Molecular structure of the human muscle-specific enolase gene (*ENO3*)

Mina PESHAVARIA and Ian N. M. DAY

University Department of Clinical Biochemistry, Level D, South Laboratory and Pathology Block, Southampton General Hospital, Tremona Road, Southampton, SO9 4XY, U.K.

The single human gene for muscle-specific enolase was isolated and its structure was characterized, from which the mature mRNA transcript and encoded protein were also deduced. The gene contains 12 exons, spans approx. 6 kb and encodes a protein of 433 residues. The gene structure is similar to that found for the rat neuron-specific enolase gene, and the deduced protein aligns precisely with other enolase sequences, including the sequence of the only published crystallized enolase, yeast eno-1. The 5' boundary of the gene includes a 5' non-coding exon and is characterized by an upstream TATA-like box and CpG-rich region. This region contains potential recognition motifs for general transcriptional regulation involving Sp1, activator protein 1 and 2, CCAAT box transcription factor/nuclear factor I and cyclic AMP, and for muscle-specific transcriptional regulation involving a CC(A + T-rich)₆GG box, M-CAT-box CAATCCT and two myocyte-specific enhancer-binding factor 1 boxes.

INTRODUCTION

Enolase (EC 4.2.1.11) catalyses the glycolytic step interconverting 2-phosphoglycerate and phosphoenolpyruvate. Although this step appears to be at 'equilibrium' *in vivo* and is not a classical site of metabolic pathway regulation [1], tissue-specific isoforms of the enzyme are found and these display minor kinetic differences [2,3]. Additionally, the absolute cellular level of enolase activity is 20-fold higher in muscle (per mg of total soluble protein) than in other tissues [1], possibly reflecting the heavy metabolic demands of skeletal muscle. The possible roles of, and differential regulation of, a wide range of isoenzyme systems and their cognate genes have been extensively investigated. Instances of mechanistic value to the organism have been identified at substrate, kinetic, macromolecular, translational and transcriptional levels, and these roles become clear as a complete description of the system is obtained (for a recent overview, see [4]). Three major tissue-specific isoforms of enolase have been identified in mammals and birds, γ or neuron-specific enolase (NSE), expressed primarily in neurons, β or muscle-specific enolase (MSE), expressed in striated muscle, and α or non-neuronal enolase (NNE), expressed in foetal cells and other adult cell types. Detailed structural information concerning rat and human NSE and NNE has emerged from cDNA cloning [5–11], reaffirming the existence of three separate genes, of which only the rat NSE gene has been characterized so far [12]. The human gene loci have been named *ENO1* (NNE), *ENO2* (NSE) and *ENO3* (MSE), and chromosome *in situ* hybridization data are also suggestive of at least three loci [13].

It is established that there is transcriptional switching of the rat NSE gene during neuronal development [12], and earlier translation studies *in vitro* of chicken muscle had suggested a comparable event for the MSE gene during muscle development [14]. Extensive general information has emerged concerning

myogenesis and muscle gene expression, including the identification of master genes such as myoblast determination gene number 1 (Myo D1) which can trigger undifferentiated cells into adopting a muscle phenotype [15], and the recognition that a common pattern of regulation of genes responsible for the phenotype involves transcriptional regulation by specific protein factors binding specific motifs near the transcriptional start site [16–21]. This work set out to isolate the human MSE gene, both to elucidate its gene structure and investigate its possible transcriptional regulation, and also to deduce the mRNA and protein sequences. Recently, complete rat and human MSE cDNA sequences have been presented [22,23] which correlate well with the mRNA deduced from the gene [24].

MATERIALS AND METHODS

Reagents

All radiochemicals were from Amersham International (Amersham, Bucks., U.K.). Restriction enzymes and other DNA-modifying enzymes were purchased from either Northumbria Biologicals Ltd., Cramlington, Northumbria, U.K., or Anglian Biotech Ltd., Colchester, Essex, U.K. Chemicals and other reagents were from Sigma Chemical Co., Poole, Dorset, U.K., and oligonucleotides were synthesized by the β -cyanoethyl phosphoramidite method using an Applied Biosystems 381A synthesizer.

