ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Biotechnology 4 (4): 316-324, 2005
ISSN 1682-296X

A +1 Programmed Ribosomal Frameshifting of a Human C2H2 Zinc Finger Gene Discovered by Computational Analysis

Shaowu Meng, Zhuo Zhang and Limsoon Wong
Institute for Infocomm Research, 21 Heng Mui Keng Terrace, 119613, Singapore

Abstract: In this study, we discover a +1 programmed ribosomal frameshifting in a human C2H2 zinc finger (C2H2-ZNF) gene after testing total 226 C2H2-ZNF genes in Human Chromosome 19 through using an original combination of standard computational tools. Our evidences are as follows: Firstly, this zinc finger gene, denoted Z3-7, has significant C2H2-ZNF domains (e-value<0.05) in its two ORFs in different reading frames. Secondly, the two ORFs are overlapped, with a common promoter, a common transcription start site, but with different start codons, different Kozak patterns, different PolyA sites and different PolyA signals. Thirdly, in the mRNA of this gene, we found a significant pseudoknot and a likely frameshifting sites UUUCCU, five upstream nucleotides before this pseudoknot; Moreover, each of the two ORFs of Z3-7 significantly matches two reading frames of seven human C2H2-ZNF ESTs (e-value<10^{-20}) and of three human C2H2-ZNF cDNAs (e-value<10^{-30}) that correspond to human C2H2-ZNF proteins. More importantly, each of the two ORFs of Z3-7 significantly matches human C2H2-ZNF proteins (e-value<10^{-30}). These facts indicate that each of the two ORFs of Z3-7 could be transcribed and translated *in vivo* and that the ribosomal frameshifting of Z3-7 is likely efficient. The present discovery would be helpful for deeper understanding of the regulatory mechanism of a gene with tandem repetitive domains and should have potential to understand and cure diseases caused by abnormality in such transcription factors as C2H2-ZNF proteins.

Key words: +1 programmed ribosomal frameshifting, human C2H2 zinc finger gene, pseudoknot, frameshifting site, significant match with human C2H2-ZNF ESTs, cDNAs, proteins

INTRODUCTION

A frameshift mutation is a mutation by which some nucleotides (not divisible by three) are inserted in or deleted from a gene's nucleotide sequence. Such an insertion or deletion shifts the ORF in the nucleotide sequence; hence the gene cannot be usually translated into a functional protein[1]. However, in viruses and transposons, some mRNAs could be translated into a new fused protein after a frameshift modification[2]. This phenomenon is called programmed ribosomal frameshifting.

So, more specifically, a programmed ribosomal frameshifting is that a ribosome shifts a reading frame at a particular mRNA site to yield a new fused protein encoded by two overlapped open reading frames[3]. Normally, a programmed ribosomal frameshifting is a mechanism to produce two functional proteins from one mRNA[4]. In some cases, a programmed ribosomal frameshifting also functions as a regulatory role in metabolic pathways[4]. In general, a gene in which a programmed ribosomal frameshifting occurs has following characteristics: (a) the two overlapping ORFs sit in different reading frames; (b) a stretch of seven nucleotides, within mRNA of this gene, functions as a frameshifting site (or called slippery site), which is often stimulated by a downstream secondary structure, such as a pseudoknot[5,6]. So, the identification of overlapping ORFs, frameshifting sites and pseudoknots for a gene is our primary focus in this study.

The known cases about the ribosomal frameshifting are either +1 programmed ribosomal frameshifting or -1 programmed ribosomal frameshifting, in which a ribosome reads into the Reading Frame +1 or the Reading Frame -1 by slipping one base toward the 3′ direction or toward the 5′ direction, respectively[6-8]. In -1 programmed ribosomal frameshifting, the frameshifting site is generalized as X XXY YYZ. Here, X, Y and Z represent any nucleotide. As for +1 programmed ribosomal frameshifting, the slippery site is modeled as UUU Ynn pattern. Here, Y is pyrimidine,

Corresponding Author: Dr. Shaowu Meng, Department of Biology, University of Toronto at Mississauga, 3359 Mississauga Road North, Missisauga, Ontario L5L 1C6, Canada
Tel: +1 905 569 4400 Fax: +1 905 828 3792

such as cytosine (C), thymine (T) and uracil (U); n represents any nucleotide. The UUU Ynn pattern has been well studied and confirmed as highly frameshift-prone site for +1 programmed ribosomal frameshifting in both eubacteria and mammalian cells[4,9,10]. Besides, the pattern CUU-ANN-N (with N representing any nucleotide), identified by a computational method, was also used to identify +1 ribosomal slippery sites in yeast[11].

Usually, a programmed ribosomal frameshifting occurs in viruses, but rarely in prokaryotic and eukaryotic cellular organisms[8,12]. A case of -1 programmed ribosomal frameshifting in the Edr gene of mouse was reported by Shigemoto *et al.*[12], which is the first example of -1 translational recoding in a eukaryotic organism. Through the use of DNA databases, Shigemoto *et al.*[12] again found that a human cDNA clone KIAA1051 has a >80% identity with Edr and that the -1 slippery sequence are highly conserved between KIAA1051 and Edr. As a consequence, they concluded that KIAA1051 is the human orthologue of Edr. The phenomenon of +1 programmed ribosomal frameshifting was reported in antizyme genes ranging from yeast to human[13-16]. This antizyme gene encodes ornithine decarboxylase (ODC) antizymes for mammalian including human[14-16]. To date, this gene family encoding antizymes 1, 2 and 3 is the only known gene family in human genome that has programmed ribosomal frameshifting[16].

