Rabbit β-Globin Is Extended Beyond Its UGA Stop Codon by Multiple Suppressions and Translational Reading Gaps†

Harold S. Chittum,‡ William S. Lane,§ Bradley A. Carlson,‡ Peter P. Roller,|| Feng-Di T. Lung,|| Byeong Jae Lee,⊥ and Dolph Lee Hatfield*,‡

Laboratory of Basic Research and Laboratory of Medicinal Chemistry, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Harvard Microchemistry Facility, 16 Divinity Avenue, Cambridge, Massachusetts 02138, and Laboratory of Molecular Genetics, Institute for Molecular Biology and Genetics, Seoul National University, Seoul 151-742, Korea

Received May 6, 1998; Revised Manuscript Received June 25, 1998

ABSTRACT: Translational reading gaps occur when genetic information encoded in mRNA is not translated during the normal course of protein synthesis. This phenomenon has been observed thus far only in prokaryotes and is a mechanism for extending the reading frame by circumventing the normal stop codon. Reading frames of proteins may also be extended by suppression of the stop codon mediated by a suppressor tRNA. The rabbit β-globin read-through protein, the only known, naturally occurring read-through protein in eukaryotes, was sequenced by ion trap mass spectrometry to determine how the reading frame is extended. Seven different proteolytic peptide fragments decoded by the same sequence that spans the UGA stop codon of rabbit β-globin mRNA were detected. Three of these peptides contain translational reading gaps of one to three amino acids that correspond to the UGA stop codon site and/or one or two of the immediate downstream codons. To our knowledge, this is the first reported example of the occurrence of reading gaps in protein synthesis in eukaryotes. This event is unique in that it is associated with bypasses involving staggered lengths of untranslated information. Four of the seven peptides contain serine, tryptophan, cysteine, and arginine decoded by UGA and thus arise by suppression. Serine is donated by selenocysteine tRNA, and it, like the other tRNAs, has previously been shown to suppress UGA in vitro in mammals, but not in vivo.

Genetic information encoded in mRNA is sometimes bypassed during translation. This phenomenon is called ribosomal hopping or translational bypassing, and its purpose is to extend the reading frames of proteins by circumventing a stop codon (reviewed in refs *1* and *2*). Other mechanisms that extend reading frames beyond a termination codon include in-frame suppression of termination codons by a suppressor tRNA and ribosomal frame shifting that alters the reading frame upstream of the termination codon either in the +1 or in the −1 frame. In-frame suppression and ribosomal frame shifting have been observed in a variety of organisms (see above reviews), while ribosomal hopping has only been observed in prokaryotes (*3*−*5*). Hops involving as few as two or as many as 50 nucleotides have been reported.

A mammalian protein, mammalian antizyme, was shown to arise by a +1 frame shift that occurs just before a UGA stop codon (*6*), and more recently, the antizyme gene in *Drosophila* was found to require +1 frame shifting for expression (*7*). In addition, many other examples of frame shifting in the +1 direction as well as in the −1 direction, and many examples of in-frame suppression, have been well-characterized in eukaryotes (*1*, *2*). However, the only known, naturally occurring read-through protein that arises from a gene within the genome of a eukaryote is the rabbit β-globin read-through protein (*8*, *9*). It was first shown to be a naturally occurring read-through protein by labeling intact rabbit reticulocytes with [^{35}S]methionine and observing a labeled spot on a two-dimensional gel in the area where the protein was expected to migrate (*8*). The occurrence of this protein in rabbit reticulocytes was further substantiated when antibodies were generated to its read-through portion and the protein was isolated and partially characterized (*9*). Since the termination codon for rabbit β-globin mRNA is UGA and UGA is recognized as a "leaky" codon (*10*, *11*), it was assumed that this protein arose by suppression of its stop codon (*8*, *9*), mediated by tryptophan tRNA misreading (*8*). Sequencing the read-through protein at the termination site, however, is required to demonstrate whether the protein arises by suppression of its stop codon or by another mechanism.

In this study, we have isolated the rabbit β-globin read-through protein from rabbit reticulocytes to determine how the reading frame is extended. Surprisingly, the read-through

† This work was supported in part by the Genetic Engineering Research Fund from the Ministry of Education of Korea.

