854–858 *Nucleic Acids Research, 1999, Vol. 27, No. 3*

The two subunits of human molybdopterin synthase: evidence for a bicistronic messenger RNA with overlapping reading frames

Joan Sloan[1], James R. Kinghorn[1,2] and Shiela E. Unkles[1,*]

[1]Department of Microbiology, Monash University, Clayton, Victoria 3168, Australia and
[2]School of Environmental and Evolutionary Biology, University of St Andrews, St Andrews, Fife KY16 9TH, UK

Received August 31, 1998; Revised November 6, 1998; Accepted December 4, 1998 DDBJ/EMBL/GenBank accession no. AF117815

ABSTRACT

Molybdoenzymes are ubiquitous and require a prosthetic group called the molybdenum cofactor for activity. We provide evidence here that the two heteromeric subunits (MOCO1-A and MOCO1-B) of human molybdopterin synthase, which is involved in the conversion of precursor Z to molybdopterin in the molybdenum cofactor biosynthetic pathway, are specified by a single bicistronic mRNA with overlapping reading frames. The transcript is in low abundance and shows variable tissue distribution. We propose that leaky scanning of the first translational initiation codon for *MOCO1-A* by 40S ribosomal subunits occurs, allowing recognition of the AUG for the downstream *MOCO1-B* reading frame. Such a genetic arrangement may result in a constant ratio and close proximity of lowly expressed enzyme subunits which should, a priori, be especially advantageous for assembly in complex mammalian cells. The *MOCO1* locus resides on human chromosome 5.

INTRODUCTION

Molybdoenzymes play essential roles in the carbon, sulphur and nitrogen cycles in most organisms. In humans, sulphite oxidase is required for the breakdown of sulphur amino acids, methionine and cysteine (1). In the fungus *Aspergillus nidulans*, a model eukaryote, another molybdoenzyme nitrate reductase is required for the important ecological process of nitrate assimilation (reviewed in 2). Additionally, a few molybdoenzymes exist in both eukaryotic groups, including xanthine dehydrogenase which is important in the catabolic process of purine breakdown to uric acid (3,4).

The molybdenum cofactor, a prosthetic group which consists of a novel pterin called molybdopterin linked by its 6-alkyl sidechain to a dithiolene group which coordinates molybdenum, is required for the catalytic activity of these enzymes. Its chemical structure and its likely biosynthetic pathway was proposed by Rajagopalan (reviewed in 5). Little information is available on the molecular biology of the biosynthesis of this cofactor in eukaryotes, although its presence has been demonstrated indirectly in a variety of biological material such as cow milk, rabbit and fowl liver and *A.nidulans* (6,7).

Pleiotropic loss of human molybdoenzymes, including sulphite oxidase and xanthine dehydrogenase, results in a severe clinical disease for which no known therapy exists (8,9). We describe here the molecular characteristics of the human locus encoding the small and large subunit of molybdopterin synthase which is involved in the conversion of precursor Z to molybdopterin (Fig. 1), and its unusual genetic organisation. The results of these studies should aid further investigation of the genetic basis of molybdenum cofactor deficiency in humans.

MATERIALS AND METHODS

cDNA clones and DNA sequencing

Following identification of human sequences in the Expressed Sequence Tag Database by deduced amino acid sequence comparison with *A.nidulans* CnxH using tBlastn (10), cDNA clones (ATCC 960768 from adult uterus and ATCC 331184 from fetal liver and spleen) were purchased from ATCC. The DNA sequence of clones was determined in both strands by automated DNA sequencing using an ABI 373 A fluorescent sequencing apparatus and a PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems). Sequences were assembled using Sequencher (Gene Codes Corp.). Neither clone is full-length, ATCC 331184 lacking 373 5′ nucleotides and ATCC 960768 lacking 12 nt from the 5′-end.

