BIOINFORMATICS DISCOVERY NOTE

Vol. 16 no. 5 2000
Pages 478–481

A homolog of mammalian antizyme is present in fission yeast Schizosaccharomyces pombe but not detected in budding yeast Saccharomyces cerevisiae

Chang Zhu[1], Kevin Karplus[2], Leslie Grate[2] and Philip Coffino[1,3,*]

[1]Department of Microbiology and Immunology, [2]Department of Computer Engineering, Jack Baskin School of Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064, USA and [3]Department of Medicine, University of California, San Francisco, San Francisco, CA 94143-0414, USA

Received on November 17, 1999; revised on January 6, 2000; accepted on February 21, 2000

Abstract

Motivation: *The antizymes (AZ) are proteins that regulate cellular polyamine pools in metazoa. To search for remote homologs in single-celled eukaryotes, we used computer software based on hidden Markov models. The most divergent homolog detected was that of the fission yeast* Schizosaccharomyces pombe. *Sequence identities between* S.pombe *AZ and known AZs are as low as 18–22% in the most conserved C-terminal regions. The authenticity of the* S.pombe *AZ is validated by the presence of a conserved nucleotide sequence that, in metazoa, promotes a* +1 *programmed ribosomal frameshift required for AZ expression. However, no homolog was detected in the completed genome of the budding yeast* Saccharomyces cerevisiae. *Procedural details and supplementary information can be found at http://itsa.ucsf.edu/~czhu/AZ.*

Contact: *pcoffin@itsa.ucsf.edu*

Introduction

Antizyme (AZ) was first identified in mammalian cells as a protein inhibitor of the enzyme ornithine decarboxylase (ODC) (reviewed in Hayashi, 1989; Coffino, 1998). Surprisingly, AZ is translated from two overlapping open reading frames (ORF), an ORF1 terminated by an UGA stop codon and an ORF2 lacking an AUG initiation codon (Matsufuji *et al.*, 1995). Translation of the functional protein is achieved by a +1 programmed ribosomal frameshift, which is stimulated by elevated polyamines. Antizyme activities are thus responsive to cellular polyamine levels. Antizyme inactivates ODC by dissociating the enzymatically active ODC dimer (Murakami *et al.*, 1985; Li and Coffino, 1992) and targeting ODC degradation by the 26S proteasome (Murakami *et al.*, 1992). In addition it inhibits polyamine transport into cells (Mitchell *et al.*, 1994; Suzuki *et al.*, 1994).

It is now clear that AZ is ubiquitous in vertebrates and present in multiple non-allelic copies. Recently homologs of vertebrate AZs have been reported in the fruit flies *Drosophila melanogaster* and *D.virillis* (Ivanov *et al.*, 1998; Salzberg *et al.*, 1996) and the nematode *Caenorhabditis elegans* (Wilson *et al.*, 1994), suggesting that AZ is widespread in metazoa. These genes share with vertebrate AZs similarities not only in amino acid sequences but also in structure. They have two overlapping ORFs and require a +1 (or −2) frameshift for translation.

Although ODC-inhibitory proteins exist in the prokaryote *Escherichia coli*, they are not mammalian AZ homologs and play no role in polyamine feedback regulation (Canellakis *et al.*, 1993). It is not known whether AZ is present in single-celled eukaryotes. In the filamentous fungus *Neurospora crassa*, polyamines decrease ODC activity by accelerating degradation (Barnett *et al.*, 1988) and also by reducing the ODC mRNA level (Williams *et al.*, 1992; Pitkin *et al.*, 1994). In the budding yeast *Saccharomyces cerevisiae*, ODC is negatively regulated by polyamines (Tyagi *et al.*, 1981; Toth and Coffino, 1999). The rate of ODC degradation increases and the enzyme level falls (Toth and Coffino, 1999). This response is inhibited by the protein synthesis inhibitor cycloheximide, suggesting a requirement for the synthesis of a protein with AZ-like function. Biochemical approaches have, however, failed to find direct evidence of polyamine-induced synthesis of an inhibitory protein that associates with *S.cerevisiae ODC* (Toth and Coffino, 1999).

