REPORT

Diverse splicing mechanisms fuse the evolutionarily conserved bicistronic *MOCS1A* and *MOCS1B* open reading frames

TODD A. GRAY and ROBERT D. NICHOLLS

Department of Genetics, Case Western Reserve University School of Medicine, Center for Human Genetics, University Hospitals of Cleveland, Cleveland, Ohio 44106, USA

ABSTRACT

Molybdenum is an essential cofactor in many enzymes, but must first be complexed by molybdopterin, whose synthesis requires four enzymatic activities. The first two enzymes of this pathway are encoded by the *MOCS1* locus in humans. We describe here a remarkably well-conserved novel mRNA splicing phenomenon that produces both an apparently bicistronic *MOCS1A-MOCS1B* transcript, as well as a distinct class of monocistronic transcript. The latter are created by a variety of splicing mechanisms (alternative splice donors, alternative splice acceptors, and exon-skipping) to bypass the normal termination nonsense codon of *MOCS1A* resulting in fusion of the *MOCS1A* and *MOCS1B* open reading frames. Therefore, these “no-nonsense” transcripts encode a single bifunctional protein embodying both *MOCS1A* and *MOCS1B* activities. This coexpression profile was observed in vertebrates (human, mouse, cow, rabbit, opossum, and chicken) and invertebrates (fruit fly and nematode) spanning at least 700 million years of evolution. Our phylogenetic data also provide evidence that the bicistronic form of *MOCS1* mRNA is likely to only produce *MOCS1A* protein and, combined with Northern analyses, suggests that *MOCS1B* is translated only as a fusion with *MOCS1A*. Taken together, the data presented here demonstrate a very highly conserved and physiologically relevant dynamic splicing scheme that profoundly influences the protein-coding potential of the *MOCS1* locus.

Keywords: evolution; molybdenum; mRNA splicing; polycistronic; translation regulation

INTRODUCTION

To date, only a few cellular mammalian genes have been described as producing a bicistronic transcript with the potential to encode two independent proteins from adjacent open reading frames (ORFs) (Lee, 1991; Blumenthal, 1998; Reiss et al., 1998b; Gray et al., 1999; Sloan et al., 1999; Stallmeyer et al., 1999a). In humans, the molybdenum cofactor synthesis-step 1 (*MOCS1*) locus has been reported to produce two enzymes (*MOCS1A* and *MOCS1B*) from non-overlapping ORFs within a bicistronic transcript (Reiss et al., 1998b). These two enzymatic activities initiate the synthesis of an organic structure that chelates molybdenum (Mo) for use in the active sites of Mo-dependent enzymes. *MOCS1A* and *MOCS1B* orthologs have been identified in plants, fungi, and prokaryotes, and the *moaA* and *moaC* prod-

ucts of the *moa* operon catalyze these enzymatic steps in *Escherichia coli* (Hoff et al., 1995; Unkles et al., 1997; Reiss et al., 1998b). Mutations affecting the *MOCS1A* and *MOCS1B* ORFs have been identified in a subset of patients with molybdenum cofactor deficiency (Reiss et al., 1998b). Even though there is no in vivo evidence supporting the independent translation of *MOCS1A* and *MOCS1B* polypeptides (Reiss et al., 1998a, 1998b), it has been inferred that because mutations have been found in both ORFs, they are each independently translated (Reiss et al., 1998b).

The predominant scanning-ribosome model of translation in eukaryotes (Kozak, 1999), wherein the first suitable AUG codon encountered by the 40S ribosomal subunit serves as the site of translation initiation, cannot apply for downstream ORFs. Because an alternative to this scanning model need not be invoked for translation of the upstream *MOCS1A* ORF, only translation of *MOCS1B* necessitates alternative mechanisms. Ribosome reinitiation, in which 40S ribosomal subunits do not dissociate upon termination of the up-

Reprint requests to: Robert D. Nicholls, Center for Neurobiology and Behavior, University of Pennsylvania, Room 528, Clinical Research Building, 415 Curie Blvd., Philadelphia, PA 19104, USA.

stream ORF but continue tracking 3' until translation is reinitiated at the next suitable AUG codon (Kozak, 1999), has been proposed as the mechanism of translation of the downstream *MOCS1B* ORF (Reiss et al., 1998b). However, additional unconventional translational mechanisms can be envisaged that could produce *MOCS1B*, including leaky ribosome scanning or internal ribosome entry (Kozak, 1999).