Southern blot and PCR analysis

Human genomic DNA was isolated from peripheral blood leukocytes using a Nucleon BACC2 Kit (Amersham) following the manufacturer's instructions. A Southern blot prepared with restriction endonuclease-digested DNA was probed at high stringency (11) using a 520 bp *Nco*I–*Eco*RV fragment of clone ATCC 331184. To show the absence of introns in the region of overlap of the ORFs, PCR amplification of 100 ng *Eco*RI- or *Bam*HI-digested genomic DNA was carried out using 0.4 μM primers D1 (5′-ACTCGACATCCTGGATTGGC-3′) and gene-specific primer GSP1 (5′-TGCACCACAGAGCGGAG-3′), 1.5 mM $MgCl_2$, 0.1 mM dNTPs and 2.5 U *Taq* polymerase (Boehringer Mannheim) for 30 cycles of 94°C for 30 s, 55°C for 20 s, 72°C for 30 s.

*To whom correspondence should be addressed. Tel: +61 3 9905 4323; Fax: +61 3 9905 4811; Email: shiela.unkles@med.monash.edu.au

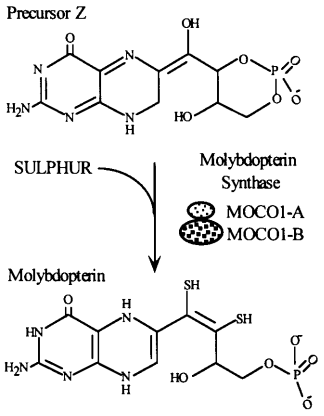

Figure 1. The conversion of precursor Z to molybdopterin during the biosynthesis of the molybdenum cofactor required for molybdoenzymes. The small and large subunits in humans, MOCO1-A and MOCO1-B, respectively, together form molybdopterin synthase which adds sulphur to precursor Z.

Northern analysis and 5′-RACE

Human adult northern blots were purchased from Clontech, hybridised at high stringency as previously described (11) and washed with 1× SSC at 65°C. The probe was synthesised by PCR using ATCC 960768 as template and primers 7338 (5′-GCCAAGAATTCGGCACGAGG-3′) and 7350 (5′-AAACAGAATTCATTAACTGTTGGATG-3′) to give an 800 bp fragment encompassing only the coding regions of *MOCO1-A* and *MOCO1-B*. Autoradiographic exposure times at –70°C were 72 h for *MOCO1* with Kodak Biomax MS film and 2 h for actin with Fuji RX film. mRNA from adult heart was purchased from Clontech and the 5′-end of the *MOCO-1* transcript determined using a 5′ RACE System for Rapid Amplification of cDNA Ends v.2.0 (Life Technologies) and GSP1 with the Abridged Anchor Primer (AAP) for the first round PCR and nested primer GSP2 (5′-TCTCCAGGCTGAAGCACGAGG-3′) or GSP3 (5′-CTCTGAACGAACTCCTG-3′) with the Abridged Universal Amplification Primer (AUAP) for second round PCR. Conditions for first round PCR were 30 cycles of denaturation for 30 s at 94°C, annealing for 20 s at 54°C and elongation for 1 min at 72°C. For second round PCR with GSP2 and AUAP, conditions were 30 cycles of denaturation for 30 s at 94°C, annealing for 20 s at 65°C and elongation for 1 min at 72°C, and with GSP3 and AUAP were 30 cycles of denaturation for 30 s at 94°C, annealing for 20 s at 52°C and elongation for 1 min at 72°C. Reaction mixtures followed recommendations by Life Technologies. The procedure was repeated using thermostable reverse transcriptase (C. therm. Polymerase; Roche) to allow first strand cDNA synthesis at 60°C using GSP2. For first round PCR, GSP2 and AAP were used with GSP3 and AUAP for second round PCR. Fragments of ~260 and 100 bp obtained with GSP2 and AUAP and GSP3 and AUAP, respectively, were gel purified and the products sequenced as above from primers GSP2 or GSP3.

Fungal transformation and analysis

The transformation procedure was essentially that described previously (12) with selection for transformants on osmotically stabilised minimal medium containing 10 mM sodium nitrate as the sole source of nitrogen. The transforming DNA was a mixture of 5 µg *MOCO1* cDNA clone ATCC 960768 with 1 µg pHELP, a plasmid which promotes autonomous replication, greatly enhances transformation frequencies in *A.nidulans* (12) and possesses gratuitous promoter activity (13). Putative transformants were purified by subculture on selective medium without osmotic stabiliser. Mycelia were grown in liquid cultures for 16–18 h at 25°C, harvested by filtration and genomic DNA prepared using a Nucleon BACC2 Kit (Amersham) following grinding of the mycelium in liquid nitrogen. DNA from human blood was obtained using the same kit and following the manufacturer's instructions. Southern blotting and hybridisation were as described (11) using as probe the same PCR fragment as that described above for northern hybridisation.