In one of earlier research[17], we have reported that there are 12 C2H2 zinc-finger genes on human chromosome 19 that can be each translated into the same type of protein after frameshifts. We also pointed out that programmed ribosomal frameshifting may cause this phenomenon in two of the 12 genes[17]. To continue this important research, we carefully study the mechanism of programmed ribosomal frameshifting for the C2H2 zinc finger (C2H2-ZNF) gene family in human chromosome 19. A significant result is that we identified a +1 programmed ribosomal frameshifting occurring in one of those genes.

Our method can be simply extended to search the whole human genome. We choose C2H2-ZNF genes to start up because of the following reasons: C2H2-ZNF genes make up one of the largest human gene families with over 700 repetitive members in the human genome[18] and a C2H2-ZNF protein also has many repetitive C2H2-ZNF domains in a tandem array[23]. A sequence pattern of a C2H2-ZNF domain is #-X-C-X(1-5)-C-X3-#-X5-#-X2-H-X(3-6)-H/C[20]. Here, X represents any amino acid, # denotes hydrophobic residues and the numbers indicate the numbers of residues. H/C says this position's amino acid is either histidines or cysteines; C2H2-ZNF proteins usually play a critical role as transcriptional regulators by mediating interactions between DNAs and proteins[21]. Moreover, the abnormal expression of C2H2 zinc finger genes, such as ZNF^{EB}, associates with the oncogenic transformation of host cells[22]. So, this study will be helpful in the deeper understanding of the regulatory mechanism of a gene with tandem repetitive domains and have potential to understand and cure a disease that is caused by abnormality in transcription factors, such as C2H2-ZNF proteins.

MATERIALS AND METHODS

The gene sequences that we used in this study are from a public website http://bahama.jgi-psf.org/pub/ch19/. From this website, we downloaded a total 226 C2H2-ZNF genes of human chromosome 19. These genes were originally identified by Dehal *et al.*[23] using Genscan and Genewise programs. For each of the 226 genes, we will test whether it has a programmed ribosomal frameshifting.

Our method is a combination of some existing algorithms that have been implemented by peer colleagues. This combination is new and it is also one of our contributions in this study. Our method consists of three sequential processes: discovery of overlapping ORFs from different reading frames, structural analysis and functional analysis. The discovery of overlapping ORFs proceeds in three steps: Step A1, Step A2 and Step B (see below). The structural analysis includes Step C for analyzing the primary structure of a gene and Steps D and E for analyzing the secondary structure of a putative mRNA of the gene. Steps F1, F2 and F3 are used to detect whether these ORFs of the gene are functional. We describe each of these steps in detail.

Discovery of overlapping ORFs from different reading frames:

- **Step A1:** This is a step to find out ORFs from different reading frames of a gene, that is, from the Reading Frames 0, +1, +2, -0, -1 and -2. Note that Reading Frame 0 is the original sequence, Reading Frame +1 is the sequence after deleting one nucleotide from Reading Frame 0, and Reading Frame +2 is the sequence after deleting two nucleotides from Reading Frame 0. The sign - indicates the reverse and complementary sequence of a given sequence. We use ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to detect the ORF in each reading frame.

- **Step A2:** This step is used to detect C2H2-ZNF domains in each reading frame. We use a program called Translate (http://tw.expasy.org/tools/dna.html) to translate the six reading frames of a gene into amino acid sequences; then we search Pfam (http://www.sanger.ac.uk/Software/Pfam/) using HMMER (http://hmmer.wustl.edu/) as searching tool and the six amino acid sequences as query sequences. The e-value cutoff level is 1.0 according to this program's default.

- **Step B:** This step operates on the outcome of Steps A1 and A2. It is used to find out long overlapping ORFs (with more than 23 amino acids) that sit in different reading frames and that also contain significant C2H2-ZNF domains. If such ORFs cannot be found, we move back to Step A1 to examine a different gene.

Structural analysis:

- Step C is used to conduct an analysis on the primary structure of a gene. We examine the 5′and 3′ non-coding regions of the long ORFs, which are derived from Step B, to find whether these ORFs each have a promoter, a transcription start site, a start codon, a Kozak pattern, a PolyA site and a PolyA signal. A promoter and a transcription start site are decided by using Dragon Promoter Finder in a public website http://research.i2r.a-star.edu.sg/promoter/DGE_main.htm[24], the sensitivity value is set according to this program's default; A start codon and a Kozak pattern are predicted by using ATG^{pr}[25] in a public website http://www.hri.co.jp/atgpr/; A PolyA site and a PolyA signal are found by using Hcpolya[26] in a public website http://l25.itba.mi.cnr.it/~webgene/wwwHC_polya.html. If such ORFs do not exist, we return to Step A1.
- Step D is used to analyze the secondary structure of RNAs, aiming to find a pseudoknot in a putative mRNA of this gene. We use GeneBee-NET[27] to perform this task. This tool analyzes the secondary structure of the first sequence in an alignment (other sequences in the alignment function as comparisons) and can find out significant pseudoknots[27]. Following the requirement of this tool, we produce two distinct mRNAs by deleting introns (the sequences between domains) separately from two long ORFs (from Step B) in different reading frames of a gene[28] and then use the same tool (http://www.genebee.msu.su/services/rna2_reduced.html) to analyze the secondary structure of the mRNA in Reading Frame 0 (original sequence). If no pseudoknot is found, we jump back to Step A1 to examine next gene.
- Step E follows Step D sequentially if a pseudoknot is found in Step D. The main purpose of this step is to analyze the upstream sequence of the pseudoknot to find a frameshifting site. If no frameshifting site exists, we return to Step A1.