If *MOCS1B* is not translated as an independent entity by atypical translational initiation mechanisms (as discussed above), another possibility is that *MOCS1A* and *MOCS1B* are translated as a single bifunctional protein. Such a fusion protein could result from selenocysteine suppression (Low & Berry, 1996) of the terminating UGA nonsense codon for *MOCS1A*, which would permit continued translation into the in-frame *MOCS1B* ORF. As the human *MOCS1* mRNA appears to lack a candidate SECIS (selenocysteine insertion site) element (Grundner-Culemann et al., 1999), and the invertebrate orthologs cannot use this mechanism, this latter hypothesis seems unlikely, but cannot be excluded without empirical testing. In a second model, a *MOCS1A*–*MOCS1B* bifunctional protein could be produced from an alternatively spliced transcript that circumvents the *MOCS1A* termination nonsense codon ("no-nonsense" transcript). In support of the alternative splicing model above, initial database searches utilizing a partial human *MOCS1A*–*MOCS1B* cDNA sequence centered on the intercistronic junction identified a single mouse EST (GenBank accession number (acc. no.) AI049337) that had a 15-nt deletion with respect to the human sequence, removing the nonsense codon. Phylogenetic and molecular studies of the *MOCS1* locus were therefore performed on eight vertebrate and invertebrate organisms revealing that all produce no-nonsense transcripts in addition to transcripts with a bicistronic structure. This finding, coupled with the conventional scanning model of translation initiation, indicates that the no-nonsense splice forms give rise to a fused *MOCS1A*–*MOCS1B* enzyme, whereas the bicistronic mRNA would produce a monofunctional *MOCS1A* protein. We further propose a model arising from analyses of the phylogenetic data suggesting that only *MOCS1A*, but not *MOCS1B*, is translated from the bicistronic mRNA.

RESULTS AND DISCUSSION

Alternative splicing of *MOCS1* orthologs

Degenerate consensus (human/mouse/fly) primers were designed in well-conserved regions of *MOCS1A* and *MOCS1B* based on human (Reiss et al., 1998b) and predicted mouse and fly sequences (see Materials and Methods), and used to amplify *MOCS1* orthologous transcripts from a variety of vertebrate sources. The resulting products of ~1 kb were subjected to sequence analysis, which verified that the recovered prod-

ucts were vertebrate *MOCS1A* and *B* orthologs (Fig. 1). Interestingly, the direct sequences of PCR products also showed a multiplicity of nucleotide peaks limited to a small region flanking the *MOCS1A* terminating nonsense codon, implying that each PCR sample pool had heterogeneity in the vicinity of the *MOCS1A* and *MOCS1B* junction (data not shown). Cloning and sequencing of products from each species revealed two forms of *MOCS1* transcripts: those with a *MOCS1A* terminating nonsense codon ("bicistronic") and those without ("no-nonsense"). To assess the relative abundance and tissue or species-specific distribution of each splice form, primers were designed to produce RT-PCR products that would enhance electrophoretic differences. As expected, products amplified from all vertebrates show at least two sizes of cDNA (Fig. 2A, and data not shown).

Human heart cDNA produces three discernible products when amplified by primers in exons 8 and 10 (Fig. 2A,B, lane 1). Cloning and sequence analysis of the largest cDNA fragment and comparison to the corresponding genomic structure (Reiss et al., 1998a) reveals that it contains exons 8–10 and represents the nonsense-containing bicistronic splice form (designated Type I splice form) previously reported. The smallest fragment has a 63-nt deletion corresponding to an omission of exon 9, which circumvents the nonsense codon and fuses the *MOCS1A* and *MOCS1B* ORFs in-frame (designated Type III splice form; Fig. 2). The size of the RT-PCR product designated as a Type II splice form (Fig. 2A, lane 1) is consistent with the 15-nt deletion observed in the mouse EST clone, and was better resolved from the Type I mRNA splice form using primers in exons 9 and 10 (Fig. 2A, lane 2). Sequence analysis verified that this is the human equivalent of the 15-nt deleted form (Type II) which also fuses the *MOCS1A* and *MOCS1B* ORFs (Fig. 2B). Type I and II splice forms were also found by RT-PCR and sequence analyses in all tissues examined in the mouse (Figs. 2A, lanes 3–6, and 2B), although the relative level of Type II is low in testis whereas heart shows comparable levels of both isoforms. Similarly, Type I and II splice forms were identified in cow, rabbit, and marsupial brain (Figs. 2A, lane 7, and 2B; and data not shown), as well as from chicken d13 embryo retinal tissue (Figs. 2A, lane 8, and 2B). While Type III splicing has not been examined in the latter five species, we can conclude that all tested vertebrate species produce both bicistronic and monocistronic *MOCS1* transcripts, in which the latter encodes only a fused *MOCS1A*–*MOCS1B* protein product.

