The EMBO Journal vol.5 no.10 pp.2559-2567, 1986

A 5′ duplication of the α-cardiac actin gene in BALB/c mice is associated with abnormal levels of α-cardiac and α-skeletal actin mRNAs in adult cardiac tissue

Ian Garner, Adrian J.Minty[1], Serge Alonso, Paul J.Barton and Margaret E.Buckingham

Department of Molecular Biology, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris, Cedex 15, France

[1]Present address: Institut de Pathologie Moleculaire, 24 rue de Faubourg Saint Jacques, 75674 Paris, Cedex 14, France

Communicated by M.E.Buckingham

We describe the structure and transcriptional activity of the 5′ portion of the α-cardiac actin gene of BALB/c mice. Southern blotting and DNA sequencing reveal that the promoter and first three exons of the gene are present as perfect repeats in a direct duplication of 9.5 kbp situated immediately upstream of the gene. Both promoters are active in adult cardiac tissue. Transcripts from the partial gene duplication give rise to novel RNAs that are spliced correctly in the actin region and polyadenylated. The level of mature α-cardiac actin mRNA is only 16.5% that found in mice that do not possess the duplication. This is due, at least in part, to interference at the transcriptional level. Transcripts from the α-skeletal actin gene accumulate to abnormally high levels in the hearts of such mutant mice. This result suggests tight regulatory coupling for this actin gene pair.

Key words: cardiac actin/5′ duplication/transcriptional interference/skeletal actin mRNA induction/novel mRNAs

Introduction

Myogenesis leads to the establishment of terminally differentiated muscle fibres. Characteristically, these synthesise a number of muscle-specific proteins including isoforms of the contractile protein multigene families (reviewed in Buckingham, 1985). The actins constitute such a family whose highly conserved members are differentially expressed during development and in adult muscle and non-muscle tissues (Vandekerckhove and Weber, 1981; Minty *et al.*, 1982; Buckingham, 1985). In the mammals six isoforms have thus far been detected: two non-muscle (β- and γ-cytoplasmic), two smooth muscle (α- and γ-smooth) and two striated muscle (α-skeletal and α-cardiac) (Vandekerckhove and Weber, 1979). Each of these is encoded by a single gene (Minty *et al.*, 1983; Ponte *et al.*, 1983) which is not linked to loci encoding either other members of the actin family (Czosnek *et al.*, 1983; Minty *et al.*, 1983; Gunning *et al.*, 1984) or other contractile proteins (Czosnek *et al.*, 1982; Robert *et al.*, 1985). It is probable that 'muscle type' actin genes first appeared during chordate evolution (Vandekerckhove and Weber, 1984; Alonso *et al.*, 1986). The striated muscle actin genes appear to have arisen by gene duplication and dispersion events during amphibian evolution.

Typically, the actins are expressed in pairs with variable relative levels. Thus, during striated muscle development these isoforms are co-expressed at high levels and in the adult skeletal muscle or heart of small mammals the α-skeletal or α-cardiac isoform predominates, respectively (>95%). Co-accumulation of these two actin mRNAs has been demonstrated in fetal and newborn muscle *in vivo* (Minty *et al.*, 1982; Gunning *et al.*, 1983; Mayer *et al.*, 1984) and in certain myogenic cell lines when myotubes form as a result of myoblast cell fusion (Minty *et al.*, 1982; Bains *et al.*, 1984) and heterokaryons formed between mouse muscle cells and human fibroblasts *in vitro* (Blau *et al.*, 1985). This has recently been shown to be the case for the corresponding proteins as well (Vandekerckhove *et al.*, 1986). The expression of these genes is thought to be regulated primarily at the transcriptional level. This is supported, at least in the case of the α-skeletal actin gene, by studies using DNase I digestion to investigate chromatin conformation (Carmon *et al.*, 1982) and cell transfection with promoter-CAT gene fusions (Melloul *et al.*, 1984).