Functional analysis:

- **Step F1:** This step is used to detect whether each of the two overlapping ORFs, which are derived from Step B, could match with different reading frames of Human C2H2-ZNF ESTs. If so, the two ORFs could be transcribed. First we extract a C2H2-ZNF EST dataset from Human EST database through using Step A2 to find out ESTs that have C2H2-ZNF domains in different reading frames. Then, we search against this dataset using TBLASTX (ftp://ftp.ncbi.nih.gov/blast/)[29] as searching tool and Z3-7 as query sequence. E-value cutoff is set at 10^{-20} that is much less than the default e-value 10. We keep those ESTs whose different reading frames match with Reading Frame 0 and Reading Frame +1 of Z3-7.

- **Step F2:** This step is to test whether each of the overlapping ORFs could match with different reading frames of cDNAs that are corresponding to real Human C2H2-ZNF gene or protein. We extract a Human C2H2-ZNF cDNAs dataset from Ensembl database[30] through using Step A2 to find out cDNAs that have C2H2-ZNF domains in different reading frames. Note that each cDNA in this cDNA dataset is corresponding to a real human C2H2-ZNF gene or protein according to Ensembl's records[30]. Then, we search against this cDNA dataset using TBLASTX as searching tool and Z3-7 as query sequence. E-value cutoff is set at 10^{-20} that is much less than the default e-value 10. We keep those cDNAs whose different reading frames match with Reading Frame 0 and Reading Frame +1 of Z3-7.

- **Step F3:** This step is used to test whether each of the overlapping ORFs could be translated. We search against SWISS-PROT using Wu-Blast (http://www.ebi.ac.uk/blast2/)[31] as searching tool and the amino acid sequences of the ORFs as query sequences. We record top 50 matched proteins.

If a gene meets all the conditions above, then this gene has a programmed ribosomal frameshifting.

RESULTS

Using this method, we tested every gene of the 226 C2H2-ZNF genes in human chromosome 19. We found one gene that has the +1 programmed ribosomal frameshift. We name this gene Z3-7 since it is the seventh gene in the third cluster of the C2H2-ZNF genes in human chromosome 19. Z3-7 has 2769 nucleotides and it is located at clone 21743738 spanned from positions 208015 to 210634.

Table 1 shows that Z3-7 has two ORFs that are in different reading frames but that contain the same C2H2-ZNF domains with significant e-values. Moreover, the two ORFs are overlapping (Fig. 1). Here, we denote the ORF in Reading Frame 0 as ORF1 and the ORF in Reading Frame +1 as ORF2 (Fig. 1).

We examined the 5′ and 3′ non-coding regions of the two ORFs. The primary structure of the two ORFs is as follows:

(I) The ORF1 spans from position 1465 to position 1869, with Kozak pattern AXXATGc, PolyA signal ACATGAAAAG and with transcription start site at position 1249, PolyA site at position 1961;
(ii) The ORF2 spans from position 1541 to position 2371, with Kozak pattern tXXATGa, PolyA signal CTAACAAATG and with transcription start site at position 1249, PolyA site at position 2424.

Observe that the two ORFs have different start codons, Kozak patterns, PolyA sites and PolyA signals. From Fig. 1, we can also see that the two ORFs share a promoter, a transcription start site. In light of these primary structural facts and based on the definition of a gene by Portin[32], we believe that each of the two ORFs of Z3-7 can be viewed as a gene.

We also carefully studied the secondary structure of a putative mRNA of Z3-7 and found a significant pseudoknot[27] with energy -2.000000 Kkal/mol (Fig. 2). This pseudoknot consists of two stems and two loops (Fig. 2). The 30 upstream nucleotides before the pseudoknot are CAAUGUGGGAAAGCCUUUGUU UCCUUCA CU. Among the 30 nucleotides, a sequence UUUCCU is likely to be a frameshifting site.

Seven human ESTs have C2H2-ZNF domains in two reading frames that significantly match with Reading Frame 0 and Reading Frame +1 of Z3-7. The e-values of most matches are much less than 10^{-20} (Table 2). The GenBank ID numbers of the seven ESTs are 1448736, 1740254, 1858722, 2555712, 3228157, 575744 and 6713901 and the likely frameshifting site TTTCCT (it is UUUCCU in mRNA) is found in 1858722, 2555712, 3228157, 575744 and 6713901 (Table 2 and Fig. 1). It indicates that each of the two ORFs can be transcribed. We also list the corresponding proteins of the seven ESTs according to the description in GenBank (http://www.ncbi.nlm.nih.gov/). As well, we found that Z3-7 match EST 2555712 with coverage being 98%, identity being 98% and e-value being zero when we use BLASTN as the searching tool. When we use TBLASTX as searching tool, the identity is much less (Fig.1), as TBLASTX translate nucleotide sequences into peptides and then compare them[24].