* To whom correspondence should be addressed: National Cancer Institute, National Institutes of Health, Bldg. 37, Rm. 2D09, Bethesda, MD 20892. Telephone: (301) 496-2797. Fax: (301) 435-4957. E-mail: hatfield@dc37a.nci.nih.gov.

‡ Laboratory of Basic Research, Division of Basic Sciences, National Cancer Institute, National Institutes of Health.

§ Harvard Microchemistry Facility.

|| Laboratory of Medicinal Chemistry, Division of Basic Sciences, National Cancer Institute, National Institutes of Health.

⊥ Seoul National University.

S0006-2960(98)01042-3 CCC: $15.00
Published on Web 07/11/1998

protein exists as a heterogeneous population of proteins. The sequences of seven different proteolytic peptide fragments that are all decoded by the same mRNA sequence, UAC CAC *UGA* GAU CUU UUU CCC UCU GCC AAA, where *UGA* is the termination codon, were determined. Four peptides result from in-frame suppression as they contain different amino acids, serine, tryptophan, cysteine, and arginine, corresponding to *UGA*. Three peptides result from translational reading gaps as they lack an amino acid or amino acids corresponding to *UGA* and/or one or two of the immediate downstream codons.

MATERIALS AND METHODS

Reagents. The Imject Activated Immunogen Conjugation Kit and SulfoLink and AminoLink Plus kits used for coupling the read-through peptide and anti-read-through peptide antibodies, respectively, were purchased from the Pierce Chemical Co. Reagents used for ELISAs[1] were purchased from Kirkegaard and Perry, Inc. The ECL Plus detection kit used for development of Western blots was purchased from Amersham.

Synthetic Peptide and the Peptide−KLH Conjugate. A synthetic peptide 23 amino acids in length corresponding to the read-through portion of rabbit β-globin mRNA with an additional cysteine residue at the amino terminus was synthesized for the purpose of antibody production. The resulting synthetic peptide contained the sequence CDLFPSAKNYGDIMKPLEHLTSG. The synthesis was preformed on a PAL-amide resin using Fmoc solid phase synthesis technology with the aid of a PE/ABI 433A Peptide Synthesizer. Side chain deprotection and cleavage from the resin were accomplished with trifluoroacetic acid/thioanisole/water/ethanedithiol (10:0.5:0.5:0.25) for 2 h. After purification, the identity and purity of the peptide were ascertained by reverse phase HPLC and FAB-MS. The matrix for mass spectrometry was glycerol (VG 7070-EHF instrument), and the observed molecular ion was as follows: $(M + H)^+$ 2537.5 (calcd 2536.9).

For immunization purposes, the peptide was conjugated to sulfonylmaleimide-activated KLH in PBS (pH 7.2) containing 0.1 M EDTA, 0.9 M NaCl, and 0.02% sodium azide for 2 h at room temperature using an Imject Activated Immunogen Conjugation Kit. The resulting conjugate was purified by gel filtration with a molecular weight cutoff of 5000.

Antisera Preparation and Immunodetection. Polyclonal antisera which recognized the read-through peptide were generated in a sheep with peptide−KLH by Covance Laboratories, Inc. Boosters were given every 21 days to increase antibody titer. The sheep was bled initially at 14 days and then at 21 day intervals thereafter. Antibody titer was measured by ELISA as previously described (*9*).

Affinity Chromatography. The techniques used in isolating the read-through proteins were like those previously described (*9*) with the exception that the peptide was coupled to SulfoLink agarose gel (using the instructions provided in the kit) through the amino-terminal cysteine. The resulting antibodies were then coupled to agarose using Pierce's AminoLink Plus kit and the instructions included with the kit. The read-through proteins were isolated from rabbit reticulocytes (New Zealand white males) as described (*9*). The care of all animals was in accordance with National Institutes of Health institutional guidelines under the expert direction of G. Lidl, D.V.M. (National Cancer Institute, National Institutes of Health).