Abnormal gene structures can offer insights into the mechanisms of gene diversification and of differential gene regulation within multigene families. Here, we describe the 5′ region of the α-cardiac actin gene of BALB/c mice. We show that the promoter and coding region up to amino acid 149 have been perfectly duplicated in these mice within a direct repeat of 9.5 kbp immediately upstream of the gene. We demonstrate that this duplication gives rise to transcripts which appear to be spliced normally in the actin coding region but generate abnormal mRNA species due to aberrant splices and processing at their 3′ ends. The presence of the upstream promoter is associated with reduced levels of both primary and mature α-cardiac actin transcripts in adult cardiac tissue. An increased accumulation of α-skeletal actin mRNA is seen in this situation which may be due to a compensatory response to the disrupted α-cardiac gene expression.

Results

The α-cardiac actin gene of BALB/c mice contains a tandem duplication of the 5′ region

We have previously reported that BALB/c mice possess a 5′ duplication of the α-cardiac actin gene which originally manifested itself as a *BclI* restriction fragment length polymorphism (Minty *et al.*, 1983). We have further characterised this phenomenon in BALB/c and other mice using restriction analysis and molecular cloning. Thus far, we have identified two strains of *mus*1 mice that possess this feature (BALB/c and DBA2) and four that do not (C3H, AKR, 129 and C57BL/6). In addition, *Mus spretus* (*mus*3) mice do not possess the duplication. For the purposes of this study we confine our interest to the structure of the mutant locus in BALB/c mice. Southern blot analysis of this region reveals clear differences between BALB/c and control mice (Figure 1). We made use of restriction fragments as probes from pAF81, a cDNA recombinant containing an insert complementary to 93% of the coding region of cardiac actin mRNA, amino acids 28-375 (Minty *et al.*, 1982). When hybridised with the 5′ *PstI-BglII* fragment from this insert (containing information for amino acids 28-85) blots reveal clear differences between mutant and wild-type loci (Figure 1). For example, *BclI* generates a fragment of at least 23 kbp with BALB/c DNA

(B/*Bcl*I) and only 15 kbp with C3H DNA (C/*Bcl*I). Similarly, *Hin*dIII generates bands of 10 kbp and 9.5 kbp in BALB/c (B/*Hin*dIII) and 10 kbp in C3H (C/*Hin*dIII) whereas *Eco*RI generates bands of 5.3 kbp and 9 kbp in BALB/c (B/*Eco*RI) and 5.3 kbp in C3H (C/*Eco*RI). The 3′ *Pvu*II−*Pst*I fragment from pAF81 (complementary to mRNA coding for amino acids 230−375) hybridises to all of the C3H bands but only to the ≥23 kbp *Bcl*I, 10 kbp *Hin*dIII and 5.3 kbp *Eco*RI fragments of BALB/c (data not shown).

The fact that both probes hybridise to a ≥23 kbp *Bcl*I fragment in mutant mice suggests that the duplicated region is close to the α-cardiac actin gene. To define this region more precisely, we isolated recombinant phage containing the duplicated region (λgA8) and the gene itself (λIG10), by screening genomic banks of BALB/c DNA with the 5′ and 3′ probes derived from pAF81. These two phage were restriction mapped and probes derived from them and pAF81 were used to analyse Southern blots of single, double and partial digests of BALB/c DNA (data not shown). This analysis indicated that the duplication is organised in tandem with the gene and enabled us to construct a restriction map of ~17 kbp of this region (Figure 2). We conclude from this that BALB/c mice possess a tandem direct repeat of at least 8 kbp of the 5′ region of the α-cardiac actin gene which does not extend beyond codon 230.

