cellular compartments and changing domain contents—are negatively associated and chosen (but not both) by the N-term truncated isoforms of ATIEs to diversify the biology roles in cells.

DISCUSSION

Alternative translation initiation is an interesting biological process, which is also an important mechanism to improve the protein function diversity as proven by present statistical studies. By choosing a different translational initiation site, a new biological role of the protein is evolved by way of affecting protein subcellular localization or domain contents. As a mechanism for protein function diversity, ATIEs produce different protein isoforms which are either located in different subcellular sites or varied in domain formation. However, some problems must be further considered. Some computational methods were used in additional evaluation on potential domains and subcellular localizations, so the two premises should be considered in our results: Psort can be used to compare different cellular compartment targets of alternatively translated proteins accurately[29,30]; and the Pfam program can match potential protein domains correctly.[31] Otherwise, because of the small sample volume, our results just represented the situation described for ATIEs available from the current protein databank and literature, instead of the whole ATIEs for all the eukaryotic proteins. In the studies by Kochetov and Sarai,[27] a leaky scanning mechanism and Kozak consensus context were used to construct a new truncated isoform in order to overcome the problem of the sample volume of ATIEs. Conversely, our study indicated that many nucleotide contexts surrounding the alternative in-frame starting codon for Met did not conform properly to the Kozak consensus translation initiation (shown in Fig. 1). In the cases such as VEGF, the initiation at the ATG does not require a leaky scanning mechanism because it is driven by a ribosomal entry site.[7] Additionally, many starting codons in-frame with conserved Kozak consensus context cannot effectively initiate the translation of alternative isoforms.[1,38] Therefore, the method of constructing ATIE samples under Kozak consensus context leads to its limits that false-positive samples join into the dataset and true alternative initiation codons without Kozak consensus context are excluded. Therefore, we were cautious to collect all the true ATIE samples from the current protein databank and literature in our study, even if the ATIE volume seems to be small.

If the context of a proximal in-frame ATG is suboptimal, 40S ribosomal subunits can initiate translation at downstream in-frame ATG codons. It can result in a synthesis of protein forms possessing the same function, thereby increasing total translation efficiency as the Rx/rax homeobox gene does.[39] However, previous research also reported that the eyeless mouse mutation (ey1) removed an alternative start codon from the Rx/rax homeobox gene.[40] The redundant isoforms of the Rx/rax homeobox gene with reduplicate biological roles were mutated and eliminated by evolution in the process of the biology system. The isoforms in ATIEs that have little or no contribution to diversity of protein function are "meaningless" in evolution. No more than 20 samples of a total of 140 ATIEs encode the alternative isoforms with the same function to increase total translation efficiency. Under the natural assumption that mutational events leading to the emergence of multiple alternative translational variants are randomly distributed, the observed nonrandom ATIEs, tending to alter domain architecture or localization sites of proteins, suggest that positive Darwinian selection favors the events of alternative translation. Such positive selection is strong evidence of the importance of ATIEs in increasing functional diversity of proteomes.

CONCLUSIONS

The use of alternative translation initiation codons expresses several proteins from a single gene. ATIEs are attracting increasing attention because they greatly contribute to the gene function diversity as well as alternative splicing. Experimental studies of individual genes revealed cases of functional differences (attributed to insertions, deletions, or substitutions of functional protein domains) between alternative initiated isoforms and differences in the cellular localizations. Alternative translation tends to increase the diversity of gene functions by affecting domain contents or subcellular localizations of the proteins. Therefore, we collected all known ATIEs in eukaryotes from the current protein databank and literature. Then, we reported a statistical analysis of experimentally verified cases of translational polymorphism. Our results revealed that 120 of 140 (85.7%) ATIEs generated biological diversity, attributed to different subcellular localizations and distinct domain contents in alternative isoforms. In our studies, four variants were considered in these ATIEs: (I) alternative isoforms both possess different domain contents and target at different compartments; (II) alternative isoforms only change their domain contents; (III) isoforms with conserved domain contents target at different subcellular localizations; (IV) there

is no change of domain contents or subcellular localizations for alternative isoforms in ATIEs. According to our results, only four ATIEs caused both protein subcellular relocation and domain formation. Analysis on pI values revealed that most N-term truncated isoforms significantly lowered their pI values targeted at different subcellular localizations, whereas they had conserved domain contents the same as the full-length isoforms. Lastly, Fisher's exact test performed for these four variants indicated that the negative association between targeting at different subcellular localizations and containing distinct domain structures for the alternatively translated isoforms was considered to be extremely statistically significant. As our novel viewpoint of protein function diversity, ATIEs produced different protein isoforms which either target at different subcellular compartments or vary in domain formation. To explain this phenomenon, it was proposed that "meaningless" ATIEs would likely be negative-selected and thus eliminated in the evolution. Increasing functional diversity of proteomes became strong evidence of the importance of ATIEs.