Table 1: C2H2-ZNF domains within two ORFs in different reading frames of Z3-7

Reading Frame	ORF	ORF-From	ORF-To	Score	E-value
0	ORF1	1513	1581	14	0.22
0	ORF1	1597	1665	19.7	0.0046
0	ORF1	1681	1746	13.7	0.23
0	ORF1	1762	1830	27.4	2.2e-05
+1	ORF2	2066	2134	25.9	6.0e-05
+1	ORF2	2150	2218	8.9	0.71
+1	ORF2	2234	2302	19	0.0075

Reading Frame 0 is the original sequence; Reading Frame +1 is the sequence after deleting one nucleotide from Reading Frame 0. ORF-From means the start position of an open reading frame; ORF-To indicates the end position

Table 2: Human C2H2-ZNF ESTs and cDNA hitting different Reading Frames of Z3-7

GI/Ensembl ID	Compared Items	RF 0 Z3-7	RF +1 Z3-7	PRFSite	Proteins
gi1448736	Matched RF	0	+1	No	P17097
	e-value	2.0e-48	7.0e-34		
gi1740254	Matched RF	+1	0	No	P15621
	e-value	1.0e-34	3.0e-21		
gi1858722	Matched RF	+2	0	Yes	ZNP
	e-value	2.0e-40	6.0e-36		
gi2555712	Matched RF	-0	-2	Yes	P52737
	e-value	0	0		
gi3228157	Matched RF	-1	-0	Yes	P15621
	e-value	1.0e-31	4.0e-31		
gi575744	Matched RF	0	+2	Yes	ZNP
	e-value	2.0e-24	1.0e-31		
gi6713901	Matched RF	-1	-0	Yes	P17039
	e-value	2.0e-29	1.0e-14		
ENSG00000164631	Matched RF	+2	0	No	P17014
	e-value	8.0e-69	2.0e-73		
ENSG00000167637	Matched RF	0	+2	No	Q14591
	e-value	0	2.0e-33		
ENSG00000171295	Matched RF	0	+2	Yes	Q9UC07
	e-value	7.0e-80	1.0e-76		

GI means GenBank identification number. Ensembl ID means Ensembl identification number. RF indicates reading frame. RF 0 is the original sequence; RF +1 is the sequence after deleting one nucleotide from RF 0; RF +2 is the sequence after deleting two nucleotides from RF 0; Mark- indicates the reverse and complementary sequence of a given sequence. PRFSite is the likely programmed ribosomal frameshifting site TTTCCT. Protein is corresponding C2H2-ZNF proteins of ESTs and cDNAs. ZNP is for zinc finger protein, i.e., we have not found specific proteins yet.