Gene isolation, sequencing and analysis

A human genomic library in the vector λ 2001 was screened as described previously [25]. One genomic clone (λ HGME) which hybridized at high stringency (0.1 $\times$ SST, 65 °C; SST is standard saline Tris) specifically to muscle mRNA, and which contained targets for probes used at low stringency (2 $\times$ SST, 50 °C) representing the entire coding region of human NSE cDNA, was

Abbreviations used: NSE, neuron-specific enolase; NNE, non-neuronal enolase; MSE, muscle-specific enolase; SST, standard saline Tris (2 $\times$ SST = 0.3 M-NaCl/0.03 M-Tris/HCl (pH 7.5)/5 mM-EDTA); AP-1, activator protein 1; AP-2, activator protein 2; CTF, CCAAT box transcription factor; NF1, nuclear factor I; cAMP, cyclic AMP; CAR_G, CC(A + T-rich)₆GG box; M-CAT, muscle-CAT heptamer CAATCCT; MEF-1, myocyte-specific enhancer-binding factor 1; Myo D1, myoblast determination gene number 1; MCBF, M-CAT-binding protein factor.

The nucleotide sequence data reported will appear in the EMBL, Genbank and DDBJ Nucleotide Sequence Databases under the accession number X55937.

Fig. 1. Northern blots using NNE and NSE cDNA probes and subfragments derived from genomic clone λ HGME

Ribosomal RNAs (6.33 and 2.36 kb) acted as size markers; 30 μ g of total RNA was loaded per track. Probes used and stringency of washing are listed below. Autoradiography was at -70°C with one intensifying screen overnight, except (c), which was a 3 h exposure. (a) Probe = full-length human NNE cDNA; washed at 42°C in $2 \times \text{SST}$; (b) probe = full-length human NSE cDNA; washed at 42°C in $2 \times \text{SST}$; (c) probe = λ HGME region A (Fig. 2a); washed at 65°C in $0.1 \times \text{SST}$; (d) probe = λ HGME region B_1 (Fig. 2a); washed at 65°C in $0.1 \times \text{SST}$; (e) probe = λ HGME region B_2 (Fig. 2a); washed at 65°C in $0.1 \times \text{SST}$. Note that NNE/NSE cross-hybridization is not observed in (a) and (b), although the protein-coding nucleotide-sequence divergence between human NNE and NSE and that deduced from λ HGME appears to be evolutionarily equidistant, being in the range 22–24% for each pairwise comparison. Longer exposures do display NNE/NSE cross-hybridization, and the greater signal from muscle RNA probably indicates that a greater proportion of total RNA in muscle is enolase mRNA. B, brain; M, muscle.

selected for further analysis. Restriction mapping, Southern and Northern blotting and subcloning into plasmids pUC18 and pUC19 and bacteriophages M13mp18 and M13mp19 were by standard methods [26,27]. All DNA sequencing was performed by the Sanger method [27]. Sonication shotgun [28] and restriction-enzyme-directed clones were created and sequencing templates were either single-stranded M13 preparations or double-stranded plasmid denatured by alkali [29]. All sequences were determined on both strands, with overlaps. Computer analysis was performed using the Staden programs as supplied by Amersham International. More detailed subsequent analyses were undertaken using the U.K. 'Seqnet' facility of software and databases implemented on a microVax 3600 at Daresbury, Warrington, U.K.

RNA isolation, Northern blot analysis and 5' primer extension

Total RNA was extracted from *post-mortem* (24–36 h) human skeletal muscle, cerebral cortex and liver, which were stored at -70°C prior to RNA preparation. Frozen tissues were homogenized in guanidinium thiocyanate and RNA was separated by ultracentrifugation through a CsCl cushion [30]. RNA samples of approx. 20 μ g were electrophoresed in 2.2 M-formaldehyde/1.2% agarose gels [26] and blotted on to Hybond-N (Amersham International).