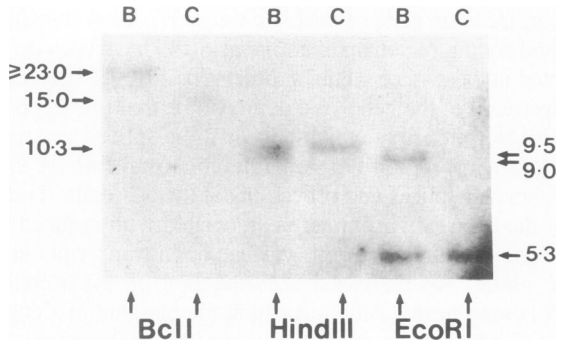


Fig. 1. Southern blot analysis of restricted mouse DNA hybridised to the 5′ *Pst*I−*Bgl*II fragment of pAF81. DNA samples from BALB/c **(B)** and C3H **(C)** mice were digested with *Bcl*I, *Hin*dIII and *Eco*RI, run on a 1% agarose gel and transferred to a nitrocellulose filter. This was hybridised with a nick-translated probe of the 5′ *Pst*I−*Bgl*II fragment of pAF81 (Minty *et al.*, 1982) and washed in 0.1 × SSC/0.1% SDS at 70°C. Sizes of fragments, as estimated from λ markers, are indicated in kbp.

The duplication contains identical repeats of the promoter region and first three exons of the α-cardiac actin gene

The pattern of restriction sites in the repeat indicates that gross rearrangements have not taken place within it. Three homologous regions of the duplication and the gene which hybridise to the 5′ pAF81 probe were sequenced using M13mp8 recombinant phage generated by the 'shotgun' approach: x/x′ (1.1 kbp), y/y′ (0.5 kbp) and z/z′ (1.0 kbp) (Figure 2). Regions x and x′ are identical, as are regions y and y′ (Figure 3A and B, respectively). Regions z and z′ are identical up to nucleotide 671 (Figure 3C) but diverge from each other after this point following a run of seven alternating purine/pyrimidine residues, ACACACA. No significant degree of homology exists beyond this point. We therefore conclude that this position defines the 3′ limit of the duplicated region. To investigate the 5′ extent of the duplication, we analysed Southern blots of digests of genomic BALB/c and C3H DNA (data not shown) using a probe homologous to the region immediately downstream of the 3′ limit of the duplication (nucleotides 670−840, Figure 3C). This sequence is unique in the C3H genome hybridising to a 12-kbp *Eco*RI fragment. However, it is duplicated in the BALB/c genome hybridising to 9-kbp and 15-kbp *Eco*RI fragments. Both of these regions are represented in λgA8 (Figure 2). We conclude from this that the duplication is essentially a perfect direct repeat of at least 9.5 kbp and is located immediately upstream of its homologous sequence.

Comparison of our sequences with the published sequences of pAF81 and both the human and chicken α-cardiac actin genes (Minty *et al.*, 1982; Hamada *et al.*, 1982a; Chang *et al.*, 1985; Eldridge *et al.*, 1985) allowed us to define the α-cardiac actin exons that had been duplicated, to locate them on our genomic map (Figure 2) and to identify a potential promoter region. Thus the first two coding exons containing information up to amino acid 149 are present in region A (Figure 2, exons II and III; Figure 3B and C). As noted above, these duplicated exons and flanking sequences are identical to their homologues in the gene itself and therefore might be expected to function faithfully in splicing and translational events should they be transcribed. Similarly, we can identify a putative promoter sequence within the 5′ 1.1-kbp sequences (x and x′) of the gene and the duplication (Figure 3A). It contains consensus TATA and CAAT boxes characteristic of many eucaryotic promoters and homologous to those proposed for the chicken promoter (Chang *et al.*, 1985; Eldridge *et al.*, 1985). Homology with the chicken gene extends beyond these regions but is confined to within 245 nucleotides upstream