*To whom correspondence should be addressed.

© Oxford University Press 2000

The mechanisms of ODC regulation in microorganisms remain to be determined. Since the presence of AZ suggests the conservation of ODC regulation, we wish to find the evolutionary origin of AZ. We set out to search for an AZ gene in *S.cerevisiae* using a computational approach. We reasoned that if AZ is present in the now-completed *S.cerevisiae* genome, we should be able to find the sequence by using computer programs designed to search for remote homologs. Hidden Markov Models (HMM) are among the most sensitive and robust computational tools designed for seeking remote homologs (Amitai, 1998; Krogh *et al.*, 1994; Hughey and Krogh, 1996). We used a revised version of the software SAM-T98 that has been reported to find more true homologs than other leading software (Karplus *et al.*, 1998; Park *et al.*, 1998). The frameshift signal gives us an independent criterion to validate whether a positive hit is an AZ. To our surprise, we found that AZ could be recognized in the fission yeast *Schizosaccharmoyces pombe* but not in the budding yeast *S.cerevisiae*.

Results and discussion

BLAST 2.0 or Psi-BLAST search of the GenBank NR (Non-Redundant) protein database with an *E* value cut-off of 0.001 failed to detect any AZ-like genes in bacteria or single-celled eukaryotes. The large number of AZs with known sequences allowed us to search for additional proteins with remote sequence similarities to vertebrate AZs using a more sensitive computation tool, SAM-T98. SAM-T98 is an HMM-based search algorithm specifically designed for remote protein homologies (Karplus *et al.*, 1998; Park *et al.*, 1998). It creates an HMM automatically starting with a single query sequence. An HMM model was built using mouse AZ1 as the seed. The model was then used to search the NR database. It identified an unnamed ORF (GenBank Accession #BAA13889) (Yoshioka *et al.*, 1997) of 172 amino acids from the fission yeast *S.pombe* with sequence similarities to all known AZ proteins, and with highest conservation for the most conserved residues. Figure 1 shows the resulting sequence alignment generated by the HMM. Among previously described AZs, those in nematodes most closely resemble the putative *S.pombe* AZ. Using BLAST search alone, it would have been difficult to identify BAA13889 as a worthy AZ candidate, since sequence identities between BAA13889 and known AZs are as low as 18–22% in the most conserved regions. BAA13889 has *E* values greater than 30 when compared with human or *C.elegans* AZs, far in excess of the 0.001 threshold value employed in our naive BLAST search. Although BAA13889 and known AZs share low sequence identities, the sequence elements most highly conserved in the *S.pombe* ORF are those most generally conserved in the C-terminal portion of AZ (Figure 1). This region, approximately half of the molecule, is sufficient for binding to ODC and is required for the other known functions: ODC degradation and inhibition of polyamine transport (Li and Coffino, 1994).

Vertebrates and invertebrates AZs share not only homologous protein sequences but also homologous gene structure. In general, AZs have two overlapping ORFs, a short ORF1 with a translation start codon and a longer ORF2 lacking an initiation codon. Relative to ORF1, ORF2 is in +1 reading frame; thus translation of the whole protein requires a +1 (or −2) ribosomal frameshift. One exception could be the silkworm AZ: EST sequences predict that ORF1 and ORF2 are in the same reading frame, but this could be due to a sequencing mistake. Conservation of frameshifting is further supported by the conservation of nucleotide sequences surrounding the UGA stop codon of ORF1: a region of approximately 30 nucleotides flanking the UGA is highly conserved for all AZs with known sequences (Figure 2). In addition, a pseudoknot structure that stimulates frameshifting is present in all vertebrates AZs examined except for AZ3.