MS−MS Peptide Sequencing. The direct sequence analysis by ion trap mass spectrometry was performed on rabbit β-globin read-through peptides resulting from the *in-gel* reduction, S-carboxyamidomethylation, and tryptic digestion of the protein isolated with SDS−PAGE . The resulting mixture was introduced directly into the electrospray ionization (ESI) source of a quadrupole ion trap mass spectrometer (Finnigan LCQ) by a reverse phase microcapillary column (75 μm inside diameter × 180 μm outside diameter, SGE), packed *in-house* with POROS R2 (Perseptive Biosystems) (*12*). Peptides were eluted at a nominal flow rate of 500 nL/min with a gradient of 0 to 50% acetonitrile in 0.05 M acetic acid over the course of 25 min. The ion trap's online data-dependent scans allowed the automatic acquisition of high-resolution (zoom scan) spectra for determining charge state and exact mass and MS−MS spectra for peptide sequence information. Identification of the spectra corresponding to known peptide sequences in the NCBI (National Center for Biotechnology Information) nr and dbest databases was facilitated using the algorithm Sequest (*13*) and then confirmed by manual inspection. The remaining unidentified MS−MS spectra were manually interpreted with the aid of a program we developed for de novo peptide sequence interpretation. The program "FuzzyIons" is a web browser-based, interactive workspace that facilitates manual de novo interpretation of peptide MS−MS spectra by autohighlighting visual cues in a manual interpretation while removing the burden of arithmetic calculations (*14*). This program improved our ability to rapidly review unidentified spectra and reveal the sequences of the novel stop codon-spanning peptides.

RESULTS AND DISCUSSION

To determine the sequence of the rabbit β-globin read-through protein at its termination site, and thus elucidate the protein's synthesis, it was isolated from an immunoaffinity column containing purified antibodies generated against the read-through portion of the protein. The immunopurified protein was further purified by SDS−PAGE, and the Coomassie Blue-stained band corresponding to the read-through protein (Figure 1) was excised, S-carboxyamidomethylated, and digested with trypsin *in-gel*. The majority of the adult β-globin sequence was identified as described in Materials and Methods (see Table 1). The sequences of residues 2−60, 67−105, and 122−147 as well as the entire read-through region were determined, confirming the identity of the β-globin read-through protein. Seven novel peptides that spanned the termination codon, YHDLFPSAK, YHLFPSAK, YHFPSAK, YHWDLFPSAK, YHSDLFPSAK, YHRDLFPSAK, and YHCDLFPSAK, were also sequenced (see Tables 1 and 2). MS−MS spectra used for the sequencing of two representative peptides (YHLFPSAK and YHWDLFPSAK) are shown in Figures 2 and 3. Thus, the

[1] Abbreviations: Sec, selenocysteine; ELISA, enzyme-linked immunosorbent assay; KLH, keyhole limpet hemocyanin; FAB-MS, fast atom bombardment mass spectrometry; MS−MS, mass spectrometry−mass spectrometry.

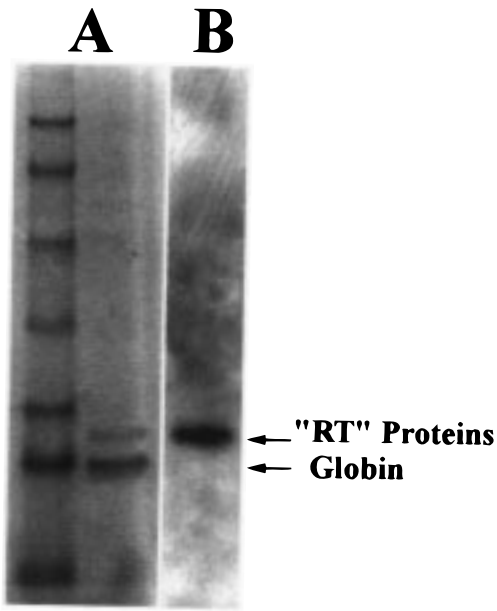

FIGURE 1: Isolation and Western blot analysis of rabbit β-globin read-through proteins. In panel A, protein molecular mass standards with molecular masses ranging from 6 to 62 kDa are shown in the left-hand lane and the immunopurified sample from rabbit reticulocytes (see Materials and Methods) is shown in the right-hand lane after resolution on polyacrylamide gels and staining with Coomassie Blue. α and β chains of rabbit globin were also isolated from immunoaffinity columns [presumably by nonspecific attachment (see also ref *9*)] as shown by the faster migrating band in the right-hand lane of A. In panel B, a Western blot following transfer of the proteins (see panel A) to a PVDF membrane is shown. Western blots were generated as described in Materials and Methods. The arrow designating "RT" proteins shows the position of the read-through proteins, and that designating globin shows the position of the α and β chains of rabbit globin.

immunopurified band obtained on gels (Figure 1) exists as a heterogeneous population of rabbit β-globin "read-through" proteins.