Fragments derived from human NSE cDNA [9] and human NNE cDNA [8,10] were labelled by random priming for use as probes, as were genomic fragments derived from the clone λ HGME described here. The latter probes were pre-reannealed with excess total human DNA prior to use [31]. Hybridizations were performed by standard methods, using empirical and theoretical data to relate percentage sequence identity to probe/target duplex melting temperature during washing of filters [32].

Fig. 2. Map of λ HGME (a) and relationship of splice sites to apparent exon structure (b)

(a) Restriction and exon map of clone λ HGME. Roman numerals indicate the exons, marked either as solid blocks (protein coding regions) or by cross-hatching (non-coding exon regions). An arrow above exon II indicates the location of the oligonucleotide used for mRNA 5' analysis (text and Fig. 5). The sequence of the region designated B is presented in Fig. 4. The regions designated A, B_1 , and B_2 were used as probes on Northern blots (Fig. 1), B_1 and B_2 being divided by the *HincII* site shown in Fig. 4. (b) The intron/exon junction sequences are shown with intron sequence in lower case and exon sequence in upper case. Exons are numbered in Roman numerals, and their nucleotide lengths are indicated. Protein coding exon sequence includes gaps between codons, and the initiator methionine codon in exon II is marked 'M₁'. All splice acceptor and donor sites display obligatory AG and GT dinucleotides respectively.

A 17-mer oligonucleotide GTTGCCCTGGAGTCCA complementary to the apparent *N*-terminus-encoding exon of λ HGME was end-labelled using T4 polynucleotide kinase and [γ - ^{32}P]ATP [26]. A 25 μ g portion of total muscle RNA or brain RNA (as a control) was annealed with the labelled oligonucleotide exactly as described [30] but in the presence of 10 mM-vanadyl ribonucleoside complexes. A murine-leukaemia-virus-reverse-transcriptase reaction (final volume 100 μ l) was performed in the supplier's buffer in the presence of 1 mM of each of the four deoxynucleotide triphosphates and 1000 units of enzyme for 1 h at 37°C . The labelled extension products were then sized against a sequencing ladder. A variety of modified conditions were tested [e.g. double enzyme concentration (see right-hand tracks in Fig. 5), different temperatures, different incubation period], since more than one extension product was obtained.

RESULTS

One genomic clone, designated λ HGME, contained sequences which hybridized to a muscle RNA estimated to be 1.7 kb, but not to liver or brain RNA at high stringency (Figs. 1c and 1e). At stringencies of washing corresponding to 80% sequence

Fig. 3. The protein sequence of human muscle (β) enolase deduced from a genomic clone λ HGME derived from locus *ENO3* is aligned with that of yeast enolase-1

★ indicates identity of match; ● indicates conservation of amino-acid type (acidic, basic or hydrophobic). The large letters designate α -helices, and the large numbers designate β -strands, as identified in yeast *eno-1* crystal structure and using the same nomenclature [33]. Helices A-H surround β -strands 1-8 to form an α/β barrel which contains the active site residues marked by ▼.

identity or lower [32], both NNE and NSE cDNA probes also hybridized to an RNA of 1.7 kb in muscle total RNA (Figs. 1a and 1b). Additionally, at these stringencies cDNA probes representing the N- or C-terminus of NSE [9] identified *Xba*I fragments of 5 kb and 0.8 kb respectively on Southern blots of λ HGME (results not shown).

The map of λ HGME is shown in Fig. 2(a) and the relationship of splice sites to the apparent exon structure is shown in Fig. 2(b). The deduced mRNA structure has been presented in brief [24]; the deduced protein structure can be aligned amino acid for amino acid with other enolase proteins, including species as remote as *Saccharomyces cerevisiae* enolase-1. This alignment is displayed in Fig. 3 and includes a schematic alignment of the crystal structural elements now established for the latter enzyme [33].