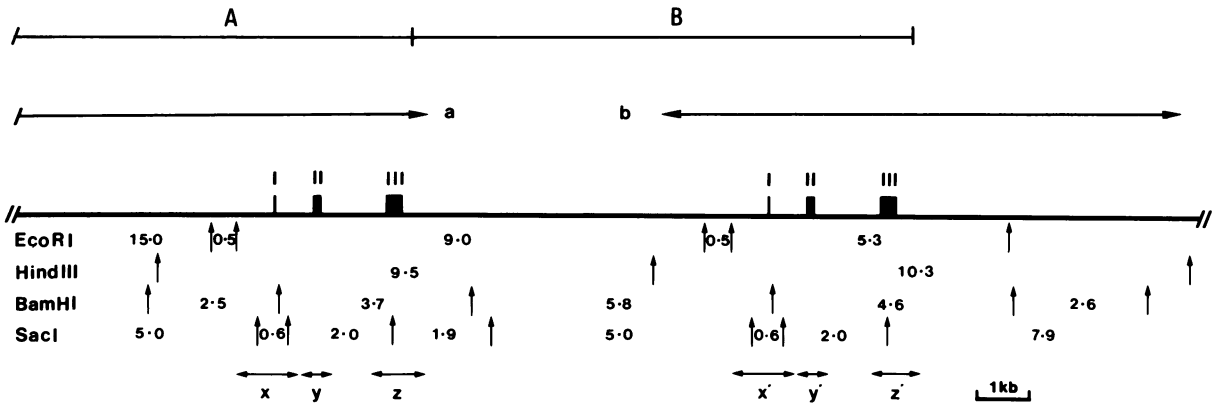


Fig. 2. Partial physical map of the α-cardiac actin locus in BALB/c mice. The black boxes numbered I, II and III represent the real and duplicated exonic sequences of the cardiac actin gene. The horizontal bars above the map, A and B, indicated the duplicated region and its homologous sequence in the *bona fide* gene, respectively. The upper horizontal arrows define the limits of the inserts in λgA8 (a) and λIG10 (b), the lower ones (x/x′, y/y′ and z/z′) the regions sequenced. The sizes of restriction fragments are given in kbp.

of the putative murine cap site. This site is perfectly conserved between the chicken and mouse sequences (nucleotides −5 to +2, Figure 3A). Upstream of this, our proposed TATA and CAAT boxes are perfectly conserved within larger regions of homology (nucleotides −33 to −21 and −110 to −91, Figure 3A; 11/12 and 18/21 conserved nucleotides respectively). Beyond these sequences two other regions of homology are noteworthy. One extends from nucleotide −148 to −140 (Figure 3A) where eight out of nine nucleotides are homologous at exactly the same position relative to the cap site in both genes. The other extends from nucleotide −242 to −226 (Figure 3A) where 16 out of 17 residues are identical to the chicken sequence between