The peptides YHWDLFPSAK, YHSDLFPSAK, YHRDLFPSAK, and YHCDLFPSAK correspond identically to the RNA sequence encoding this peptide (see Table 2). Tryptophan, serine, arginine, and cysteine occur at the position occupied by the UGA termination codon. Thus, multiple tRNAs are responsible for suppressing UGA in rabbit β-globin mRNA. Tryptophan, arginine, and cysteine have been shown to serve as suppressor tRNAs of a single UGA stop codon in vitro in rabbit reticulocyte lysates (*15*). This study confirms that this observation also occurs in vivo. The most likely tRNA that carries out suppression by donating its serine in response to UGA is selenocysteine (Sec) tRNA (reviewed in ref *16*). Sec tRNA specifically decodes UGA, is initially aminoacylated with serine, and is therefore designated tRNA$^{[Ser]Sec}$. Sec tRNA$^{[Ser]Sec}$ occurs in rabbit reticulocytes (*17*) and is capable of suppressing UGA in rabbit β-globin mRNA in in vitro assays (*18*−*20*), and seryl-tRNA$^{[Ser]Sec}$, but not selenocysteyl-tRNA$^{[Ser]Sec}$, is recognized by the normal elongation factor, eEF-1 (*20*). Furthermore, a serine tRNA that decodes UGA was not found in mammals by using specific assays designed to detect opal suppressor tRNAs (*21*). Since seryl-tRNA$^{[Ser]Sec}$ is most certainly responsible for donating serine at the read-through site to extend β-globin, this observation means that the serylated form of tRNA$^{[Ser]Sec}$ plays a role in protein synthesis and serves as an authentic suppressor tRNA.

Table 1: Tryptic Peptide Fragment Sequences Identified by Ion Trap Mass Spectrometry from the Rabbit β-Globin Read-Through Protein

residue number[a]	sequence	retention time[b] (min)	*m*/*z*[c]	*z*[d]
	Peptides Which Span the Read-Through Site			
146−155	YHDLFPSAK	34.2	539.26	2
146−155	YHLFPSAK	34.6	481.76	2
146−155	YHFPSAK	25.0	425.22	2
146−155	YHWDLFPSAK	−	632.31	2
146−155	YHSDLFPSAK	32.8	582.83	2
146−155	YHRDLFPSAK	−	617.30	2
146−155	YHCDLFPSAK	−	619.27	2
	Carboxy-Terminal Read-Through Peptide			
156−170	NYGDIMKPLEHLTSG	50.5	837.93	2
	Peptides Which Constitute Normal β-Globin			
2−9	VHLSEEK	11.5	928.46	1
10−18	SAVTALWGK	40.1	932.51	1
19−31	VNVEEVGGEALGR	35.5	664.82	2
32−41	LLVVYPWTQR	52.2	637.85	2
42−60	FFESFGDLSSAHAVMSNPK	51.0	1038.00	2
67−83	KVLAAFSEGLSHLDNLK	54.4	921.56	2
89−96	LSELHCDK	14.1	501.24	2
97−105	LHVDPENFR	33.2	563.78	2
122−133	EFTPQVQAAYQK	30.1	705.32	2
134−145	VVAGVANALAHK	31.7	575.37	2

[a] Residue numbers for peptides beyond the normal stop position were numbered according to contiguous codon positions (i.e., the UGA at the normal stop position corresponds to residue number 148). Dashes indicate that the sequence was determined in a separate chromatographic run using selected ion monitoring (SIM) for the predicted *m*/*z*. [b] Retention time on capillary HPLC−MS. [c] Mass to charge ratio of molecular ions. [d] Charge state of protonated molecular ions.