The 1.63 kb *Pst*I region designated 'B' in Fig. 2 displays a sequence (Fig. 4) with an atypical CpG content, the first 540 nucleotides displaying 5.9% CpG dinucleotide and a CpG/GpC ratio of 0.61, in contrast with total genomic DNA which contains an average of 1% CpG and a CpG/GpC ratio of 0.23 [34]. Additionally, an 871 bp *Hinc*II-*Pst*I subfragment (designated 'B₂' in Fig. 2a) representing the 3' half of this region hybridizes at high stringency to the apparent 1.7 kb muscle enolase mRNA on Northern blots (Fig. 1e), although the CpG-rich region (designated 'B₁' in Fig. 2a) showed no such hybridization (Fig. 1d). Complementary 5' primer extension using the oligonucleotide described above produced fragments of 114 and 157 nucleotides

using total muscle RNA, but not total brain RNA (Fig. 5). This constellation of findings is sufficient to deduce the likely main patterns of transcription and splicing at the 5' end of this gene (see the Discussion section). A variety of structural features of this region (Fig. 4) which may be important to this process are discussed below.

DISCUSSION

Identification of λ HGME as the human gene encoding muscle-specific enolase

At the protein level it is well established that skeletal muscle expresses an enolase that is electrophoretically and immunologically distinct from those expressed in liver, brain and other tissues, and is apparently the product of a separate gene [2,35]. Previous analyses of cDNA encoding rat NSE, NNE [5-7] and most recently rat MSE [22] have affirmed the existence of separate genes expressed in a tissue-specific fashion. Analogous studies in humans using NNE and NSE cDNA [8-11] for cross-hybridizations (e.g. Figs. 1a and 1b) have demonstrated a related mRNA of 1.7 kb expressed in muscle. Clone λ HGME displays hybridization properties (Fig. 1) establishing muscle-specific expression of at least part of the sequence of the clone studied. The deduced exon structure encodes a protein which can be aligned amino acid for amino acid with human NSE and NNE, and rat NSE, NNE and MSE, without frameshifts or stop codons (results not shown) and with preservation of established

Fig. 4. Sequence of a *PstI* subfragment designated 'B' from λ HGME (see Fig. 2a) which represents the 5' end of the human muscle enolase gene

CpG dinucleotides are marked by solid circles, and a direct repeat of 10 nucleotides is marked by a solid square with horizontal arrow. The TATA-like box is shown by a double-lined box, and the obligatory splice donor and acceptor nucleotides of intron I are labelled as such and doubly underlined. The *HincII* and *PstI* sites used to obtain the probe used for the Northern blot in Fig. 1(e) are marked. The first three codons, which are found in exon II (Fig. 2b), are underlined and the amino acids encoded are indicated in one-letter format, the initiator methionine bearing the subscript 'i'. The region initially deduced (see the Discussion section) to contain exon I is indicated by open arrowheads in the right margin, and the apparent start site of transcription is marked by a vertical arrowhead. A region displaying a 9/12 match with the likely 3' boundary of rat MSE exon I is underlined and is immediately followed by a potential GT splice donor, underlined by a zig-zag (see the Discussion section). Ten motifs of potential significance to transcriptional regulation (see the Discussion section) are enclosed in boxes and numbered sequentially: 1, an AP-1 consensus; 2, an AP-2 consensus; 3, an Sp 1 consensus; 4, a CArG box consensus; 5, an Sp 1 consensus; 6, a minimal consensus for cyclic AMP regulation; 7, an M-CAT box; 8, a MEF/Myo D1 consensus; 9, a CTF/NF-I consensus; 10, an MEF/Myo D1 consensus. Motifs 8 and 10 form a perfect long-range inverted repeat of 12 nucleotides, as discussed in the text. The whole sequence was determined on both strands, with overlaps.

active-site residues (Fig. 3). The deduced mRNA [24] of 1.5 kb corresponds with the results of Northern blots (allowing for a polyadenylate tail) and is smaller than the mRNAs for human NNE and NSE (Figs. 1a and 1b). Additionally, restriction subfragments from λ HGME used as probes correspond uniquely with fragments of the same size on Southern blots of restriction digests of total human DNA (results not shown), both

establishing the physical veracity of the cloned region and excluding the possibility of two related enolase genes both specifically expressing a 1.7 kb mRNA in skeletal muscle.