ACKNOWLEDGMENTS

J. Cai and Y. Huang thank Tsinghua University for the Ph.D. Grant support.

REFERENCES

1. Touriol C, Bornes S, Bonnal S, et al. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biol Cell 2003;95:169–178.
2. Parker KL, Carroll MC, Shreffler DC, Capra JD. Identification of H-2-controlled structural variants of the murine Slp protein and demonstration of cis-regulation of its expression. J Immunol 1981;126:995–997.
3. Bruening W, Pelletier J. A non-AUG translational initiation event generates novel WT1 isoforms. J Biol Chem 1996;271:8646–8654.
4. Yazgan O, Pfarr CM. Differential binding of the menin tumor suppressor protein to JunD isoforms. Cancer Res 2001;61:916–920.
5. Becerra SP, Rose JA, Hardy M, Baroudy BM, Anderson CW. Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. Proc Natl Acad Sci USA 1985;82: 7919–7923.
6. Curran J, Kolakofsky D. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J 1988;7:245–251.
7. Huez I, Bornes S, Bresson D, Creancier L, Prats H. New vascular endothelial growth factor isoform generated by internal ribosome entry site-driven CUG translation initiation. Mol Endocrinol 2001;15:2197–2210.
8. Packham G, Brimmell M, Cleveland JL. Mammalian cells express two differently localized Bag-1 isoforms generated by alternative translation initiation. Biochem J 1997;328:807–813.
9. Prats AC, De Billy G, Wang P, Darlix JL. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. J Mol Biol 1989;205:363–372.
10. Vagner S, Gensac MC, Maret A, et al. Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. Mol Cell Biol 1995;15:35–44.
11. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. Gene 2002;299:1–34.
12. Kozak M. Alternative ways to think about mRNA sequences and proteins that appear to promote internal initiation of translation. Gene 2003;318:1–23.
13. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene 1999;234:187–208.
14. Kozak M. An analysis of vertebrate sequences: initiation of translational control. J Cell Biol 1991;115:887–903.
15. Acland P, Dixon M, Peters G, Dickson C. Subcellular fate of the int-2 oncoprotein is determined by choice of initiation codon. Nature 1990;343:662–665.
16. Hann SR. Regulation and function of non-AUG-initiated protooncogenes. Biochimie 1994;76:880–886.
17. Hann SR, King MW, Bentley DL, Anderson CW, Eisenman RN. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell 1998;52:185–195.
18. Peabody DS. Translation initiation at non-AUG triplets in mammalian cells. J Biol Chem 1989;164:5031–5035.
19. RajBhandary UL. More surprises in translation: initiation without the initiator tRNA. Proc Natl Acad Sci USA 2000;97:1325–1327.
20. Hann SR, Dixit M, Sears RC, Sealy L. The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA-binding site. Genes Dev 1994;8:2441–2452.
21. Gandre S, Bercovich Z, Kahana C. Mitochondrial localization of antizyme is determined by context dependent alternative utilization of two AUG initiation codons. Mitochondrion 2003;2:245–256.
22. Perry ME, Mendrysa SM, Saucedo LJ, Tannous P, Holubar M. P76MDM2 inhibits the ability of p90MDM2 to destabilize p53. J Biol Chem 2000;275:5733–5738.
23. Lu NZ, Cidlowski JA. The origin and functions of multiple human glucocorticoid receptor isoforms. Ann NY Acad Sci 2004;1024:102–123.
24. Lock P, Ralph S, Stanley E, Boulet I, Ramsay R, Dunn AR. Two isoforms of murine hck, generated by utilization of alternative translational initiation codons, exhibit different patterns of subcellular localization. Mol Cell Biol 1991;11:4363–4370.
25. Luque CM, Lallena MJ, Alonso MA, Correas I. An alternative domain determines nuclear localization in multifunctional protein 4.1. J Biol Chem 1998;273:11643–11649.
26. Goss KH, Trzepacz C, Tuohy TMF, Groden J. Attenuated APC alleles produce functional protein from internal translational initiation. Proc Natl Acad Sci USA 2002;99:8161–8166.
27. Kochetov AV, Sarai A. Translational polymorphism as a potential source of plant proteins variety in *Arabidopsis thaliana*. Bioinformatics 2004;20:445–447.
28. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 2000;28:45–48.
29. Nakai K, Kanehisa M. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 1992;14:897–911.
30. Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 1999;24:34–36.
31. Bateman A, Birney E, Cerruti L, et al. The Pfam protein families database. Nucleic Acids Res 2002;30:276–280.
32. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 1986;44:283–292.
33. Kozak M. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 1987;15:8125–8148.
34. Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotide in positions +5 and +6. EMBO J 1997;16:2482–2492.
35. Sillero A, Ribeiro JM. Isoelectric points of proteins: theoretical determination. Anal Biochem 1989;179:319–325.
36. Schwartz R, Ting CS, King J. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. Genome Res 2001;11:703–709.
37. Agresti A. A survey of exact inference for contingency tables. Stat Sci 1992;7:131–153.
38. Tailor CS, Marin M, Nouri A, Kavanaugh MP, Kabat D. Truncated forms of the dual function human ASCT2 neutral amino acid transporter/retroviral receptor are translationally initiated at multiple alternative CUG and GUG codons. J Biol Chem 2001;276: 27221–27230.
39. Kochetov AV. AUG codons at the beginning of protein coding sequences are frequent in eukaryotic mRNAs with a suboptimal start codon context. Bioinformatics 2005;21:837–840.
40. Tucker P, Laemle L, Munson A, et al. The eyeless mouse mutation (ey1) removes an alternative start codon from the Rx/rax homeobox gene. Genesis 2001;31:43–53.