Splicing of invertebrate *MOCS1* orthologous loci

Analysis of *Drosophila* genomic BAC sequence revealed a four exon structure with a similar Type I bi-

cistronic arrangement of the respective *MOCS1A* and *MOCS1B* orthologs (Figs. 1, 2B). Accordingly, a *Drosophila* EST was identified (GenBank acc. no. AI293256), sequenced, and identified as a Type I splice form. Notable differences with the vertebrate orthologs are that the fly has two consecutive nonsense codons terminating the *MOCS1A* ORF (Fig. 1), and that the downstream *MOCS1B* is not in frame with the *MOCS1A* ORF upstream. However, the annotation accompanying the genomic sequence (AC006562.8) predicted an exon 4 splice acceptor downstream of the nonsense codons (Type IV; Fig. 2B). Analysis of *MOCS1* cDNA from *Drosophila* egg mRNA confirmed the presence of approximately equal levels of both Types I and IV splice forms (Figs. 2A, lane 9, and 2B). Through the use of the predicted downstream splice acceptor, the Type IV splice form lacks 62 nt that remove the nonsense codons and fuse the respective *MOCS1A* and *MOCS1B* ORFs in-frame (Fig. 2B,C).

Caenorhabditis elegans represents an evolutionarily more distant species to further test the application of the no-nonsense splicing model. The annotated genomic sequence (GenBank acc. no. Z46267.1) predicts an orthologous 10-exon gene that would produce a fused *MOCS1A*–*MOCS1B* protein. However, based on genomic sequence inspection, we predicted that nematodes may also include a novel exon (denoted here as exon 6) not listed in the annotation that would produce a Type I bicistronic transcript when this exon is utilized. This prediction was empirically tested by exon 4 to 7 amplification, which shows two sizes consistent with two splice forms (Figs. 2A, lane 10, and 2B). The presence of the Type I form predicted by our model was verified by sequence analysis (Figs. 1, 2C). Therefore, the *C. elegans* locus is comprised of 11 exons with exon 6 encoding the C-terminus of *MOCS1A*, terminated by a UAA nonsense codon. Surprisingly, the transcript form predicted in the GenBank annotation was not detected. The larger RT-PCR product observed represents another splice form utilizing exon 6 that uses

an alternative splice acceptor 7 nt upstream of the Type I splice acceptor. The mRNA bearing the 7-nt insertion shifts the reading frame through exon 6, such that the *MOCS1A* nonsense termination codon is no longer in-frame. Not only does this frameshift avoid the *MOCS1A* nonsense codon, but it simultaneously brings the *MOCS1B* ORF in register, allowing continuous translation of both ORFs (designated Type V splice form; Fig. 2B,C). It is well-documented that *C. elegans* produces a significant proportion of polycistronic transcripts that are subsequently reduced by *trans*-splicing so that translation occurs monocistronically (Blumenthal, 1995). This esoteric mechanism, however, is superfluous to the *MOCS1* analysis here, because the observed presence of bicistronic and no-nonsense splice forms clearly shows that both monofunctional and bifunctional protein products are likely to be produced. Furthermore, sequence data suggests that *C. elegans* *MOCS1B* does not have a candidate initiating methionine required for independent translation (see below), even if *trans*-spliced.

Therefore, in addition to Type I bicistronic mRNAs, four forms of no-nonsense splicing are documented here: Type II (alternative splice donor), Type III (exon skipping), Type IV (alternative splice acceptor), and Type V (alternative reading frame resulting from an alternative splice acceptor), all of which fuse the upstream and downstream ORFs in register to encode a single bifunctional protein. Collectively, these data demonstrate a striking evolutionary conservation over at least 700 million years of the pattern of *MOCS1* mRNA splicing to produce bicistronic and no-nonsense isoforms, which has important functional implications for the encoded proteins. Notably, the *Aspergillus cnxABC* locus has been reported to produce a transcript with a fused ORF orthologous to *MOCS1A*–*MOCS1B* (Unkles et al., 1997). Genomic inspection, however, indicates that a 4-nt shift in splice donor position could also produce a bicistronic transcript with a *MOCS1A* nonsense codon that maximizes the C-terminal homology to other