The other three peptides identified in this study that span the stop codon in rabbit β-globin mRNA lack an amino acid corresponding to the termination site and, in addition, the next one or two downstream amino acids. It would seem highly unlikely that the corresponding parent proteins would arise by processing rabbit β-globin mRNA through mechanisms such as editing (*22*, *23*) or splicing (*24*−*27*). Clearly, a staggered mechanism of removing three nucleotides at a time in three separate stages would be required to synthesize the necessary mRNAs. There is no precedent in either the RNA editing or splicing event that would result in such multiple mRNA forms. Furthermore, it seems unlikely that the occurrence of these proteins would arise by protein splicing (*28*, *29*) as there is no precedent in protein splicing that would result in such multiple protein forms. Therefore, the unfilled site or sites in the peptides sequenced herein most likely result from translational reading gaps that are associated with varying lengths of uncoded genetic information. Reading gaps in genetic information or ribosomal hops have been described thus far only in bacteria and bacteriophage (*2*−*5*). The launch and landing sites in prokaryotes involve identical, or at least similar, codons at each site, and the length of the bypass involving any single event, which may include as few as two or as many as 50 nucleotides, is always constant (reviewed in refs *1* and *2*). On the other hand, the reading gap in mammals seems to reflect translational stuttering after the initial bypass of the stop codon in that the length of the abyss may involve only the termination codon or may include, in addition, the next one or two downstream codons.

Table 2: Tryptic Peptides from the Rabbit β-Globin Read-Through Protein That Span the UGA Stop Codon[a]

sequence										MW	relative ion abundance
UAC	CAC	UGA	GAU	CUU	UUU	CCC	UCU	GCC	AAA		
Y	H		D	L	F	P	S	A	K	1076.51	1.0
Y	H			L	F	P	S	A	K	961.51	3.1
Y	H				F	P	S	A	K	848.43	4.2
Y	H	*W*	D	L	F	P	S	A	K	1262.61	11.0
Y	H	*S*	D	L	F	P	S	A	K	1163.65	3.6
Y	H	*R*	D	L	F	P	S	A	K	1232.59	0.6
Y	H	*C*	D	L	F	P	S	A	K	1236.53	0.4

[a] Sequences of the tryptic peptides which span the read-through site are aligned with the corresponding nucleotide sequence written as codons. The relative abundance of the full scan doubly protonated ion for each peptide is expressed relative to that of the first peptide. Observed ion intensities in part reflect the ability of a particular peptide to be ionized and do not necessarily correspond to relative quantities.