If the *S.pombe* ORF BAA13889 is an AZ, it should be expected to share a requirement for frameshifting. We therefore examined the nucleotide sequences for frameshift signals. As shown in Figure 2, sequences similar to known frameshift sites were found at the beginning of the *S. pombe* ORF BAA13889. There are overlapping ORFs, with the 3′ ORF (BAA13889, ORF2) in the +1 frame relative to the 5′ ORF (ORF1). Translation of both ORFs through a +1 frameshift results in a protein of 226 amino acids, a size similar to known AZs and slightly improves its similarities to other AZs. Based on the sequence similarities to mammalian AZ and the frameshift signal, we conclude that we have identified the *S.pombe* AZ. Biochemical data has confirmed the ODC-inhibitory activity of the *S.pombe* ORF BAA13889 protein (personal communication, I.P.Ivanov, S.Matsufuji, Y.Murakami, R.F.Gesteland, and J.Atkins, submitted).

The *S.pombe* AZ is unique in that ORF2 has an AUG initiation codon 5′ of the presumptive frameshift site, which raises the question whether frameshifting is required for its expression and function. Two sequence features suggest that it is. First, the conserved frameshifting signal is present in the *S.pombe* AZ at the right position: it is near the predicted N-terminus where frameshifting occurs in other AZ genes and a +1 frameshift aligns ORF1 and ORF2. Second, translation initiation from ORF2, if possible, will be inefficient. Of the 13 codons 5′ of the frameshift site in ORF2, six are rare codons and four are rarest (with usage frequencies 3–5.3 per thousand compared with an average of 16 per thousand).

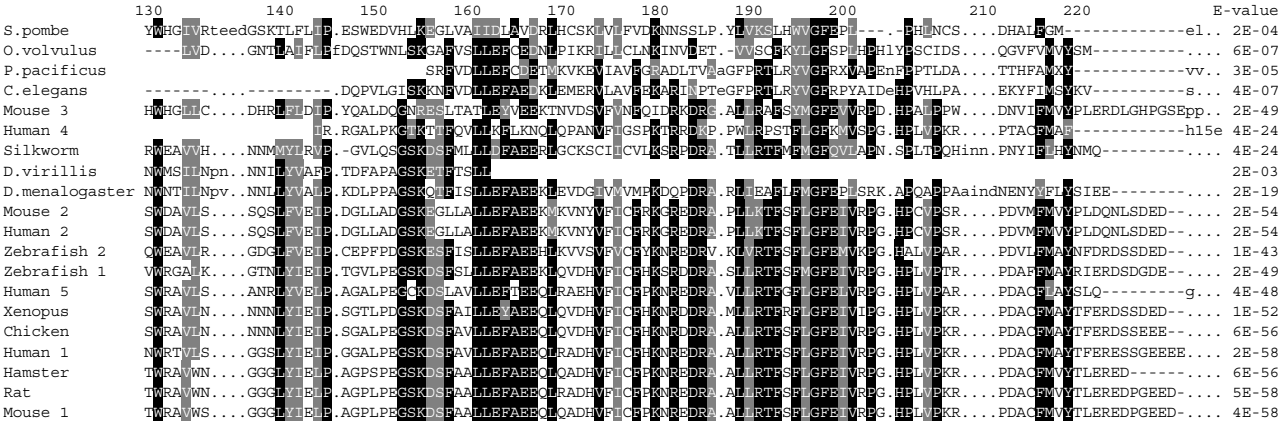

Fig. 1. Multiple alignment of AZ protein sequences generated by the HMM. Only the most conserved carboxy-terminal region is shown. Numbers on top correspond to amino acid positions for human AZ1. Aligned sequences are in upper case, and unaligned sequences are in lower case. Gaps are indicated by dashes and unknown sequences are denoted by blank spaces. Amino acids are shaded if at least 14 out of 20 are conserved with identities shaded in dark and similarities shaded in light background. E-values for each sequence are listed on the right and are calculated using a training set that excluded *S.pombe* AZ and the NR database size of 408 494 sequences.