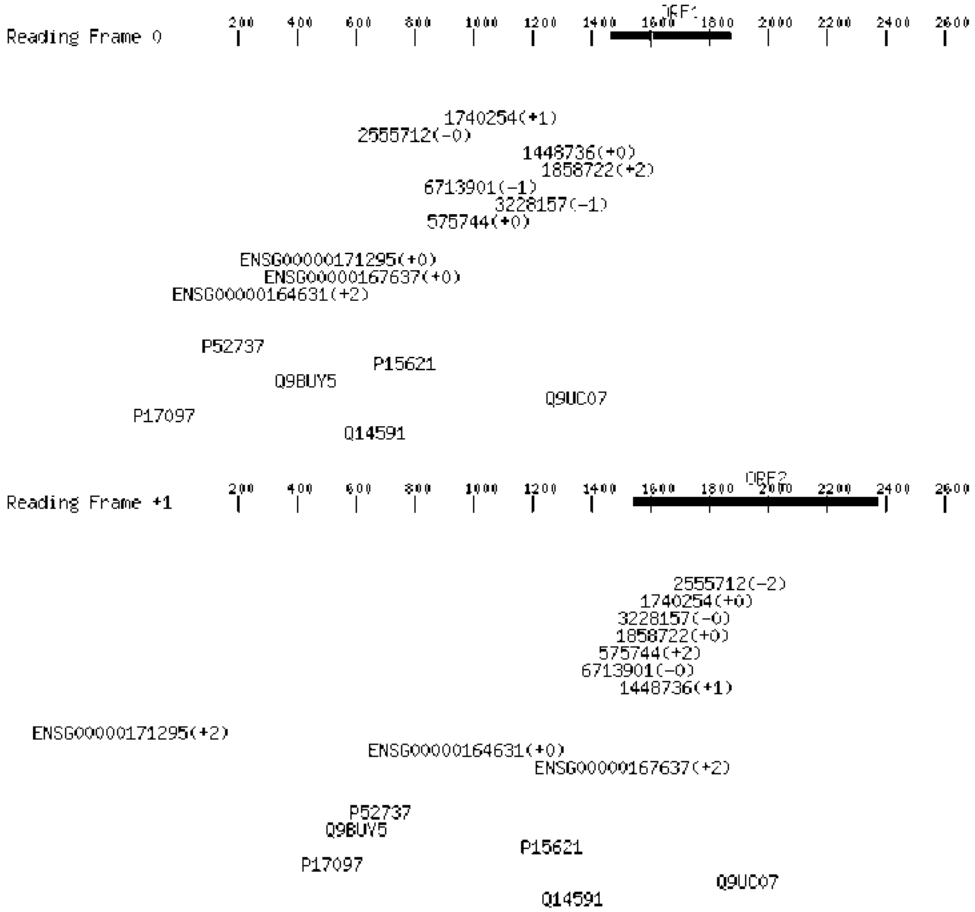

Fig. 1: ESTs, cDNAs and proteins hit two ORFs in different reading frames of gene Z3-7. Reading Frame 0 is the original DNA sequence; Reading Frame +1 is the sequence after deleting one nucleotide from Reading Frame 0. TSS indicates a transcription start site. PRFSite represents a likely programmed ribosomal frameshifting site whose sequence pattern is TTTCCT and spans from nucleotides 1601 to 1606. PolyA denotes a PolyA site. Red lines are the matched parts between ESTs and Z3-7; Pink lines are the matched parts between cDNAs and Z3-7 and between proteins and Z3-7; Yellow lines are those parts that do not match with Z3-7. A number in reading frames serves as a graduation. A number before an EST is its identification number in GenBank and the numbers in round brackets are the matched reading frames of ESTs; a mark before a cDNA is its identification number in Ensembl and the numbers in round brackets are the matched reading frames of cDNAs; a mark before a protein indicates its accession number to SWISSPROT. In this figure, proteins correspond to some of the ESTs and cDNAs.

We found that two reading frames in each of three cDNAs, ENSG00000164631, ENSG00000167637 and ENSG00000171295, can significantly match Reading Frame 0 and Reading Frame +1 of Z3-7 with e-value much less than 10^{-30}. Besides, the likely frameshifting site TTTCCT is found in ENSG00000171295 (Table 2). In Table 2, we list corresponding proteins of these three cDNAs. Furthermore, we found that the zinc finger proteins in SWISS-PROT gave hits of highly significant e-values (less than 10^{-30}) on each of the two ORFs (Fig. 1), after blasting the amino acid sequences of the two ORFs against SWISS-PROT. The two facts about cDNAs and proteins suggest that each of the two ORFs of Z3-7 might be translated.

In summary, we have observed that Z3-7 has two overlapping ORFs (in different reading frames), a pseudoknot and a likely frameshifting site near the pseudoknot. Moreover, each of the two ORFs is a gene and is possibly transcribed and translated. These characteristics of Z3-7 perfectly meet the requirements for a gene to have a ribosomal frameshifting[5,33]. And hence, it is reasonable to claim that Z3-7 is quiet likely under regulation by a +1 programmed ribosomal frameshifting.

As we have reported previously[17], in addition to Z3-7, another gene, named Z3-2, was also found to have a pseudoknot in its mRNA. However, we have not found a likely frameshifting site, either UUU Ynn or CUU-ANN-N pattern, near its pseudoknot in this study.

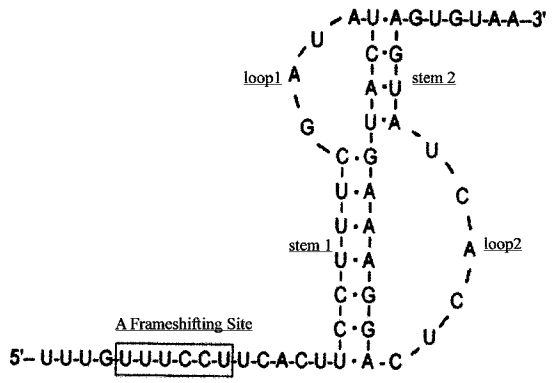

Fig. 2: The secondary structure of a pseudoknot in the mRNA of Z3-7. Elements of this pseudoknot are marked in this picture

Therefore, we do not think there is a +1 programmed ribosomal frameshifting in Z3-2 yet. So, we just put the results about Z3-2 as online supplementary in http://zincfinger.i2r.a-star.edu.sg.

DISCUSSION

The e-values here are much lower than their respective standards of significance: When searching Pfam, the default e-value cutoff is 1.0 and an e-value is considered to be significant if it is less than 0.05[20]. In Z3-7, two C2H2-ZNF domains in ORF1 of Reading Frame 0 have e-values 0.0046 and 2.2e-05; and two C2H2-ZNF domains in ORF2 of Reading Frame +1 have e-values 0.0075 and 6.0e-05. These four e-values are significant since they are less than 0.05; the other three C2H2-ZNF domains have a little bit bigger e-values, 0.22, 0.23 and 0.71, but yet have good scores (Table 1). The two ORFs of Z3-7, ORF1 and ORF2, match two reading frames of seven human C2H2-ZNF ESTs with significant e-values less than 10^{-20} and three C2H2-ZNF cDNAs with significant e-values less than 10^{-30} (The default e-value cutoff of TBLASTX is 10) and also match human C2H2-ZNF proteins with significant e-values less than 10^{-30}. All these e-value levels indicate that our results and analyses on the domain, EST, cDNA and protein matches are strongly convincible.