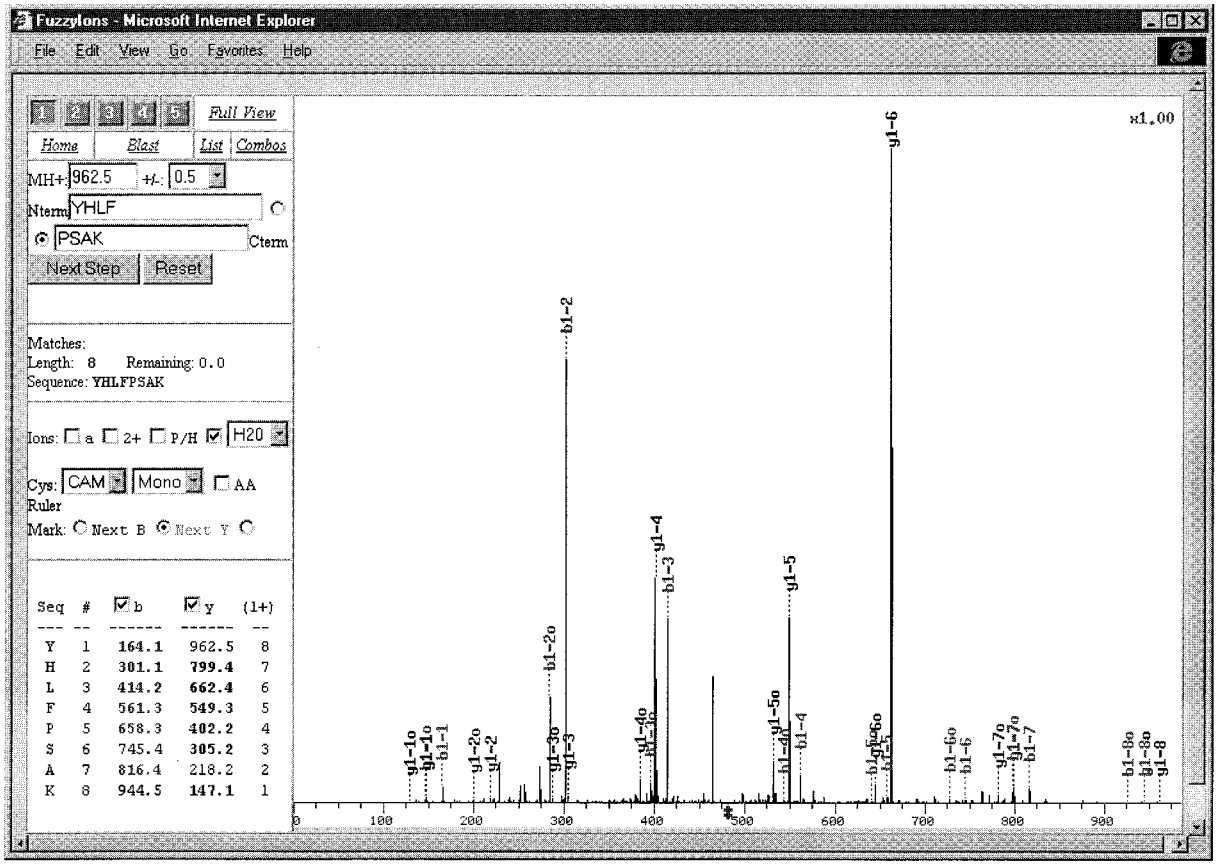

FIGURE 2: MS−MS spectrum for one of the tryptic peptides, YHLFPSAK, that bypasses the UGA stop codon in rabbit β-globin mRNA. The completed de novo interpretation using the program FuzzyIons (see Materials and Methods) is shown.

The stop signal that dictates cessation of protein synthesis at the carboxyl terminus of the rabbit β-globin extension proteins described in this study is comprised of two tandem UAA codons (*9*). The occurrence of such tandem stop codons would seem to imply that protein synthesis must absolutely terminate at this site. The presence of tandem UAA codons downstream of the normal codon in globin mRNAs of other mammals may suggest that hopping and/or suppression also occur(s) in these animals. We therefore examined other mammalian β-globin genes for the presence of the tandem UAA sequence that occurs downstream of the β-globin termination codon in rabbits using the NCBI nucleotide database. Interestingly, this sequence has been conserved in other mammals. It occurs in the same reading frame as the β-globin stop codon in rabbits, sheep, cows, and lemurs, in the -1 reading frame with respect to the β-globin stop codon in humans and rats, and in the $+1$ reading frame in pigs. Hopping into a different reading frame has been observed in a number of translational bypass events in prokaryotes (*2*, *3*, *5*). Whether these other mammals contain a protein or proteins that arise by an extension of β-globin remains to be established.

UGA serves as both a stop and a Sec codon in the genetic code that is most widely used in nature (*30*). Thus, translation of UGA as Sec is not a mechanism for extending the reading frame beyond a stop codon, but is the actual decoding of UGA as Sec. The mechanism involved in bypassing the termination codon in the rabbit β-globin read-through protein clearly involves translational stuttering in selecting a landing site and provides a means for extending the reading frame. In addition, the launch and landing codons are not similar to each other as has been found in prokaryotes (*2*−*5*). Thus, translational reading gaps that are associated with a fissure of different lengths appear to be a

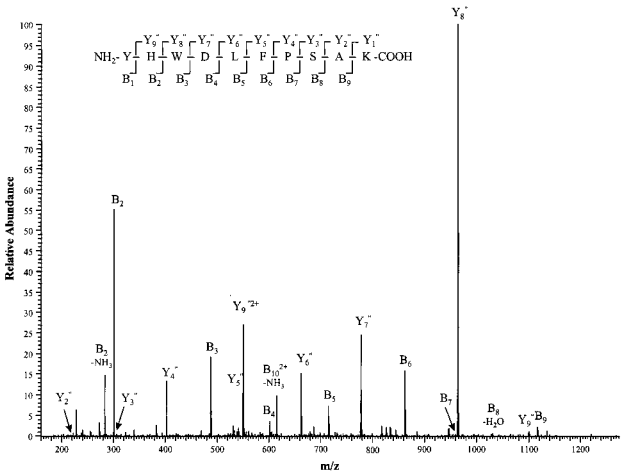

FIGURE 3: MS−MS spectrum of a peptide resulting from suppression of the UGA stop codon in rabbit β-globin mRNA. The spectrum used to derive the sequence of the tryptic peptide YHWDLFPSAK is shown. The upper left corner indicates the source of the ion fragment peaks identified as either B or Y ions.

novel mechanism for circumventing termination codons and extending proteins. The underlying mechanism(s) that causes reading gaps and the resulting, apparently dysfunctional, length of the bypass and the role of the rabbit β-globin readthrough proteins described in this study await further elucidation.