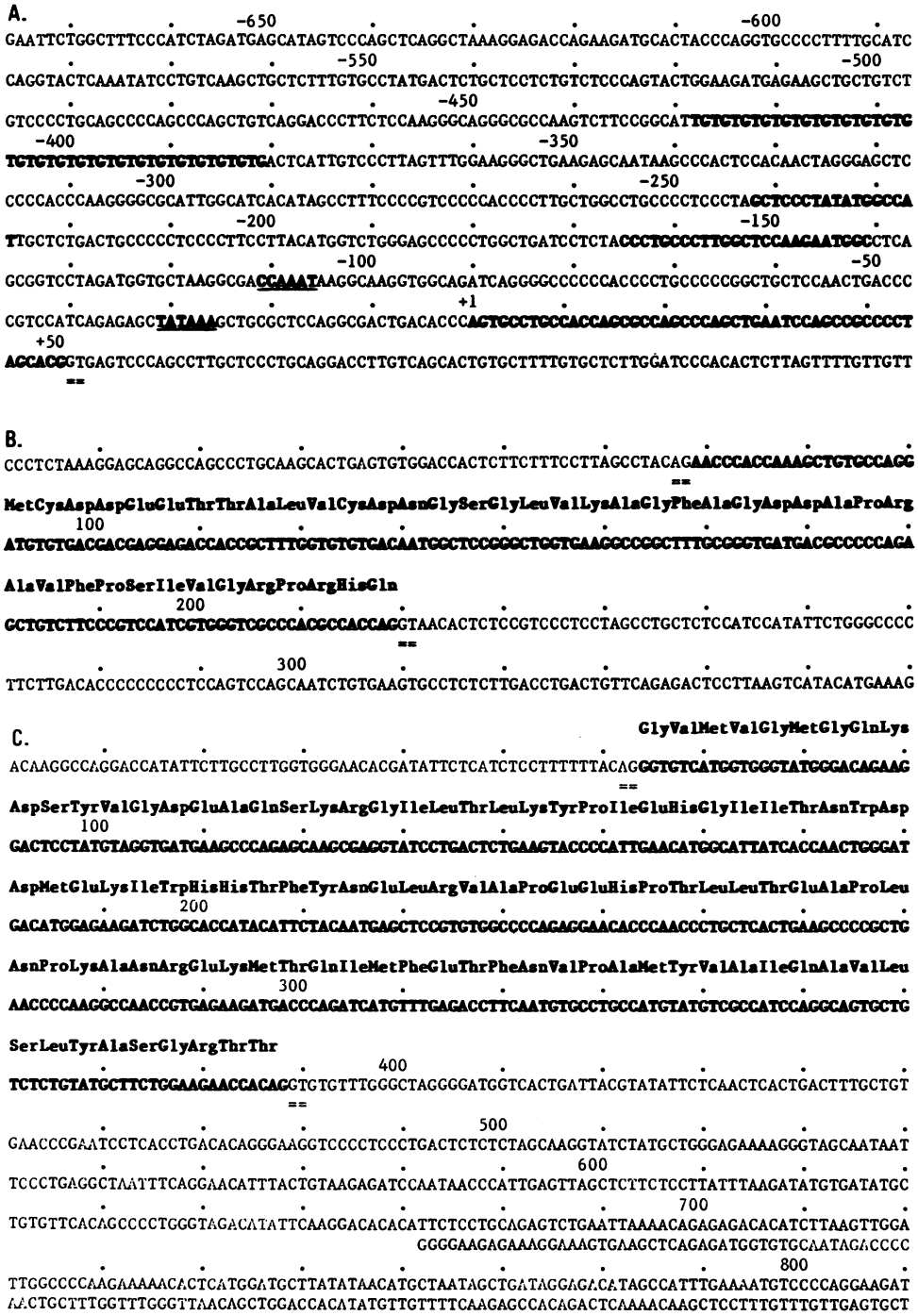


Fig. 3. Comparison of the partial nucleotide sequences of the first three exons and promoter region of the α-cardiac actin gene in BALB/c mice with their homologous regions in the duplication. Regions of the gene and the duplication homologous to the 5′ *Pst*I−*Bgl*II fragment of pAF81 were subcloned into the *Sma*I site of M13mp8 by the 'shotgun' approach. Dideoxy sequencing of recombinant phages generated three pairs of homologous sequences (see x and x′, y and y′, z and z′, Figure 2). Sequences presented represent: **(A)** the promoter region and exon I, x/x′; **(B)** exon II, y/y′; and **(C)** exon III and the 3′-limit of the duplication, z/z′. Promoter features referred to in the text, exon sequences and derived amino acid sequences are presented in bold type. The CAAT and TATA boxes are underlined. Nucleotides in the promoter region are numbered from the cap site (the A residue designated +1). Other sequences are numbered from the first nucleotide presented. Splice junctions bordering exons are underlined (==). The 3′ acceptor site upstream of exon II **(B)** is a proposed signal which conforms to the consensus of Mount (1982). The upper line corresponds to the sequence present in the *bona fide* gene. The lower sequence beginning at nucleotide 672 **(C)** shows where the duplication differs from this.