RESULTS

Sequence and organisation of *MOCO1*

A full-length human cDNA (designated *MOCO1*) was sequenced, which contains open reading frames for two proteins, MOCO1-A and MOCO1-B (Fig. 2A) with high amino acid similarity to the small and large subunits, respectively, of the enzyme molybdopterin synthase (below). The 5′-end of the transcript was determined by RACE following first strand cDNA synthesis at either 42 or 60°C, the latter to reduce possible mRNA secondary structure effects. DNA sequence determination of PCR products yielded the same end sequence regardless of the first strand synthesis conditions or the primer sets used (S.E.Unkles, unpublished results). Twenty-eight nucleotides from the transcript 5′-end (Fig. 2B) lies the open reading frame (on ORF 1) for the inferred MOCO1-A protein. Overlapping this by 80 nt (on ORF 2) is the coding sequence for the larger subunit of molybdopterin synthase. Verification that the original clone was not the result of reverse transcriptase error or another cloning artefact resulting in a change in the reading frame was obtained by comparison with the sequences of further independent cDNA clones (S.E.Unkles, unpublished data). In addition, the pattern of Southern hybridisation at high stringency to human genomic DNA digested with several restriction endonucleases with a probe spanning the two ORFs is indicative of a single gene (S.E.Unkles, unpublished data), the location of which has been mapped by UniGene cluster analysis to chromosome 5, interval D5S628–D5S474 (14). Finally, PCR analysis of human genomic DNA (S.E.Unkles, unpublished data) demonstrates the absence of introns between nt 154 and 324 (Fig. 2B), making unlikely any mechanism involving intron splicing in the formation of this transcript.

Protein sequence comparison of MOCO1-A and MOCO1-B with CnxG and CnxH

The MOCO1-A protein, of molecular size 9.7 kDa, shows considerable similarity with lower eukaryotic *A.nidulans* CnxG (size 9.6 kDa; S.E.Unkles, unpublished data). Overall the deduced MOCO1-A and CnxG proteins have 35% identical residues (Fig. 3A) and similarity of 64%. The second open reading frame encodes the 20.9 kDa MOCO1-B protein which shows an overall identity of 31% (Fig. 3B) and 55% similarity with the corresponding fungal CnxH protein (size 21.6 kDa; S. E. Unkles, unpublished data). In *A.nidulans*, the *cnxG* and *cnxH* genes encoding the two subunits are genetically unlinked to each other (2). Essential amino acid residues (boxed in Fig. 3) in