The frameshifting site is quite likely: The UUU Ynn pattern (Y is pyrimidine such as cytosine (C), thymine (T) and uracil (U); n represents any nucleotide) has been well studied and confirmed as highly frameshift-prone site for +1 programmed ribosomal frameshifting in both eubacteria and mammalian cells[4,9,10]. The sequence pattern UUUCCU that we identified at five upstream nucleotides before the pseudoknot precisely matches the model pattern of UUU Ynn. Moreover, we also find this likely frameshifting site in C2H2-ZNF ESTs 1858722, 2555712, 3228157, 575744 and 6713901 and a C2H2-ZNF cDNA ENSG00000171295. So, UUUCCU is a very likely frameshift site in Z3-7.

Further, let us discuss the possibility that a ribosome shifts into another reading frame at this slippery site, UUUCCU. An mRNA consists of only exons after RNA processing[28]. So the mRNA of Z3-7 includes only those domains in ORF1. Here, a domain is considered as an exon. From Table 1, observe that the first nucleotide of the pseudoknot is at the 112th nucleotide of the mRNA of Z3-7, so the pseudoknot covers the end of the second domain and the start of the third domain in ORF1 (as one C2H2-ZNF domain ranges only 69 nucleotides) and the frameshifting site that we identified lies in the second domain of ORF1. The second domain (from position 1597 to 1665 in its original DNA sequence) of ORF1 is at the upstream of the first domain (from position 2066 to 2134) of ORF2 (Table 1 and Fig. 1). Hence, from this likely slippery site, a ribosome can slip into a position at the upstream of the first domain of ORF2 in Reading Frame +1 so that a +1 ribosomal frameshifting occurs. In this sense, the slippery site UUUCCU identified in this study is also likely.

Nevertheless, wet experiments are needed to verify whether this site is really functional to cause the +1 programmed ribosomal frameshifting.

The +1 ribosomal frameshifting in Z3-7 is likely efficient: A frameshifting-prone site is a main factor to cause a ribosomal frameshifting[4]. Moreover, if a pseudoknot is found to present in the same mRNA of this slippery site, the efficiency of a ribosomal frameshifting can be greatly elevated[4]. As seen before, we found a significant pseudoknot and a likely frameshift-prone site that are in the same mRNA of Z3-7, so we believe that the efficiency of the programmed ribosomal frameshifting in Z3-7 may be high. On the other hand in literature, a study by Matsufuji *et al.*[14] about rat antizyme 1 and a study by Ivanov *et al.*[15] about human antizyme 2 have both reported wet-laboratory confirmed efficiency on how 3′ RNA pseudoknot of a frameshifting site stimulates a +1 programmed ribosomal frameshifting.

Another factor, the spacer length between a frameshifting site and a pseudoknot, is also important to affect the efficiency of ribosomal frameshifting[5]. As mentioned above, the spacer length between the frameshifting site and the pseudoknot in the present study is five nucleotides. Next, we discuss how the spacer length affects the efficiency and explain why our spacer length of five nucleotides is reasonable. Although the

known studies about this factor are only focused on -1 programmed ribosomal frameshifting, we can infer some knowledge about +1 programmed ribosomal frameshifting through analyzing these reported -1 programmed ribosomal frameshifting since both +1 frameshifting and -1 frameshifting are caused by one base slip of ribosomes, forward 3' direction and backward 5' direction, respectively and since both +1 frameshifting and -1 frameshifting can be describe by an integrated model[34]. Brierley *et al.*[35] had a mutational analysis on how spacer lengths in a coronavirus Infectious Bronchitis Virus (IBV) affect the efficiency of ribosomal frameshifting and considered six nucleotides as the optimal length. Napthine *et al.*[36] also conducted a similar study. They changed the spacer lengths from three to nine nucleotides in a plasmid pKA13 and tested the frameshifting efficiency for each mutant. They found that the optimal length is eight nucleotides and also found that three, four, five, six, seven, eight or nine nucleotides are all functional to produce a frameshifting though with different efficiencies. From the studies of Brierley *et al.*[35] and Napthine *et al.*[36] on the -1 riobosomal frameshifting, we can see that the spacer length between a slippery site and a pseudoknot is not a fixed number for all organisms, it may be different for different organisms. However, we emphasize that in one organism, the precise length of this spacer must be maintained in order to keep maximal framshifting efficiency[37].

Moreover, as discussed earlier, Z3-7 has two overlapping ORFs, ORF1 and ORF2, in different reading frames (Fig.1). ORF1 is within the original sequence (Reading Frame 0) and ORF2 is a product from ribosomes shifting one base into Reading Frame +1. Furthermore, these two ORFs of Z3-7 significantly match two reading frames of seven human C2H2-ZNF ESTs with e-values less than 10^{-20} and five such ESTs have the likely frameshifting site UUUCCU (Table 2). As well, these two ORFs significantly match two reading frames of three human C2H2-ZNF cDNAs (e-value<10^{-30}) that correspond human C2H2-ZNF proteins (Table 2). Also, the amino acid sequences of these two ORFs significantly match the human C2H2-ZNF proteins of SWISS-PROT with e-values less than 10^{-30} (Fig. 1). These facts about the transcription and translation of the two ORFs further give us confidence that the +1 programmed ribosomal frameshifting in Z3-7 is quite likely efficient.