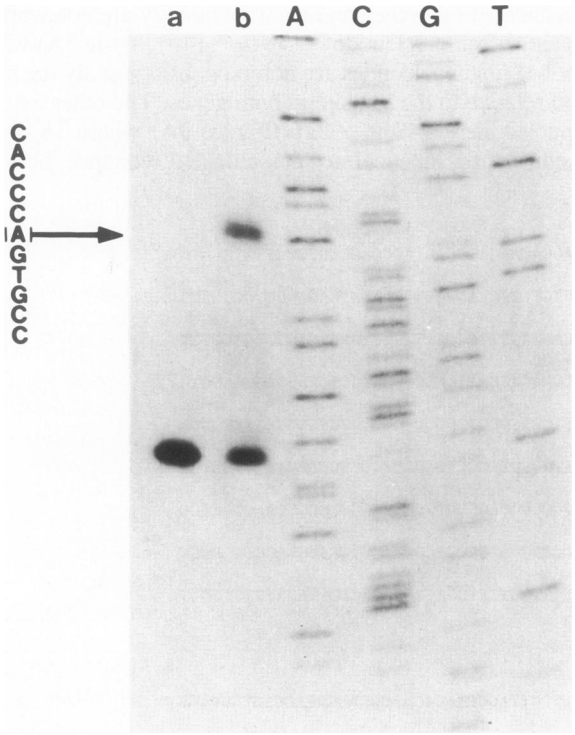


Fig. 4. Analysis of the cap site of α-cardiac actin mRNA. A labelled probe was generated from an M13mp8 recombinant phage containing nucleotides −250 to +53 (Figure 3A) cloned in its *Sma*I site. Extension *in vitro* of the M13 17-mer (-20) primer (NEB) and *Pvu*II restriction yielded a probe which was homologous to nucleotides +27 to +53 and contained 48 nucleotides of M13 sequence. This was hybridised to total adult BALB/c cardiac mRNA **(lane b)** and control yeast tRNA **(lane a)**. Both were then extended with reverse transcriptase. The products were migrated on a sequencing gel next to markers generated from the same recombinant phage (lanes ACGT). For clarity, the labelling of these tracks has been reversed to give the same sequence as the DNA sequence presented in Figure 3A. The cap site is indicated on the partial sequence presented to the left of the figure.