Deduced human MSE

Chicken muscle enolase protein has been sequenced by conventional Edman degradation of overlapping peptides [36], and

Fig. 5. MSE mRNA 5' end analysis

The ^{32}P -end-labelled oligonucleotide GTTGCCCTGGAGTCCA complementary to part of exon II (Fig. 2a) was used to prime a reverse transcription reaction using total brain (track B) or total muscle (track M) RNA as the template (see the Materials and methods section for conditions). The denatured products were electrophoresed alongside a sequencing ladder as shown in the resulting autoradiograph. The deduced limit(s) of exon I are shown in Fig. 4 and discussed in the text.

the rat muscle enolase sequence has been deduced from cDNA clones [22]. The former sequence was used in conjunction with secondary structure prediction algorithms to attempt to define the physical structure of enolase [37]. From the enolase sequences now known, it is clear that direct alignment enables useful comparison and prediction [9]. The direct superposition of human MSE α positions upon the known crystal structure α positions for yeast enolase-1 illustrated in Fig. 3 opens the possibility of direct predictive modelling. No major kinetic differences have been identified between NNE, NSE and MSE (e.g. [3]). Our modelling accords with this, in that the protein sequence variations correlate with surface residues, whereas the core and active-site residues within the α/β barrel (Fig. 3) are highly conserved. The variations may yield an insight into evolutionary patterns and can now be investigated as the possible explanation of known differences in cytoskeletal and other macromolecular interactions displayed by these three isoenzymes [38–41].

Boundaries of the MSE gene

Relative to the deduced protein sequence, the splice boundaries of the human MSE gene correspond precisely with those of the rat NSE gene [12], reinforcing the assumption that these mammalian enolase genes are related by an event involving duplication of chromosomal material, such as the tetraploidization believed to have taken place over 300×10^6 years ago, rather than a retropositional event involving mRNA [42].

The apparent splice sites are all bounded by typical splice consensus [43], and an apparent polyadenylation signal noted previously [24] aligns with that found in cDNA representing the equivalent mature transcript in rat muscle [22].

A region with a high CpG content and CpG/GpC ratio is typical of the 5' end of many mammalian genes [44,45] and is typically found upstream of the TATA box [46] which directs transcriptional initiation, although it may extend as far as the first intron. Such a region is present in λ HGME and is particularly distinctive beyond 1100 bp upstream of the initiator methionine codon (Fig. 4). This region (B_1 in Fig. 2a) showed no hybridization to muscle RNA, although the adjacent *HincII*–*PstI* fragment (B_2 in Fig. 2a; see also Fig. 4) did hybridize at high

stringency (Figs. 1d and 1e). It was therefore likely that the 5' non-coding region of MSE mRNA, calculated to be 63 nucleotides (shorter species) and 106 nucleotides (longer species) from 5' primer extension experiments, was reported in the latter range either as a separate 5' non-coding exon [47] or as a 5' continuation of the first coding exon. Attempts to combine an appropriately placed TATA-like sequence with these lengths appear to rule out the latter possibility. For there to be a 5' non-coding exon, the only likely splice site for the first coding exon is the sequence TGCCTGCAG, placed 2 nucleotides upstream of the initiator codon, and the most likely disposition of a TATA box and splice donor site (i.e. TATA...20–30 nucleotides... transcript start...61 or 104 nucleotide exon I... GT of splice site) is within the lines marked by open arrowheads in Fig. 4. Recent analyses of rat [22] and human [23] MSE cDNAs show nucleotide alignments within this region, although the latter displays an inversion (probably a cDNA cloning artefact [9]) relative to the genomic sequence. Therefore we believe our deduced TATA box to be correct. It is noted that the rat and human cDNA alignments place the exon I 3' boundary adjacent to different GT potential splice donor sites within the appropriate range, and it is possible that their usage is relative rather than absolute. The longer primer extension product places a maximal length exon I with a transcriptional start site 27 nucleotides downstream of the ATAAAT motif, assuming the 3' boundary to correspond with the published human cDNA sequence [23]. We were unable to find conditions yielding only one 5' primer extension product, or one SI nuclease protection fragment (results not shown). However, we note that the shorter apparent exon I, counted from the same transcriptional start site, would have its boundary at exactly the point at which alignment with the 3' limit of the rat MSE cDNA 5' non-coding region is found, and would be followed immediately by a favourable intron donor site (Fig. 4). Since the intron acceptor site is only two nucleotides away from the initiator methionine codon, two different Kozak consensus [48], namely TAGTAAAGGCCATGG and CATCCCAAGCC-ATGG, could be formed. It is conceivable therefore that exon I in the MSE gene might give rise to differential regulation at both the transcriptional and the translational level, either through this mechanism or through some other alternative exon I. Strategies involving combinations of oligonucleotides and polymerase chain reaction have been successful in characterizing the micro-heterogeneity of 5' exons in more detail [49], and might be useful for an extended analysis of these possibilities.