A programmed ribosomal frameshifting is likely an important regulation mechanism for gene expression in human genome: We agree that a programmed ribosomal frameshifting is universal but that only a minority of mRNAs uses this mechanism[4,38] since we discover a likely +1 ribosomal frameshifting in 226 C2H2-ZNF genes in human chromosome 19. According to our potential discovery, it is also reasonable to infer that a programmed ribosomal frameshifting is likely an important regulation mechanism for some genes' expression in human genome. More concretely, in human genome, a programmed ribosomal frameshifting may function at least at two points as follows.

Firstly, it provides a mechanism to use an mRNA in high efficiency. Usually, a programmed ribosomal frameshifting produces two proteins from one mRNA. One of the two proteins is from ORF1 and the other protein is from the overlapping of ORF1 and ORF2. These two proteins can be expressed at the same time in a fixed ratio[4], or may be expressed at different time or in different conditions. Besides, the two proteins may be distinct[5] and also may be structurally relative and functionally similar[39]. In our case, Z3-7 probably expresses three proteins since each of its two ORFs has a complete primary structure of a gene and has significant matches with human C2H2-ZNF ESTs and human C2H2-ZNF proteins. Possibly, the first protein is from ORF1 in Reading Frame 0, the second protein is from ORF2 in Reading Frame +1 and the third protein is a fused protein from overlapping sequence of ORF1 and ORF2. Our inference is supported by a study of Shigemoto *et al.*[8] where a mammalian Edr gene encodes three proteins, separately from ORF1, ORF2 and the overlapping of ORF1 and ORF2. As a future research topic, we will look for a collaborator to conduct laboratory experiments to test whether the three possible proteins are translated *in vivo*.

Secondly, it also provides a mechanism to produce a novel protein. In some cases, a programmed ribosomal frameshifting only yields one protein from the combined sequence of ORF1 and ORF2[33]. Whether a programmed ribosomal frameshifting produces one protein or two proteins, the novel protein from the fused sequence of ORF1 and ORF2 are different with the standard protein product from ORF1. In our paper, the three proteins from Z3-7 should be structurally relative, functionally similar and they might belong to the same type of C2H2-ZNF protein. Because these three proteins have different numbers of C2H2-ZNF domains (Table 1), each of them may have specific functions.

Moreover, a programmed ribosomal frameshifting plays a regulatory role in metabolic pathways, too. For example, a frameshifting makes animal antizyme expression not regulated by antizyme itself, but rather by polyamines whose intracellular level is governed by antizyme[4,14].

ACKNOWLEDGMENTS

We are grateful to Charlie Hodgman, Chua Nam Hai, Ng See Kiong and Ellen Mackay for their useful comments on this study and to Jinyan Li for carefully reading the early draft.

REFERENCES

1. Nei, M. and S. Kumar, 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.
2. Li, Z., G. Stahl and P.J. Farabaugh, 2001. Programmed +1 frameshifting stimulated by complementarity between a downstream mRNA sequence and an error-correcting region of rRNA. RNA, 7: 275-284.
3. Baranov, P.V., O.L. Gurvich, A.W. Hammer, R.F. Gesteland and J.F. Atkins, 2003. RECODE 2003. Nucleic Acids Research, 31: 87-89.
4. Atkins, J.F., P.V. Baranov, O. Fayet, A.J. Herr, M.T. Howard, I.P. Ivanov, S. Matsufuji, W.A. Miller, M.B. Moore, M.F. Prère,N.M. Wills, J. Zhou and R.F. Gesteland, 2001. Overriding standard decoding: Implications of recoding for ribosome function and enrichment of gene expression. Cold Spring Harbor Symposia on Quantitative Biology. Volume LXVI, 217-232.
5. Alam, S.L., J.F. Atkins and R.F. Gesteland, 1999. Programmed ribosomal frameshifting: Much ado about knotting. Proc. Natl Acad. Sci. USA., 96: 14177-14179.
6. Atkins, J.F. and R.F. Gesteland, 1999. Intricacies of ribosomal frameshifting. Nature Structural Biology, 6: 206-207.
7. Farabaugh, P.J., 1996. Programmed translational frameshifting. Annu. Rev. Genet., 30: 507-528.
8. Shigemoto, K., J. Brennan, E. Walls, C.J. Watson, D. Stott, P.W.J. Rigby and A.D. Reith, 2001. Identification and characterization of a developmentally regulated mammalian gene that utilizes -1 programmed ribosomal frameshifting. Nucleic Acids Research, 29: 4079-4088.
9. Fu, C. and J. Parker, 1994. A ribosomal frameshifting error during translation of the argI message of *Escherichia coli*. Mol. Gen. Genet., 243: 434-441.
10. Schwartz, R., J.F. Curran, 1997. Analyses of frameshifting at UUU-pyrimidine sites. Nucleic Acids Research, 25: 2005-2011.
11. Shah, A.A., M.C. Giddings J.B. Parvaz, R.F. Gesteland, J.F. Atkins and I.P. Ivanov, 2002. Computational identification of putative programmed translational frameshift sites. Bioinformatics, 18: 1046-1053.
12. Sundararajan, A., W.A. Michaud, Q. Qian, G. Stahl and P.J. Farabaugh, 1999. Near-cognate peptidyl-tRNAs promote +1 programmed translational frameshifting in yeast. Molecular Cell, 4: 1005-1015.
13. Belcourt, M.F. and P.J. Farabaugh, 1990. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell. 62: 339-352.
14. Matsufuji, S., T. Matsufuji, Y. Miyazaki, Y. Murakami, J.F. Atkins, R.F. Gesteland and S. Hayashi, 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. Cell. 80: 51-60.
15. Ivanov, I.P., R.F. Gesteland and J.F. Atkins, 1998. A second mammalian antizyme: Conservation of programmed ribosomal frameshifting. Genomics, 52: 119-129.
16. Ivanov, I.P., A. Rohrwasser, D.A. Terreros, R.F. Gesteland and J.F. Atkins, 2000. Discovery of a spermatogenesis stage-specific ornithine decarboxylase antizyme: antizyme 3. Proc. Natl Acad. Sci. USA., 97: 4808-4813.
17. Meng S.W., Z. Zhang and J. Li, 2004. Twelve C2H2 zinc-finger genes on human chromosome 19 can be each translated into the tame type of protein after frameshifts. Bioinformatics, 20: 1-4.
18. Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, O.H. Smith, M. Yandell, C.A. Evans, R.A. Holt *et al*., 2001. The Sequence of the human genome. Science, 291: 1304-1351.
20. Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall and E.L.L. Sonnhammer, 2002. The Pfam contribution to the annual NAR database issue. Nucleic Acids Research, 30: 276-280.
21. Shannon, M. and L. Stubbs, 1998. Analysis of homologous *XRCC1*-linked zinc-finger gene families in human and mouse: Evidence for orthologous genes. Genomics, 49: 112-121.
22. Tune, C. E., M. Pilon, Y. Saiki and H.M. Dosch, 2002. Sustained expression of the novel EBV-induced Zinc Finger Gene, ZNF^{EB}, is critical for the transition of B lymphocyte activation to oncogenic growth transformation. J. Immunol., 168: 680-688.
23. Dehal, P., P. Predki, A.S. Olsen, A. Kobayashi, P. Folta, S. Lucas, M. Land, A. Terry, C.L.E. Zhou, S. Rash, Q. Zhang, L. Gordon, J. Kim, C. Elkin, M.J. Pollard, P. Richardson, D. Rokhsar, E. Uberbacher, T. Hawkins, E. Branscomb and L. Stubbs, 2001. Human Chromosome 19 and related regions in Mouse: Conservative and lineage-specific evolution. Science, 293: 104-111.