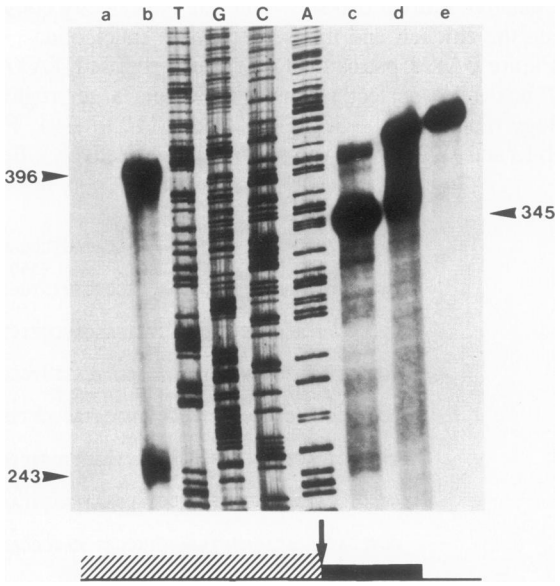


Fig. 5. S1 analysis of transcripts derived from the duplication and the *bona fide* gene. A probe (396 nucleotides) was generated from an M13mp8 recombinant phage containing nucleotides 429−773 of the duplication (Figure 3C) cloned in its *Sma*I site. Extension of the M13 17-mer (-20) primer (NEB) and *Eco*RI restriction yielded a probe of 396 nucleotides with homology to transcripts of both the gene and the duplication. Its structure is represented schematically under the photograph. The hatched box represents the 243 nucleotides of common homology to both duplication and primary gene transcripts. The arrow indicates the end of the duplication and hence of the homology in this region. The black box represents the continued homology (102 nucleotides) with transcripts derived from the duplication. The black line represents M13 sequence. This probe was hybridised to: **lane a**, 50 μg of yeast tRNA; **lane b**, 50 μg of total adult C3H cardiac RNA; **lane c**, 50 μg of total adult BALB/c cardiac RNA; and **lane d**, 0.5 μg of poly(A)$^+$ adult BALB/c cardiac RNA. Digestion with S1 nuclease results in protected bands of 242−246 nucleotides **(lanes b** and **c)** and 345−357 nucleotides **(lanes c** and **d)**. **Lane e** contains 1/10 the amount of probe used in **lanes a−d**, this was not treated with nuclease. The sequencing products of reactions using M13mp8 as template were co-migrated as size markers; nucleotides 243, 345 and 396 within this sequence are indicated.

nucleotides −182 and −198. We have previously noted a potential E1A enhancer type core sequence in the 5′ region of the murine alkali myosin light chain $MLC1_F$ promoter (Daubas *et al.*, 1985). The sequence between nucleotides −150 and −141 (Figure 3A) is homologous to this sequence at seven out of 10 positions. Immediately upstream of both of these sequences, a 17-nucleotide region is 75% conserved between the two mouse genes (Daubas *et al.*, 1985; nucleotides −149 to −165, Figure 3A). A final feature of interest in the mouse actin sequence is the dinucleotide repeat motif $(TG)_{24}$ between nucleotides −382 and −428 (Figure 3A).

To confirm that this is indeed the murine promoter region and to define the 5′ extremity of the mRNA, we analysed this region by primer extension experiments. The probe used was generated from an M13mp8 recombinant and was homologous to nucleotides +27 to +53, Figure 3A. The result of such an experiment is shown in Figure 4. A single novel band 26 nucleotides bigger than the probe is generated with BALB/c cardiac RNA (lane b). No such bands were generated with control RNA (lane a). This clearly defines the 5′ end of the BALB/c α-cardiac actin mRNA to the 'A' residue marked '+1' in Figure 3A and indicated in Figure 4. The first 'T' of the proposed TATA box and the first 'C' of the proposed CAAT box are 31 and 111 nucleotides upstream of this point, respectively. We propose, therefore, that this region constitutes the α-cardiac actin promoter in mice with transcription beginning at the 'A' residue noted above.

We next determined the size of the first non-coding exon using probes similarly generated from recombinant M13mp8 phage and S1 protection experiments. Here, we protected fragments of 50−54 nucleotides (data not presented). Chang *et al.* (1985) and Eldridge *et al.* (1985) both defined the homologous splice junction in the chicken to be CGGT. This sequence is conserved in the mouse gene centered at +50 (Figure 3A); around it, homology is extremely limited. We further propose that this splice signal has been functionally conserved in the mouse and that exon I is 50 nucleotides long beginning at nucleotide +1 in Figure 3A. The position of this exon is indicated in Figure 2. The sizes of these protected fragments and an inability to extend the primer used above beyond the 'A' residue noted suggests that should the duplicated promoter be active, transcripts from it do not extend through the promoter region of the *bona fide* gene.