FIGURE 1. Deduced amino acid alignment of Type I *MOCS1A* and *MOCS1B* orthologous bicistronic cDNAs. Consensus residues are shaded in medium gray, similar amino acids are lightly shaded, and all methionine (M) residues that may potentially act as translation initiation sites in *MOCS1B* are reverse shaded. Gaps introduced into the alignment are shown as dashes, and asterisks at position 392 (and 393 for fruit fly) denote the nonsense codon that terminates the *MOCS1A* reading frame in Type I transcripts. Species abbreviations: *Hsap*: *Homo sapiens*, human; *Mmus*: *Mus musculus*, mouse; *Btau*: *Bos taurus*, cow; *Ocun*: *Oryctolagus cuniculus*, rabbit; *Mdom*: *Monodelphis domestica*, South American opossum; *Ggal*: *Gallus gallus*, chicken; *Dmel*: *Drosophila melanogaster*, fruit fly; and *Cele*: *Caenorhabditis elegans*, nematode. *Cele* residues N-terminal of Asp328 and C-terminal to Leu414 in the aligned sequence were deduced from GenBank genomic sequence (protein ID CAA86430.1). The *MOCS1B* reading frame depicted in the invertebrate Type I configuration has been shifted (–1) at the splice junction preceding the *MOCS1B* exons (Asp406 in fruit fly and Val395 in nematode) to be in register with the *MOCS1A* reading frame for display purposes. Only partial amino acid sequences are shown for most vertebrates, as these were recovered by cDNA amplification from internal priming sites; human sequence was derived from GenBank acc. no. AJ224328. *MOCS1* cDNA sequences determined in this study have been deposited in GenBank for *Hsap* (Type I: acc. no. AF24015, Type II: acc. no. AF24022, Type III: acc. no. AF24023); *Mmus* (Type I: acc. no. AF24016, Type II: acc. no. AF24024); *Btau* (Type I: acc. no. AF24017); *Ocun* (Type I: acc. no. AF24018); *Mdom* (Type I: acc. no. AF24019, Type II: acc. no. AF24025); *Ggal* (Type I: acc. no. AF24020, Type II: acc. no. AF24026); *Dmel* (Type I: acc. no. AF24021, Type IV: acc. no. AF24027); and *Cele* (Type I: acc. no. AF228711, Type V: acc. no. AF228712).

species (see above), raising the possibility that this fungus produces two splice forms as well. While *Ara-bidopsis* *MOCS1A* and *MOCS1B* orthologs are sepa-rated and present on different chromosomes (Hoff et al., 1995) showing that a sequential organization is not universally required, the present work showing a phy-logenetically widespread representation of this intricate organization and splice scheme suggests that produc-tion of both forms is physiologically and evolutionarily advantageous.

Models from phylogenetic comparisons

Our comparative data in vertebrates and invertebrates reveal the absence of a phylogenetically conserved initiation methionine codon for *MOCS1B* until well into the conserved coding region (e.g., Met541; see Fig. 1). Indeed, the *Drosophila* and *C. elegans* *MOCS1B* se-quences only contain Met residues C-terminal to well-conserved blocks of *MOCS1B* residues (Fig. 1). Even if *C. elegans* *MOCS1B* is *trans*-spliced, the SL2 leader

FIGURE 2. (Legend on facing page.)

may help to select the appropriate initiating methionine but does not encode, and therefore cannot supply, an initiating methionine (Blumenthal, 1995); therefore, the initiating methionine must be intrinsically provided by the *MOCS1B* ORF. Because the invertebrate *MOCS1B* ORFs lack a candidate initiating methionine that would produce a protein including the presumably essential residues, it is highly unlikely that functional *MOCS1B* protein can be independently translated in these species.

Though modestly transcribed, Northern analyses in human (Reiss et al., 1998b), as well as in mouse, opossum, and chicken (Fig. 3) show a single transcript size consistent with mRNAs containing both ORFs, but not smaller transcripts possibly encoding either *MOCS1* subunit alone. It is therefore likely that in vertebrates, as in invertebrates, that *MOCS1B* is not independently translated because (1) there is no evidence for a small, *MOCS1B*-specific transcript in four diverse vertebrates, (2) there is no conserved *MOCS1B* translation initiation codon (see above), and (3) the conventional scanning ribosome translational model (reviewed in Kozak, 1999) only predicts efficient translation of *MOCS1A* but not *MOCS1B* from the bicistronic transcripts. Consequently, we therefore hypothesize that the only protein translated from the bicistronic Type I splice form is monofunctional *MOCS1A* protein. Evidence that *MOCS1A* protein is indeed translated from the bicistronic (Type I) transcript can be seen by noting that the codons specific to this splice form are particularly well conserved (e.g., the Gly-Gly C-terminal dipeptide; Fig. 2C and see below), attesting to their protein-coding function. We envision that the *MOCS1B* ORF of the bicistronic transcript constitutes an extended 3' UTR, similar to *NESP55-GNAS1* transcripts that are thought to produce only *NESP55* protein but not *G_sα* due to the lack of an initiating methionine for the latter ORF (Hayward et al., 1998; Ischia et al., 1997; Peters et al., 1999). In