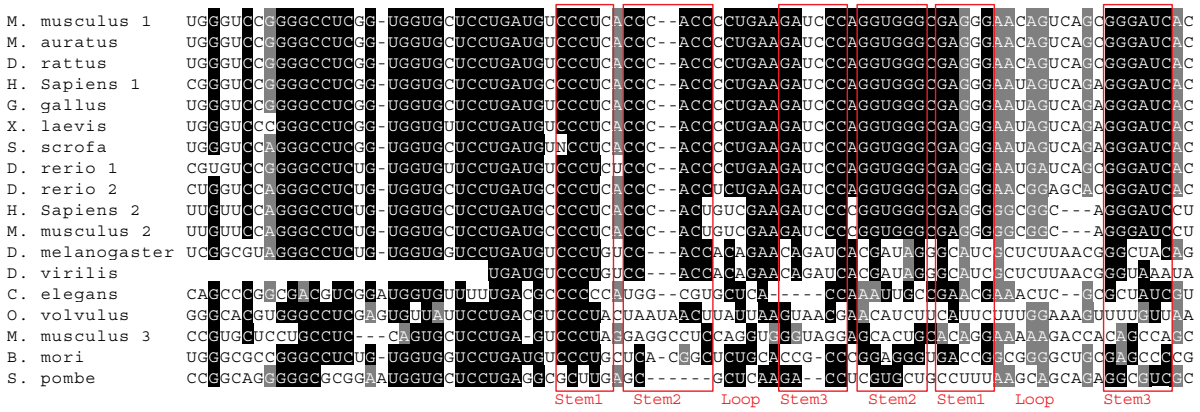

Fig. 2. Conservation of frameshifting in AZs. Multiple sequence alignments of nucleotide sequences flanking ORF1 stop codon UGA. Nucleotides are shaded if at least 14 out of 20 are conserved in the alignments with identities shaded in dark and similarities shaded in light background. Boxes show positions forming the conserved pseudoknot secondary structure in most vertebrate AZs (Matsufuji *et al.*, 1995), a structure which is absent in the AZs of lower eukaryotes.

The vertebrate AZ1 mRNA pseudoknot secondary structure (Matsufuji *et al.*, 1995) is analogous to secondary structures present in a number of -1 frameshifting genes. This pesudoknot immediately follows the stop codon UGA of ORF1 and stimulates AZ1 frameshifting by approximately 2.5-fold *in vitro* and even more in the yeast *S.cerevisiae*, presumably by positioning ribosomes for frameshifting (Matsufuji *et al.*, 1996). In invertebrate AZs, the pseudoknot secondary structure is absent. The secondary structure is also absent in the fission yeast *S.pombe* AZ, suggesting that this structure evolved late in evolution. Consistent with this hypothesis, deletion of the secondary structure in rat AZ1 diminishes but does not abolish frameshifting (Matsufuji *et al.*, 1996). Surprisingly, the same HMM that recognized *S.pombe* AZ failed to identify an AZ homolog in the now-completed *S.cerevisiae* genome. Because of frameshifting, *S.cerevisiae* AZ might not be recognized as a continuous ORF. We therefore identified all stop codons (UGA, UAA, UAG) in the *S.cerevisiae* genome, assuming possible $+1$, $+2$, or $+3$ frameshifting, and conceptually translated nucleotide sequences $5'$ and $3'$ of the stop codons to create a pseudo-orf database. The same HMM also failed to identify an AZ homolog in the peusdo-orf

database. In *S.cerevisiae*, the properties of the regulatory response of ODC to polyamines are functionally similar to those in mammalian cells (Toth and Coffino, 1999), suggesting the presence of a protein that plays an AZ-like role. Having failed to detect the presence of such a putative protein, we conclude that AZ is paralogous in *S.cerevisiae* or, alternatively that it is orthologous but is sufficiently diverged to escape detection by the methods employed.