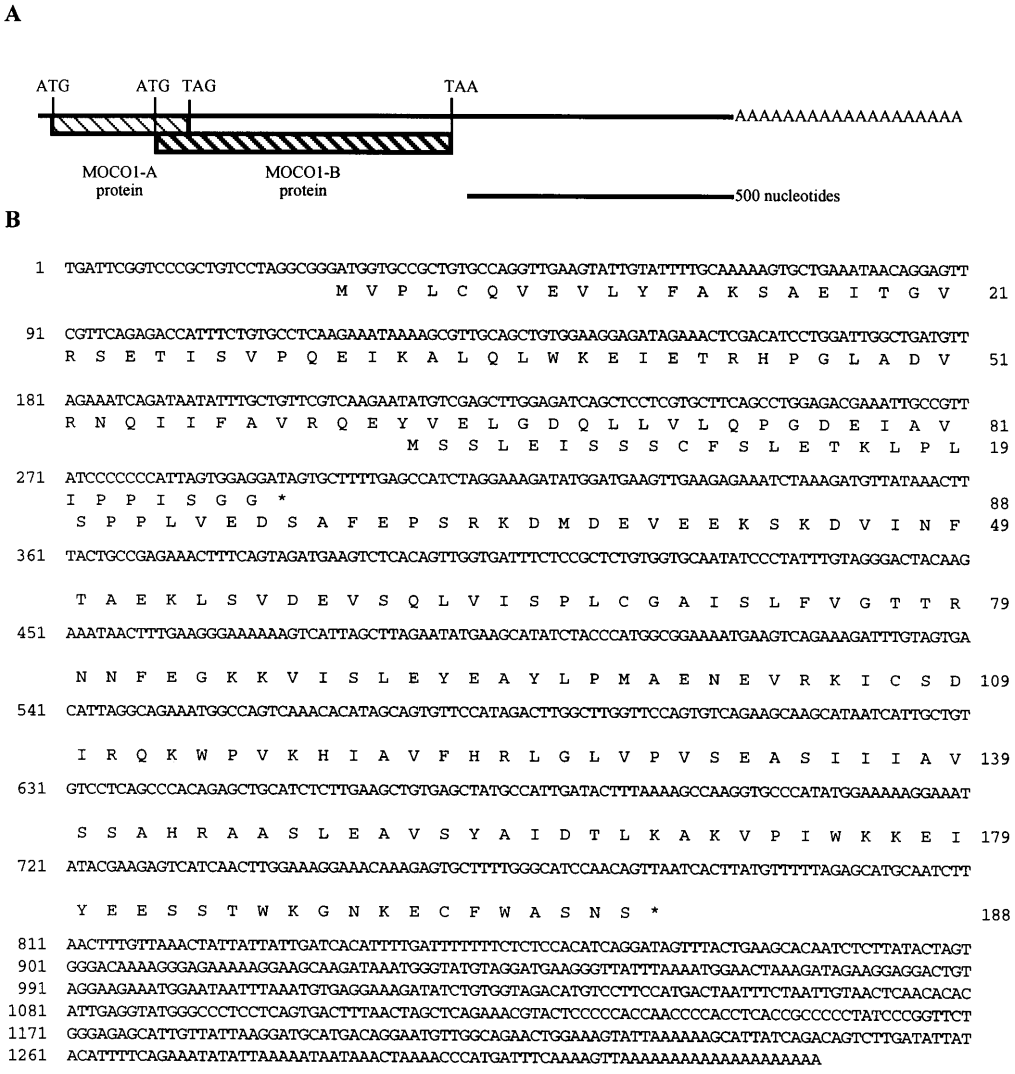

Figure 2. The genetic organisation of the *MOCO1* transcript (**A**). The relative positions of the ORFs for MOCO1-A and MOCO1-B proteins with the translational start and stop codons are shown. DNA sequence and deduced amino acid sequences of ORFs 1 and 2 for MOCO1-A and MOCO1-B, respectively (**B**). Numbers to the left refer to nucleotides and those on the right to amino acid residues.

the fungal proteins have been determined by mutational analysis and these are also present in the human counterparts.

Complementation of mutants of the lower eukaryote *A.nidulans*

Inspection of nucleotides in the neighbourhood of the proposed translational initiation for human *MOCO1-A* reveals an optimal translational context of GGGAUGGU with G at position –3 and G at position +4. This strong context would be expected to prevent leaky scanning by the ribosomes and therefore initation of *MOCO1-B* (15–17). To test if the downstream ORF encoding MOCO1-B is translated and functional, i.e. that leaky scanning takes place, human *MOCO1* was transformed into a mutant strain, *cnxH3*, of the fungus *A.nidulans*. The *A.nidulans cnxH* gene encodes the large subunit of the molybdopterin synthase and the *cnxH3* mutant lacks the ability to make the molybdenum cofactor required for activity of the enzyme nitrate reductase, loss of which results in the inability of the organism to grow on nitrate as a sole nitrogen source. Selection for transformants can be, therefore, conveniently achieved by the restoration of growth on nitrate as a nitrogen source in *cnxH* mutants. The results in Figure 4a show phenotypic complementation of *cnxH3* by human *MOCO1* following selection on nitrate. Confirmation that these were bona fide *MOCO1*-transformed *cnxH* strains was carried out by Southern blot (Fig. 4b) where a *MOCO1*-hybridising fragment was seen in DNA isolated from transformants (lanes 1–6) but not in DNA from the wild-type *A.nidulans* (lane 7).