FIGURE 3. Northern analyses of *MOCS1* transcripts in multiple vertebrates. Transcripts of ~3.3 kb are visible in total RNA prepared from mouse (*Mmus*) liver, South American opossum (*Mdom*) brain, and chicken (*Ggal*) eye. The *MOCS1* probe encompasses a significant portion of both *MOCS1A* and *MOCS1B* (see Materials and Methods and Fig. 1, residues 249–636). Approximate migration positions of marker transcripts are shown on the left.

conclusion, because phylogenetic (this paper) and human mutation (Reiss et al., 1998b) data imply that *MOCS1B* protein must be produced from this locus, we propose that the *MOCS1B* ORF is only translated from the no-nonsense transcripts (Types II–V) as part of a fused *MOCS1A*–*MOCS1B* multifunctional protein (Fig. 2B).

Recent studies indicate that gene fusion to form multidomain proteins is relatively common in evolution (Doolittle, 1999; Enright et al., 1999; Marcotte et al., 1999a, 1999b), including in other eukaryotic genes involved in Mo metabolism (Stallmeyer et al., 1999b). Gene fusion begins with a recombination event juxtaposing two ORFs such that transcription of the upstream ORF continues uninterrupted through the downstream ORF. Should

FIGURE 2. Alternative *MOCS1* splice forms in diverse species. **A:** Electrophoretic separation of *MOCS1* RT-PCR products. Roman numerals designating the splice type (see **B**) of the RT-PCR products are shown to the left of each sample. Species abbreviations are as in Figure 1. The *Mmus* panel shows the patterns from a variety of tissues: B: brain, H: heart, L: liver, and T: testis. **B:** Schematic representation of mRNA splice forms and their respective putative protein products. Genomic organization of human (exons 8–10), fruit fly (exons 3 and 4), and nematode (exons 5–7) are shown at left. Green shading indicates *MOCS1A* sequences and blue designates *MOCS1B*. The red vertical line represents a nonsense codon that terminates the *MOCS1A* ORF. Roman numerals (I–III) in human designate splice donor choices into exon 10; numerals I and IV in fly indicate splice acceptors into exon 4 from exon 3; and numerals I and V in nematode correspond to alternative splicing into exon 6, with Type V including seven additional nucleotides (open box) to shift the reading frame throughout the exon. mRNA products of different splice types are shown in the center, with black vertical lines indicating splice boundaries. Adjacent roman numerals classify splice type: inclusive exon–exon splicing (Type I), alternative splice donor (Type II), exon skipping (Type III), alternative splice acceptor (Type IV), and frameshifted (Type V; alternative reading frame indicated by hatched exon 6). Putative translation products from mRNA splice types are shown at right. According to our model (see text), Type I transcripts produce only *MOCS1A* protein, whereas Types II–V produce fused bifunctional proteins. **C:** Partial *MOCS1A* and *MOCS1B* amino acid sequence deduced from Type I transcripts through the critical splice choice region. Residue shading as in Figure 1, except as noted. Amino acids lacking in human exon 9-skipped (Type III) transcripts are shown by the purple bracket above, and the five codons missing in vertebrate Type II transcripts appear on a yellow background. The orange box indicates residues lacking in the *Dmel* Type IV no-nonsense splice form. The sequence encoded by the *Cele* Type I is shown. The *Dmel* and *Cele* Type I *MOCS1B* amino acid sequences result from a (–1) frameshift to be in register with the *MOCS1A* reading frame for display purposes (See Fig. 1 legend). The *Cele* Type V amino acid sequence is shown separately below, with the point of divergence at Val374 resulting from the frameshift due to alternative splicing into exon 6.

the respective ORFs be in register and not disrupted by an intervening stop codon in the hybrid mRNA, translation initiated for the upstream ORF would be extended to include the downstream ORF, seamlessly fusing the encoded proteins. While other Mo loci (e.g., the *Drosophila cinnamon* (*cin*) gene and its eukaryotic orthologs; Stallmeyer et al., 1999b) have undergone complete gene fusion during evolution, the *MOCS1* locus appears to have achieved a remarkably enduring balance between the no-nonsense splice form (generating a fused protein equivalent to a gene fusion product) and a bicistronic splice form (that we propose independently produces MOCS1A). Our demonstration of coexpression of both no-nonsense and bicistronic splice types over 700 million years indicates that this is not simply a transient arrangement awaiting resolution. The *MOCS1* locus therefore exists in an evolutionarily stable but functionally dynamic state.