REFERENCES

1. Farabaugh, P. J. (1996) *Microbiol. Rev. 60*, 103.
2. Gesteland, R. F., and Atkins, J. F. (1996) *Annu. Rev. Biochem. 65*, 741−768.
3. Weiss, R. B., Dunn, D. M., Atkins, J. F., and Gesteland, R. F (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 687−693.
4. Kane, J. F., Violand, B. N., Curran, D. F., Staten, N. R., Duffin, K. L., and Bogosian, G. (1992) *Nucleic Acids Res. 20*, 6707−6712.
5. Manch-Citron, J. N., and London, J. (1994) *J. Bacteriol. 176*, 1944−1948.
6. Matsufuji, S., Matsufuji, T., Murakami, Y., Atkins, J. F., Gesteland, R. F., and Hayashi, S. (1995) *Cell 80*, 51−60.
7. Ivanov, I. P., Simin, K., Letsou, A., Atkins, J. F., and Gesteland, R. F. (1998) *Mol. Cell. Biol. 18*, 1553−1561.
8. Geller, A. I., and Rich, A. (1980) *Nature 283*, 41−46.
9. Hatfield, D., Thorgeirsson, S. S., Copeland, T. D., Oroszlan, S., and Bustin, M. (1988) *Biochemistry 27*, 1179−1183.
10. Weiner, A. M., and Weber, K. (1973) *J. Mol. Biol. 80*, 837−855.
11. Lovett, P. S., Ambulos, N. P., Jr., Mulbry, W., Noguchi, N., and Rogers, E. J. (1991) *J. Bacteriol. 173*, 1810−1812.
12. Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S., and Verdine, G. L. (1996) *Curr. Biol. 6*, 968−980.
13. Eng, J. K., McCormick, A. L., and Yates, J. R., III (1994) *J. Am. Soc. Mass. Spectrom. 5*, 976−989.
14. Lane, W. S., Eng, J., Yates, J. R., III, and Baker, M. A. (1998) The 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL.
15. Feng, Y.-A., Copeland, T. D., Oroszlan, S., Rein, A., and Levin, J. G. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8860−8863.
16. Lee, B. J., Park, S., Park, J., Chittum, H. S., and Hatfield, D. L. (1996) *Mol. Cells 6*, 509−520.
17. Hatfield, D. L., Matthews, C. R., and Rice, M. (1979) *Biochim. Biophys. Acta 564*, 414−423.
18. Diamond, A. M., Dudock, B., and Hatfield, D. (1981) *Cell 25*, 497−506.
19. Hatfield, D., Diamond, A. M., and Dudock, B. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6215−6219.
20. Jung, J.-E., Karoor, V., Sandbacken, M. G., Lee, B. J., Ohama, T., Gesteland, R. F., Atkins, J. F., Mullenbach, G. T., Hill, K. E., Wahba, A. J., and Hatfield, D. (1994) *J. Biol. Chem. 269*, 29739−29745.
21. Hatfield, D. (1972) *Proc. Natl. Acad. Sci. U.S.A. 67*, 3014−3018.
22. Simpson, L., and Emeson, R. B. (1996) *Annu. Rev. Neurosci. 19*, 27−52.
23. Smith, H. C., Gott, J. M., and Hanson, M. R. (1997) *RNA 3*, 1105−1123.
24. Sharp, P. A. (1985) *Cell 42*, 397−400.
25. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) *Annu. Rev. Biochem. 55*, 1119−1150.
26. Sharp, P. A. (1988) *JAMA, J. Am. Med. Assoc. 260*, 3035−3041.
27. Ruby, S. W., and Abelson, J. (1991) *Trends Genet. 7*, 79−85.
28. Shao, Y., and Kent, S. B. H. (1997) *Chem. Biol. 4*, 187−194.
29. Perler, F. B., Olsen, G. J., and Adam, E. (1997) *Nucleic Acids Res. 25*, 1087−1093.
30. Hatfield, D., and Diamond, A. (1993) *Trends Genet. 9*, 69−70.

BI981042R