Expression of *MOCO1* in human tissue

The size and tissue distribution of *MOCO1* mRNA was analysed by northern hybridisation (Fig. 5). The *MOCO1* probe, spanning the two open reading frames for proteins MOCO1-A and MOCO1-B, reveals a single transcript of approximately the expected size of 1.35 kb which can be detected at varying levels in mRNA from all tissues sampled. The transcript is present in greatest abundance in heart and skeletal muscle, at lower levels

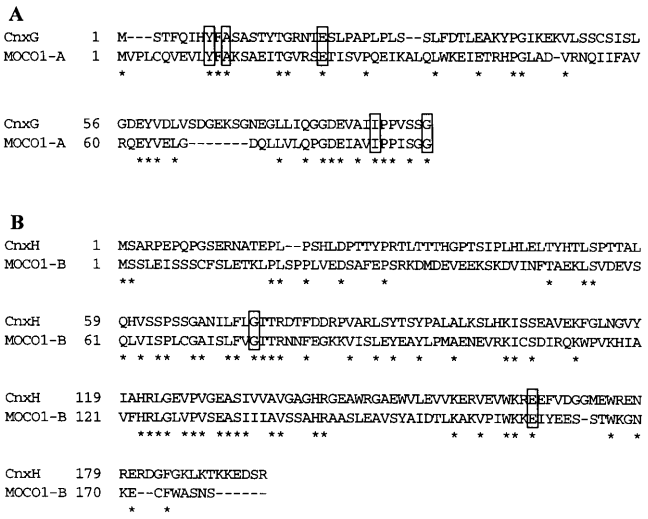

Figure 3. Amino acid comparisons of MOCO1-A with fungal CnxG (**A**) and MOCO1-B with fungal CnxH (**B**). Numbers to the left indicate amino acid residues and * below indicates identical residues. Alignments were obtained using ClustalW. Known essential amino acid residues are boxed.

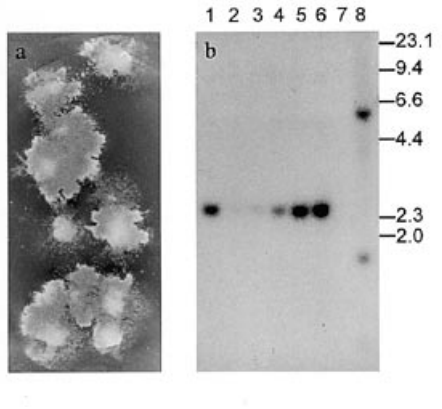

Figure 4. Functional expression of MOCO1-B. (**a**) Phenotypic complementation of the fungal (*A.nidulans*) *cnxH3* mutant (defective in the large subunit of molybdopterin synthase) to restore activity of the fungal molybdoenzyme nitrate reductase and hence permit growth of strain *cnxH3* on nitrate as the sole source of nitrogen. This is the original selection showing nitrate-utilising strains transformed with the *MOCO1* fragment together with the fungal replicating plasmid pHELP. No transformants were observed in the absence of *MOCO1*. (**b**) Southern blot of *Eco*RI-digested DNA from fungal nitrate-utilising *cnxH3* transformants (lanes 1–6) with a DNA fragment from the *MOCO1* gene as the probe. Lane 7 contains *A.nidulans cnxH3* DNA as negative control. Lane 8 contains human DNA provided by one of the authors (S.E.Unkles). *MOCO1* is present in low copy number in human DNA, at least under the high stringency conditions employed.

in brain, kidney and pancreas and is barely detectable in lung and peripheral blood leukocytes. This may reflect the distribution of molybdoenzymes within tissues. It is noteworthy that the *MOCO1* transcript requires a substantially longer autoradiographic exposure than the control β-actin suggestive of a low abundance mRNA. A faintly hybridising band of ~4 kb was also observed in some tissues with the *MOCO1* probe, but as the actin probe hybridised weakly to a band of the same size (i.e. the weak bands were