Why has this intricate organization and splicing strategy evolved and been so stably maintained? There must be selective advantages uniquely conferred by the protein products derived from each of the bicistronic and the no-nonsense splice types. Clearly, the no-nonsense splice forms encode a single multifunctional protein incorporating both MOCS1A and MOCS1B activities. A composite protein such as this may offer many benefits including enhanced reaction kinetics and coordinate regulation. What benefit could be imparted by the bicistronic splice form? Phylogenetic evidence presented and discussed above suggests that this form most likely produces MOCS1A, but not MOCS1B. If so, this would result in the synthesis of more MOCS1A relative to MOCS1B. Unlike the fused form of MOCS1A, the autonomous form of MOCS1A originating from the bicistronic mRNA is not necessarily dedicated for use with MOCS1B, and is free to diffuse for possible use in other cellular processes. Consistent with this model, the *Arabidopsis* *MOCS1A* ortholog is more widely and highly expressed than the unlinked *MOCS1B* ortholog (Hoff et al., 1995), indicating that greater relative levels of MOCS1A are required in this species. In addition, a *MOCS1A* paralogous locus (F49H6.5) has recently been identified in *C. elegans* unaccompanied by *MOCS1B*-derived sequences. Inspection of this genomic locus shows a very high degree of homology (92% amino acid identity) between its putatively encoded product and MOCS1A, making functional redundancy a plausible scenario. Although transcription from this locus has not yet been verified, its presence supports the model of relative overexpression of MOCS1A with respect to MOCS1B.

A second feature unique to the separate MOCS1A isoform produced by the bicistronic splice type is the invariant Gly-Gly dipeptide at its C-terminus (Figs. 1, 2C). This distinctive feature may be significant, as a Gly-Gly dipeptide confers potent biological activity to the C-termini of other polypeptides, such as ubiquitin

and ubiquitin-like proteins (Hochstrasser, 1996; Haas & Siepmann, 1997; Gray et al., 1999). Any functional attributes specific to MOCS1A will only be revealed by future in vivo and in vitro analyses. Nevertheless, it is evident that each *MOCS1* splice form may encode proteins with distinct properties that would justify sustaining this gene structure and bimodal splice pattern over 700 million years. Furthermore, the ratio of transcripts producing fused or separate polypeptides could differ during development or with environmental/dietary stimuli (Unkles et al., 1997), consistent with our initial mouse studies suggesting tissue specificity.

Concluding comments

The highly conserved retention of both types of *MOCS1* splice forms suggests that the loss of either type may lead to a mutant phenotype. In conventional mutation analyses, nucleic acid changes that occur in the protein-coding sequence that either do not alter the amino acid sequence, or result in a very conservative substitution may be dismissed as too subtle to create a mutant protein. Such mutations, however, may create or abolish cryptic splice sites, thereby disrupting the intricate but apparently essential splicing scheme of the *MOCS1* locus. This information will assist mutation identification and genetic counseling in human molybdenum cofactor deficiency disorders. Unrelated splice junction mutations have already been described in such patients (Reiss et al., 1999a). Of course, no-nonsense splice mutations may also be identified in other species with possible *MOCS1* mutations as well, or in other genetic loci found in the future to utilize no-nonsense splicing.

Similar to the vertebrate and invertebrate *MOCS1* loci characterized here, the human *MOCS2* locus is also composed of two ORFs that encode the remaining two enzymatic activities in the synthesis of molybdopterin (Sloan et al., 1999; Stallmeyer et al., 1999a). As with the human *MOCS1* locus, mutations in *MOCS2* have been identified in patients defined as a distinct complementation group of molybdenum coenzyme deficiency (Reiss et al., 1999b). However, because the *MOCS2A* and *MOCS2B* ORFs are not adjacent but are slightly overlapping and are offset by -2 (or $+1$) with respect to each other, it has been proposed that leaky ribosome scanning accounts for translation of the second ORF (Sloan et al., 1999; Stallmeyer et al., 1999a). Given the similarities with *MOCS1*, the *MOCS2* locus may represent another good candidate locus for the use of no-nonsense splicing.