Acknowledgments

This work was supported by Public Health Service grant GM45335 to P.C. from the National Institute of General Medical Sciences, NIH, and by the National Science Foundation Grants DBI 9808007 and EIA 9905322 to K.K.

References

Amitai,M. (1998) *Science*, **282**, 1436–1437.

Barnett,G.R., Manouchehr,S. and Davis,R.H. (1988) *J. Biol. Chem.*, **263**, 10005–10008.

Canellakis,E.S., Paterakis,A.A., Huang,S.C., Panagiotidis,C.A. and Kyriakidis,D.A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7129–7133.

Coffino,P. (1998) In Peters,J.-M. and Finley,J.R.H.D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, pp. 411–427.

Hayashi,S.I. (1989) *Ornithine Decarboxylase: Biology, Enzymology and Molecular Genetics*. Pergamon Press, New York.

Hughey,R. and Krogh,A. (1996) *Comput. Appl. Biosci.*, **12**, 95–107.

Ivanov,I.P., Simin,K., Letsou,A., Atkins,J.F. and Gesteland,R.F. (1998) *Mol. Cell. Biol.*, **18**, 1553–1561.

Karplus,K., Barrett,C. and Hughey,R. (1998) *Bioinformatics*, **14**, 846–856.

Krogh,A., Brown,M., Mian,I.S., Sjolander,K. and Haussler,D. (1994) *J. Mol. Biol*, **235**, 1501–1531.

Li,X. and Coffino,P. (1992) *Mol. Cell. Biol.*, **12**, 3556–3562.

Li,X. and Coffino,P. (1994) *Mol. Cell. Biol.*, **14**, 87–92.

Matsufuji,S., Matsufuji,T., Miyazaki,Y., Murakami,Y., Atkins,J.F., Gesteland,R.F. and Hayashi,S.I. (1995) *Cell*, **80**, 51–60.

Matsufuji,S., Matsufuji,T., Wils,N., Gesteland,R. and Atkins,J. (1996) *EMBO J.*, **15**, 1360–1370.

Mitchell,J.L.A., Judd,G.G., Bareyal-Leyser,A. and Ling,S.Y. (1994) *Biochem. J.*, **299**, 19–22.

Murakami,Y., Fujita,K., Kameji,T. and Hayashi,S. (1985) *Biochem. J.*, **225**, 689–697.

Murakami,Y., Matsufuji,S., Kameji,T., Hayashi,S., Igarashi,K., Tamura,T., Tanaka,K. and Ichihara,A. (1992) *Nature*, **360**, 597–599.

Park,J., Karplus,K., Barrett,C., Hughey,R., Haussler,D., Hubbard,T. and Chothia,C. (1998) *J. Mol. Biol.*, **284**, 1201–1210.

Pitkin,J., Perriere,M., Kanehl,A., Ristow,J.L. and Davis,R.H. (1994) *Arch. Biochem. Biophys.*, **315**, 153–1610.

Salzberg,A., Golden,K., Bodmer,R. and Bellen,H. (1996) *Genetics*, **144**, 183–196.

Suzuki,T., He,Y., Kashiwagi,K., Murakami,Y., Hayashi,S.I. and Igarashi,K. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 8930–8934.

Tyagi,A.K., Tabor,C.W. and Tabor,H. (1981) *J. Biol. Chem.*, **256**, 12156–12163.

Toth,C. and Coffino,P. (1999) *J. Biol. Chem.*, **274**, 25921–25926.

Williams,L.J., Barnett,G.R., Ristow,J.L., Pitkin,J., Perriere,M. and Davis,R.H. (1992) *Mol. Cell. Biol.*, **12**, 347–359.

Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J. *et al.*, (1994) *Nature*, **368**(6466), 32–38.

Yoshioka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H. (1997) *DNA Res.*, **4**, 363–369.