Content of the 5' flanking region of the MSE gene

Detailed information has emerged concerning the structure and transcriptional regulation of eukaryotic genes, including the definition of regulatory factors and recognition sequences most commonly placed within 1–2 kb upstream of the transcriptional start site [50]. Therefore it is now possible to make reasonable mechanistic guesses based upon searches using databases of known recognition motifs (e.g. [16]) as a guide to further empirical analysis. It is known from protein studies [2], *in vitro* translation [14] and Northern blotting [51] that mammals and birds exhibit a switch from NNE to MSE during foetal muscle development, that this switch is transcriptional in nature, and that it is influenced by changes of muscle innervation [52] and by the myogenic determinant Myo D1 [51]. In the yeast *Saccharomyces cerevisiae* it has already been shown that the 5' end of one enolase locus contains a transcriptional regulatory element susceptible to glucose regulation [53]. The 5' end of the other enolase locus is susceptible to regulation by heat shock under the control of a distal locus *HSR-1* [54].

The core consensus target GGGCGG for the general transcriptional activator protein Spl [50] is found in the human

MSE gene at positions -495 and -394 relative to the TATA box. Perfect consensus sequences are also present for the general activator proteins AP-1 [50] and AP-2 [50] at positions -706 and -770 respectively. A minimal target for cyclic AMP regulation (CGTCA) [21] is observed at position -366. At 79 nucleotides preceding the TATA box, a site with a 10/11 match with the nuclear factor I (NFI)/CCAAT box transcription factor (CTF) target is found. These proteins belong to a family of proteins involved in CCAAT box recognition [50], and often act upon a target of 40-100 nucleotides preceding the TATA site. No CCAAT box is found in this gene, but this NFI/CTF motif may subservise a similar role.

Several well-characterized, powerful, position- and orientation-independent sites of enhancement of muscle-specific expression are known. Four are present within a range of 800 bp of the transcriptional start site of the human MSE gene. There is an M-CAT element (CATTCCT) [19] at position -283 relative to the TATA box, a CArG motif [CC(A+T-rich)₆GG box] [17,18] at position -464, a myocyte enhancer factor 1 (MEF-1) target [55] at position -126 and an MEF-1 target in intron I. The M-CAT element is a perfect match for that bound by a specific protein, MCBP (M-CAT-binding protein) and responsible for muscle-specific activation of the troponin T gene [56]. The sequence CCCTAAGAGG represents a possible match with the CC(A/T)₆GG consensus known to be associated with muscle actin promoters [17,18]. The two MEF-1 consensus sequences in the human MSE gene are in opposite orientation, and also constitute a long-range inverted repeat of 12 nucleotides. Any cruciform structure involving these elements would contain the TATA box and exon I in its loop. Additionally, the most distinctive direct repeat involves the sequence CTGTCCAGC, representing the first 10 nucleotides of the transcript, which is repeated at position -605 relative to the TATA box. A number of complex structures might thus be possible in association with transcription of this gene. The MEF-1 enhancer sequence defined in the mouse muscle creatine kinase gene [55,57] binds a protein factor MEF-1, which is closely related to Myo D1, a nuclear phosphoprotein that plays a major role in early events of myogenic determination [15,58]. Molecular cloning has demonstrated that Myo D1 is a member of a family of related genes and proteins. Further analysis of the interactions between regulator and regulated genes in muscle should yield an insight into the metabolic integration of the muscle cell during development, and it appears likely that the MSE gene will be highly relevant to this integration.