The complex regulation we have shown for the *MOCS1* locus serves as a model for analyses of other candidate eukaryotic polycistronic loci and highlights the need for in vivo protein studies (Gray et al., 1999) to justify a polycistronic status. It is noteworthy that regardless of which mechanism underlies each no-

nonsense splice type, they all fuse the flanking *MOCS1A* and *MOCS1B* ORFs in-frame, clearly illustrating selection and evolutionary conservation of the splice products, rather than the splicing mechanism *per se*. The functional splice choice displayed in the *MOCS1* locus may represent the most ancient documented conserved splicing strategy observed in any orthologous locus.

MATERIALS AND METHODS

Degenerate vertebrate/fruit fly consensus primers (RN1117, 5'-[GATGTGCGSTTCWTAGAGTAYATGCC]-3', and RN1118, 5'-[ACWGCCTTGACATGTCRTACAS]-3', where R = A/G, S = C/G, W = A/T, and Y = C/T) were designed from human (Reiss et al., 1998b), mouse ESTs (data not shown), and *Drosophila melanogaster* BAC (GenBank acc. no. AC006562.8) sequences. These were used to amplify ~1 kb of *MOCS1* to provide sequence information for species-specific primer sets flanking the region in question. Primers RN946 5'-[GAGCTGCTGAGAATCATTGG]-3' and RN947 5'-[CCTGTGCTAAGGCACTTTGA]-3' were used to detect human splice Types I–III, RN1185 5'-[GAACCGGCCCATGATCCTC]-3' and RN947 for human and mouse splice Types I and II, RN1197 5'-[GTTCAACATTTCCCAGATGAAGAAC]-3' and RN1204 5'-[GGCTCTCAGCCCCTGAATGC]-3' for *Monodelphis domestica*, RN1227 5'-[GCATGTTTAACTTGCCAGATG]-3' and RN1228 5'-[GTACCACTGTTGCTCTGGTG]-3' for chicken, RN1188 5'-[GAGAGTGTTAGCGAGG AACAGTTG]-3' and RN1189 5'-[GATTGTGCGCTATAAGTTGGGTAAG]-3' for *Drosophila melanogaster*, and RN1249 5'-[GGCTGAGAATAACAGCTGATGG]-3' and RN1250 5'-[CTATGATAAAGCGAAGTATAGATGG]-3' for *Caenorhabditis elegans*.

Products derived from the vertebrate amplification with RN1117 and RN1118, as well as *C. elegans* amplified with RN1249 and RN1250, were initially directly sequenced (Cleveland Genomics), then cloned into pCR2.1 (Invitrogen). Gel purified *C. elegans* PCR products were also restricted with *MspI*, which cuts the Type I splice form but not the Type V splice form, to verify fragment identity (data not shown). All other amplification products were purified and cloned in pCR2.1 and individual clones representing each splice type were analyzed by automated sequencing (Cleveland Genomics). Murine and *Drosophila* EST clones were purchased (Research Genetics) and characterized, and the inserts were fully sequenced with vector and internal primers. All nucleic acid (Clustal) and protein (PAM 350) alignments were performed with the MacVector (v.6.5) sequence analysis software with minor modifications to the default parameters.

Northern analyses were performed by standard procedures using 20 µg of total RNA from mouse liver, South American opossum brain, and 13-day embryonic chicken eye. Following transfer, individual lanes were hybridized with species-specific ³²P-radiolabeled cDNA probes from PCR products RN1117/RN1118 (see above) in ExpressHyb (CLONTECH). Blots were washed in 0.1× SSC, 0.1% SDS at 50 °C and film exposures made at –80 °C for up to 10 days.

ACKNOWLEDGMENTS

We gratefully acknowledge Peter Harte and Greg Matera for helpful comments, and Bruce Lamb, John McCarrey, James

Bitgood, Phil Morgan, and Peter Harte for provision of tissue, mRNA, or cDNA reagents. This study was supported by funds from the National Institutes of Health (2R01HD31491 to RDN).