24. Bajic, V.B. and S.H. Seah, 2003. Dragon gene start finder: An advanced system for finding approximate locations of the start of gene transcriptional units. Genome Research, 13: 1923-1929.
25. Nishikawa, T., T. Ota and T. Isogai, 2000. Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. Bioinformatics, 16: 960-967.
26. Milanesi, L., D. D'Angelo and I.B. Rogozin, 1999. GeneBuilder: Interactive in silico prediction of gene structure. Bioinformatics., 15: 612-621.
27. Brodsky, L.I., V.V. Ivanov, Kalai, L. dzidis Ya., A.M. Leontovich, V.K. Nikolaev, S.I. Feranchuk and V.A. Drachev, 1995. GeneBee-NET: Internet-based server for analyzing biopolymers structure. Biochemistry, 60: 923-928.
28. Weaver, R.F., 1998. Molecular Biology. WCB/McGraw-Hill, Boston.
29. Gish, W. and D.J. States, 1993. Identification of protein coding regions by database similarity search. Nature Genet., 3: 266-272.
30. Clamp, M.D. Andrews, D. Barker, P. Bevan, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuffand V. Curwen, *et al.*, 2003. Ensembl 2002: accommodating comparative genomics. Nucleic Acids Research, 31: 38-42.
31. Altschul, S.F. and W. Gish, 1996. Local alignment statistics. Methods in Enzymology, 266: 460-480.
32. Portin, P., 2002. Historical development of the concept of the gene. J. Med. Philos., 27: 257-286.
33. Baranov, P.V., O.L. Gurvich, O. Fayet, M.F. Prère, W.A. Miller, R.F. Gesteland, J.F. Atkins and M.C. Giddings, 2001. RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression. Nucleic Acids Research, 29: 264-267.
34. Harger, J.W., A. Meskauskas and J.D. Dinman, 2002. An: Integrated model of programmed ribosomal frameshifting. Trends in Biochemical Sci., 27: 448-454.
35. Brierley, I., A.J. Jenner and S.C. Inglis, 1992. Mutational analysis of the slippery-sequence component of a coronavirus ribosomal frameshifting signal. J. Mol. Biol., 227: 463-479.
36. Napthine, S., J. Liphardt, A. Bloys, S. Routledge and I. Brierley, 1999. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. J. Mol. Biol., 288: 305-320.
37. Brierley, I. and S. Pennell, 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic -1 ribosomal frameshifting. Cold Spring Harbor Symposia on Quantitative Biology, LXVI: 233-248.
38. Gesteland, R.F., R.B. Weiss and J.F. Atkins, 1992. Recoding: Reprogrammed genetic decoding. Sciences, 257: 1640-1643.
39. Larsen, B., R.E. Gesteland and J.F. Atkins, 1997. Structural probing and mutagenic analysis of the stem-loop required for *Escherichia coli dnaX* ribosomal frameshifting: programmed efficiency of 50%. J. Mol. Biol., 271: 47-60.