This work was supported by the Sir Halley Stewart Trust. Related work in the laboratory of I.N.M.D. is also supported by the Medical Research Council of Great Britain and the British Epilepsy Research Foundation. We are grateful to Dr. Mark Pickett for oligonucleotide synthesis, Mrs. P. Allsop and Miss L. Hinks for assistance in the early stages of this work, Professor R. J. Thompson for Departmental space and facilities, and Mrs. Diane Brown for typing the manuscript.

REFERENCES

- Newsholme, E. A. & Start, C. (1973) Regulation in Metabolism, pp. 88-137, John Wiley & Sons, London, New York, Sydney and Toronto
- Pearce, J. M., Edwards, Y. H. & Harris, H. (1975) *Ann. J. Hum. Genet.* **39**, 263-280
- Marangos, P. J., Parma, A. M. & Goodwin, F. K. (1978) *J. Neurochem.* **31**, 727-732
- Doohan, S. (ed.) (1990) *Biochem. Soc. Trans.* **18**, 167-196
- Sakimura, K., Kushiya, E., Obinata, M., Odani, S. & Takahashi, Y. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7453-7457
- Sakimura, K., Kushiya, E., Obinata, M. & Takahashi, Y. (1985) *Nucleic Acids Res.* **13**, 4365-4378
- Forss-Petter, S., Danielson, P. & Sutcliffe, J. G. (1986) *J. Neurosci. Res.* **16**, 141-156
- Giallongo, A., Feo, S., Moore, R., Croce, C. M. & Showe, L. C. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6741-6745
- McAleese, S. M., Dunbar, B., Fothergill, J. E., Hinks, L. J. & Day, I. N. M. (1988) *Eur. J. Biochem.* **178**, 413-417
- Peshavaria, M., Quinn, G. B., Reeves, I., Hinks, L. J. & Day, I. N. M. (1990) *Biochem. Soc. Trans.* **18**, 254-255
- Oliva, D., Barba, G., Barbieri, G., Giallongo, A. & Feo, S. (1989) *Gene* **79**, 355-360
- Sakimura, K., Kushiya, E., Takahashi, Y. & Suzuki, Y. (1987) *Gene* **60**, 103-113
- Craig, S. P., Day, I. N. M., Thompson, R. J. & Craig, I. W. (1991) *Cytogenet. Cell. Genet.* in the press
- Tanaka, M., Sugisaki, K. & Nakashima, K. (1985) *Biochem. Biophys. Res. Commun.* **133**, 868-872
- Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P.-F., Weintraub, H. & Lassar, A. B. (1988) *Science* **242**, 405-411
- Wingender, E. (1988) *Nucleic Acids Res.* **16**, 1879-1901
- Gustafson, T. A. & Kedes, L. (1989) *Mol. Cell. Biol.* **9**, 3269-3283
- Ng, S.-Y., Gunning, P., Liu, S.-H., Leavitt, J. & Kedes, L. (1989) *Nucleic Acids Res.* **17**, 601-605
- Mar, J. H. & Ordahl, C. P. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6404-6408
- Devlin, B. H., Wefald, F. C., Kraus, W. E., Bernard, T. S. & Williams, R. S. (1989) *J. Biol. Chem.* **264**, 13896-13901
- Fink, J. S., Verhave, M., Kasper, S., Tsukuda, T., Mandel, G. & Goodman, R. H. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6662-6666
- Ohshima, Y., Mitsui, H., Takayama, Y., Kushiya, E., Sakimura, K. & Takahashi, Y. (1989) *FEBS Lett.* **242**, 425-430
- Cali, L., Feo, S., Oliva, D. & Giallongo, A. (1990) *Nucleic Acids Res.* **18**, 1893
- Peshavaria, M., Hinks, L. J. & Day, I. N. M. (1989) *Nucleic Acids Res.* **17**, 8862
- Day, I. N. M., Hinks, L. J. & Thompson, R. J. (1990) *Biochem. J.* **268**, 521-524
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- M13 Handbook, Amersham International, Amersham
- Anderson, S. (1981) *Nucleic Acids Res.* **9**, 3015-3027
- Hattori, M. & Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232-238
- Calzone, F. J., Britten, R. J. & Davidson, E. H. (1987) *Methods Enzymol.* **152**, 611-632
- Spoerel, N. A. & Kalatos, F. C. (1987) *Methods Enzymol.* **152**, 598-603
- Lathe, R. (1985) *J. Mol. Biol.* **183**, 1-12
- Stec, B. & Lebiada, L. (1990) *J. Mol. Biol.* **211**, 235-248
- Swartz, M. N., Trautner, T. A. & Kornberg, A. (1962) *J. Biol. Chem.* **237**, 1961-1967
- Marangos, P. J., Zis, A. P., Clark, R. L. & Goodwin, F. K. (1978) *Brain Res.* **150**, 117-133
- Russell, G. A., Dunbar, B. & Fothergill-Gilmore, L. A. (1986) *Biochem. J.* **236**, 115-126
- Sawyer, L., Fothergill-Gilmore, L. A. & Russell, G. A. (1986) *Biochem. J.* **236**, 127-130
- Brady, S. T. & Lasek, R. J. (1981) *Cell* **23**, 515-523
- Wistow, G. & Piatigorsky, J. (1987) *Science* **236**, 1554-1556
- Batke, J., Nazaryan, K. B. & Karapetian, N. H. (1988) *Arch. Biochem. Biophys.* **264**, 510-518
- Pagliaro, L., Kerr, K. & Lansing Taylor, D. (1989) *J. Cell. Sci.* **94**, 333-342
- Tolan, D. R., Niclas, J., Bruce, B. D. & Lebo, R. V. (1987) *Am. J. Hum. Genet.* **41**, 907-924
- Shapiro, M. B. & Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155-7174
- Bird, A. P. (1987) *Trends Genet.* **3**, 342-347
- Gardiner-Garden, M. & Frommer, M. (1987) *J. Mol. Biol.* **196**, 261-282
- Penotti, F. E. (1990) *J. Mol. Biol.* **213**, 37-52
- Hawkins, J. D. (1988) *Nucleic Acids Res.* **16**, 9893-9905
- Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125-8148
- Delort, J., Dumas, J. B., Darmon, M. C. & Mallet, J. (1989) *Nucleic Acids Res.* **17**, 6439-6448
- Thangue, N. B. & Rigby, P. W. J. (1988) in *Transcription and Splicing* (Hames, B. D. & Glover, D. M., eds.), pp. 1-42, IRL Press, Oxford and Washington, DC