Received January 27, 2000; returned for revision February 17, 2000; revised manuscript received April 13, 2000

REFERENCES

- Blumenthal T. 1995. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet* 11:132–136.
- Blumenthal T. 1998. Gene clusters and polycistronic transcription in eukaryotes. *Bioessays* 20:480–487.
- Doolittle RF. 1999. Do you dig my groove? *Nat Genet* 23:6–8.
- Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402:86–90.
- Gray TA, Saitoh S, Nicholls RD. 1999. An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc Natl Acad Sci USA* 96:5616–5621.
- Grundner-Culemann E, Martin GW 3rd, Harney JW, Berry MJ. 1999. Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *RNA* 5:625–635.
- Haas AL, Siepmann TJ. 1997. Pathways of ubiquitin conjugation. *FASEB J* 11:1257–1268.
- Hayward BE, Moran V, Strain L, Bonthron DT. 1998. Bidirectional imprinting of a single gene: *GNAS1* encodes maternally, paternally, and biallelically derived proteins. *Proc Natl Acad Sci USA* 95:15475–15480.
- Hochstrasser M. 1996. Ubiquitin-dependent protein degradation. *Annu Rev Genet* 30:405–439.
- Hoff T, Schnorr KM, Meyer C, Caboche M. 1995. Isolation of two *Arabidopsis* cDNAs involved in early steps of molybdenum cofactor biosynthesis by functional complementation of *Escherichia coli* mutants. *J Biol Chem* 270:6100–6107.
- Ischia R, Lovisetti-Scamihorn P, Hogue-Angeletti R, Woltersdorfer M, Winkler H, Fischer-Colbrie R. 1997. Molecular cloning and characterization of NESP55, a novel chromogranin-like precursor of a peptide with 5-HT1B receptor antagonist activity. *J Biol Chem* 272:11657–11662.
- Kozak M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* 234:187–208.
- Lee SJ. 1991. Expression of growth/differentiation factor 1 in the nervous system: Conservation of a bicistronic structure. *Proc Natl Acad Sci USA* 88:4250–4254.
- Low SC, Berry MJ. 1996. Knowing when not to stop: Selenocysteine incorporation in eukaryotes. *Trends Biochem Sci* 21:203–208.
- Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. 1999a. Detecting protein function and protein–protein interactions from genome sequences. *Science* 285:751–753.
- Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. 1999b. A combined algorithm for genome-wide prediction of protein function. *Nature* 402:83–86.
- Peters J, Wroe SF, Wells CA, Miller HJ, Bodle D, Beechey CV, Williamson CM, Kelsey G. 1999. A cluster of oppositely imprinted transcripts at the *Gnas* locus in the distal imprinting region of mouse chromosome 2. *Proc Natl Acad Sci USA* 96:3830–3835.
- Reiss J, Christensen E, Dorche C. 1999a. Molybdenum cofactor deficiency: First prenatal genetic analysis. *Prenat Diagn* 19:386–388.
- Reiss J, Christensen E, Kurlemann G, Zabot MT, Dorche C. 1998a. Genomic structure and mutational spectrum of the bicistronic *MOCS1* gene defective in molybdenum cofactor deficiency type A. *Hum Genet* 103:639–644.
- Reiss J, Cohen N, Dorche C, Mandel H, Mendel RR, Stallmeyer B, Zabot MT, Dierks T. 1998b. Mutations in a polycistronic nuclear gene associated with molybdenum cofactor deficiency. *Nat Genet* 20:51–53.
- Reiss J, Dorche C, Stallmeyer B, Mendel RR, Cohen N, Zabot MT. 1999b. Human molybdopter synthase gene: Genomic structure and mutations in molybdenum cofactor deficiency type B. *Am J Hum Genet* 64:706–711.
- Sloan J, Kinghorn JR, Unkles SE. 1999. The two subunits of human molybdopter synthase: Evidence for a bicistronic messenger

- RNA with overlapping reading frames. *Nucleic Acids Res* 27: 854–858.
- Stallmeyer B, Drugeon G, Reiss J, Haenni AL, Mendel RR. 1999a. Human molybdopterin synthase gene: Identification of a bicistronic transcript with overlapping reading frames. *Am J Hum Genet* 64:698–705.
- Stallmeyer B, Schwarz G, Schulze J, Nerlich A, Reiss J, Kirsch J, Mendel RR. 1999b. The neurotransmitter receptor-anchoring protein gephyrin reconstitutes molybdenum cofactor biosynthesis in bacteria, plants, and mammalian cells. *Proc Natl Acad Sci USA* 96:1333–1338.
- Unkles SE, Smith J, Kanan GJ, Millar LJ, Heck IS, Boxer DH, Kinghorn JR. 1997. The *Aspergillus nidulans* *cnxABC* locus is a single gene encoding two catalytic domains required for synthesis of precursor Z, an intermediate in molybdenum cofactor biosynthesis. *J Biol Chem* 272:28381–28390.