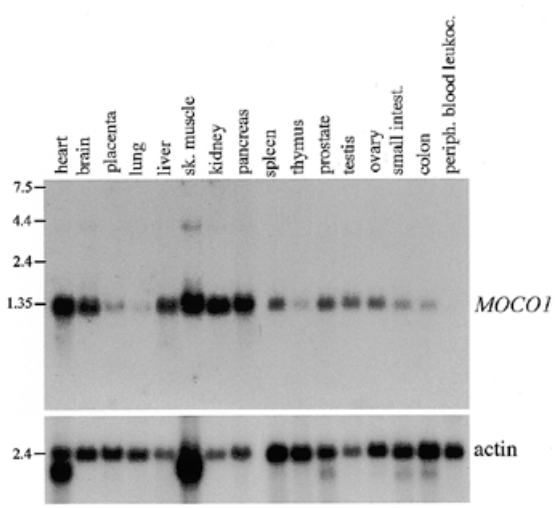

Figure 5. Transcript size and tissue expression of *MOCO1*. Human adult multiple tissue northern blots containing 2 µg poly(A)$^+$ mRNA in each lane were hybridised at high stringency using an 800 bp ^{32}P-labelled PCR fragment which spans both ORFs as probe. The size of the transcript (~1.35 kb) is indicated. Human β-actin was used as a control for equality of RNA concentration in each tissue.

superimposable) it is likely that this is due to non-specific hybridisation.

DISCUSSION

With the exception of certain viruses and retrotransposon elements, eukaryotic genes are generally transcribed to give a monocistronic mRNA which specifies a single protein or isoform variants thereof. Reports in eukaryotes of full-length proteins being translated from a single cellular bicistronic mRNA are rare (18,19) but it has been demonstrated that such transcripts can arise by differential exon splicing (19). A single mammalian transcript with tandem reading frames encoding two proteins known to function in dentin mineralisation has recently been described (20). In this study, we show that the subunits of the human enzyme molybdopterin synthase are also encoded as a single transcriptional unit but that the reading frames overlap by 80 nt. Likewise, comparison of the human cDNA to the Expressed Sequence Tag Database shows that this arrangement exists also in the mouse and rat, and most likely in the fruit fly, such conservation between species strongly suggesting that the bicistronic arrangement is not an artefact and that probably both proteins are synthesised *in vivo*. In addition, PCR analysis indicates that at the genomic DNA level the reading frames are similarly overlapping and, therefore, the bicistronic transcript does not arise from differential splicing. By contrast, in the lower eukaryote *A.nidulans*, the molybdopterin synthase subunits are encoded by unlinked genes (2) transcribed to give independent mRNAs (S.E.Unkles, unpublished data).

The strong initiation context of the first AUG for the small molybdopterin synthase subunit would at first appear to preclude the possibility of expression of the large subunit protein on a different reading frame. However the 5′-untranslated region length of the *MOCO1* transcript is very short, only 27 nt, and has the potential for formation of a small stem–loop structure (ΔG = –13.5 kcal/mol)

between nt –13 and +2. A stem–loop in such close proximity (7 nt) to the 5′-end of the mRNA in combination with the short untranslated region could interfere with the formation of an initiation complex (21). Therefore, even with a favourable initiation context, a proportion of the 40S ribosomal subunits may bypass this first *MOCO1-A* AUG and continue scanning (22). Initiation would then occur at the downstream *MOCO1-B* AUG. Cap-independent translation, recognised to function in certain cellular mRNAs (23,24), is also a possibility to permit *MOCO1-B* initiation although prediction from sequence alone of an internal ribosome entry site (IRES) is difficult. An alternative to the leaky scanning or IRES mechanisms is that of frameshifting, described in several examples of viral translation of overlapping reading frames, where ribosome pausing, often stimulated by the presence of RNA secondary structure, results in slippage at a specific site usually by –1 but occasionally +1. In *MOCO1*, a +1 slip would permit a change from the ORF encoding the small subunit to that encoding the large subunit, producing a single polypeptide. However, our recent mutational studies in *A.nidulans* have shown an absolute requirement for the C-terminal Gly residue of the small subunit to allow catalytic activity (S.E.Unkles, unpublished) and this would almost certainly be the case in humans also. Frameshifting would have to occur at the termination codon of the small subunit ORF and, following completion of the single protein, precise proteolysis would be necessary to release the catalytically active Gly. Frameshifting, therefore, is less likely than the leaky scanning or IRES mechanisms proposed above.