51. Lamandé, N., Mazo, A. M., Lucas, M., Montarras, D., Pinset, C., Gros, F., Legault-Demare, L. & Lazar, M. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4445–4449
52. Matsushita, H., Yamada, S., Satoh, T., Kato, K. & Adachi, M. (1986) *Exp. Neurol.* **93**, 84–91
53. Cohen, R., Holland, J. P., Yokoi, T. & Holland, M. J. (1986) *Mol. Cell. Biol.* **6**, 2287–2297
54. Iida, H. & Yahara, I. (1985) *Nature (London)* **315**, 688–690
55. Horlick, R. A. & Benfield, P. A. (1989) *Mol. Cell. Biol.* **9**, 2396–2413
56. Mar, J. H. & Ordahl, C. P. (1990) *Mol. Cell. Biol.* **10**, 4271–4283
57. Buskin, J. N. & Hauschka, S. D. (1989) *Mol. Cell. Biol.* **9**, 2627–2640
58. Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., Hauschka, S. D. & Weintraub, H. (1989) *Cell* **58**, 823–831
59. Wright, W. E., Sassoon, D. A. & Lin, V. K. (1989) *Cell* **56**, 607–617
60. Miner, J. H. & Wold, B. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1089–1093
61. Braun, T., Bober, E., Winter, B., Rosenthal, N. & Arnold, H. H. (1990) *EMBO J.* **9**, 821–831

Received 24 October 1990; accepted 27 November 1990