The question arises as to why an apparently lowly expressed gene should have assumed this unusual organisation in humans. Bicistronic messages are probably a means of minimising genome size in viruses but such a constraint would seem unnecessary in mammals. Instead, they may provide the means by which the cell can exert translational control over the synthesis of the different subunits of catalytic heteromeric dimers in order to achieve a constant 1:1 ratio of the two products and/or assure co-translation and close proximity for folding and assembly of the holoenzyme. Recently, it has been reported that another human gene, *MOCS1*, located on chromosome 6 encodes the two proteins catalysing the first steps in molybdenum cofactor biosynthesis, on a single transcript with tandem reading frames (25). Regardless of the reason for such transcripts with overlapping or tandem reading frames, the growing incidence of bicistronic mRNAs has implications for the analysis of genome sequences in higher eukaryotes.

ACKNOWLEDGEMENT

J.R.K. wishes to thank the Royal Society (London) for a grant to enable him to visit Australia.

REFERENCES

1 Garrett,R.M. and Rajagopalan,K.V. (1994) *J. Biol. Chem*., **269**, 272–276.
2 Cove,D.J. (1969) *Biol. Rev*., **54**, 291–303.
3 Ichida,K., Amaya,Y., Kamatani,N., Nishino,T., Hosoya,T. and Sakai,O. (1997) *J. Clin. Invest*., **99**, 2391–2397.
4 Glatigny,A. and Scazzocchio,C. (1995) *J. Biol. Chem*., **270**, 3534–3550.
5 Rajagopalan,K.V. (1996) In Neidhart,F.C. (ed.), *Escherichia coli and Salmonella Cellular and Molecular Biology.* ASM Press, Washington, DC, pp. 674–679.
6 Ketchum,P.A., Cambier,H.Y., Frazier,W.A., Madanski,C.H. and Nason,A. (1970) *Proc. Natl Acad. Sci. USA*, **66**, 1016–1023.
7 Garrett,R.G. and Cove,D.J. (1976) *Mol. Gen. Genet*., **147**, 179–186.
8 Bamforth,F.J., Johnson,J.L., Davidson,A.G.F., Wong,L.T.K., Lockitch,G. and Applegarth,D.A. (1990) *Clin. Biochem*., **23**, 537–542.
9 Mize,C., Johnson,J.L. and Rajagopalan,K.V. (1995) *J. Inherit. Metab. Dis*., **18**, 283–2909.
10 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol*., **215**, 403–410.
11 MacCabe,A.P., Riach,M.B.R., Unkles,S.E. and Kinghorn,J.R. (1990) *EMBO J*., **9**, 279–287.
12 Gems,D.H. and Clutterbuck,A.J. (1993) *Curr. Genet*., **24**, 520–524.
13 Unkles,S.E., Smith,J., Kana'n,G.J.M., Millar,L.J., Heck,I.S., Boxer,D.H. and Kinghorn,J.R. (1997) *J. Biol. Chem*., **272**, 28381–28389.
14 Deloukas,P., Schuler,G.D., Gyapay,G., Beasley,E.M., Sonderlund,C., Rodriguez-Tome,P., Hui,L., Matise,T.C., McKusick,K.B., Beckmann,J.S. *et al*. (1998) *Science*, **282**, 744–746.
15 Kozak,M. (1991) *J. Biol. Chem*., **266**, 19867–19870.
16 Kozak,M. (1992) *Annu. Rev. Cell Biol*., **8**, 197–225.
17 Kozak,M. (1995) *Proc. Natl Acad. Sci USA*, **92**, 2662–2666.
18 Lee,S.-J. (1991) *Proc. Natl Acad. Sci USA*, **88**, 4250–4254.
19 Szabo,G., Katarova,Z. and Greenspan,R. (1994) *Mol. Cell. Biol*., **14**, 7535–7545.
20 Ritchie,H. and Wang,L.-H. (1997) *Biochem. Biophys. Res. Commun*., **231**, 425–428.
21 Kozak,M. (1989) *Mol. Cell. Biol*., **9**, 5134–5142.
22 Kozak,M. (1991) *Gene Expr*., **1**, 111–115.
23 Le,S.-Y. and Maizel,J.V.,Jr (1997) *Nucleic Acids Res*., **25**, 362–369.
24 Yang,Q. and Sarnow,P. (1997) *Nucleic Acids Res*., **25**, 2800–2807.
25 Reiss,J., Cohen,N., Dorche,C., Mandel,H., Mendel,R.R., Stallmeyer,B., Zabot,M.-T. and Dierks,T. (1998) *Nature Genet*., **20**, 51–53.