Both promoters are transcriptionally active

Essentially, the structure of this mutant locus is that of a gene with tandem identical promoters upstream. Two questions immediately arise; is the upstream promoter transcriptionally active, and, if so, what effect does this have on the downstream promoter? To address these questions we analysed RNAs produced from this locus using S1 protection experiments. Transcripts traversing the 3′ limit of the duplication would diverge from those of the *bona fide* gene at nucleotide 672 (Figure 3C). We there-

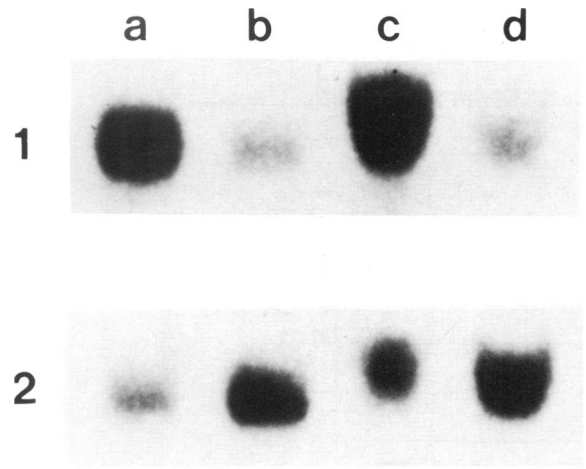


Fig. 6. Northern analysis of cardiac RNAs. 500 ng equivalents of poly(A)$^+$ adult cardiac RNA were denatured using glyoxal, run on an agarose gel and blotted onto a diazotised Biodyne A membrane (Pall, US). This filter was sequentially hybridised to an α-cardiac actin mRNA-specific probe **(1)** and an α-skeletal actin mRNA-specific probe **(2)**. The former has homology to nucleotides 1−50 (Figure 3A) and was generated from an M13 recombinant phage. The latter is the *Pst*I−*Pst*I fragment from pAM91.1 and is uniquely homologous to the 3′ non-coding portion of the α-skeletal actin mRNA (see Minty *et al.*, 1982). **Lanes: a**, poly(A)$^+$ C3H RNA; **b**, poly(A)$^+$ BALB/c RNA; **c**, total C3H RNA; **d**, total BALB/c RNA. The region of the blot containing the mature α-cardiac messages (1600 nucleotides) is presented.

fore used an M13mp8 recombinant homologous to the duplicated region (nucleotides 429−773, Figure 3C) to generate a probe of 396 nucleotides. This contained 345 nucleotides with homology to RNA derived from the duplication but only 243 residues homologous to transcripts from the gene itself. When hybridised to BALB/c and C3H cardiac RNA and digested with S1 nuclease, two sizes of protected species are observed (Figure 5). Total RNA from hearts of mice which do not possess the duplication [C3H, 1% poly(A)$^+$] protects 242−246 nucleotides of the probe (lane b). This is consistent with protection expected from transcripts of the *bona fide* gene. In contrast to this, total RNA from BALB/c hearts [2% poly(A)$^+$] protects both 242−246 and 345−357 nucleotides of the probe (lane c). Poly(A)$^+$ RNA from adult BALB/c hearts protects only 345−357 nucleotides of the probe (lane d).

These data indicate that both of these regions are transcribed in adult BALB/c cardiac tissue. In view of our analysis of the cap site, it is likely that the two regions are transcribed from separate promoters. The fact that apparently normal α-cardiac actin mRNA is produced in BALB/c hearts (see below) indicates that the gene promoter itself is functional. It remains possible however that transcripts of the duplication initiate at a third, cryptic promoter located within it. Our analysis of abnormal actin RNAs present in BALB/c cardiac tissue (see below) suggests that this is not the case and leads us to conclude that the two α-cardiac actin promoters are functional in this situation. However, they appear to be operating with widely different efficiencies. Transcripts of the duplication appear to accumulate to much higher levels than primary transcripts of the gene (compare band intensities in lane c, Figure 5). Moreover, we detect more gene precursor RNA in C3H cardiac RNA than in BALB/c cardiac RNA (compare the relative intensities of the 242−246 nucleotide protected species in Figure 5). Densitometry scans indicate that there is 5-fold more α-cardiac actin precursor per μg of poly(A)$^+$ RNA in C3H preparations than in BALB/c preparations. Thus, although both promoters are active, the *bona fide* gene promoter appears to be inhibited in this situation.

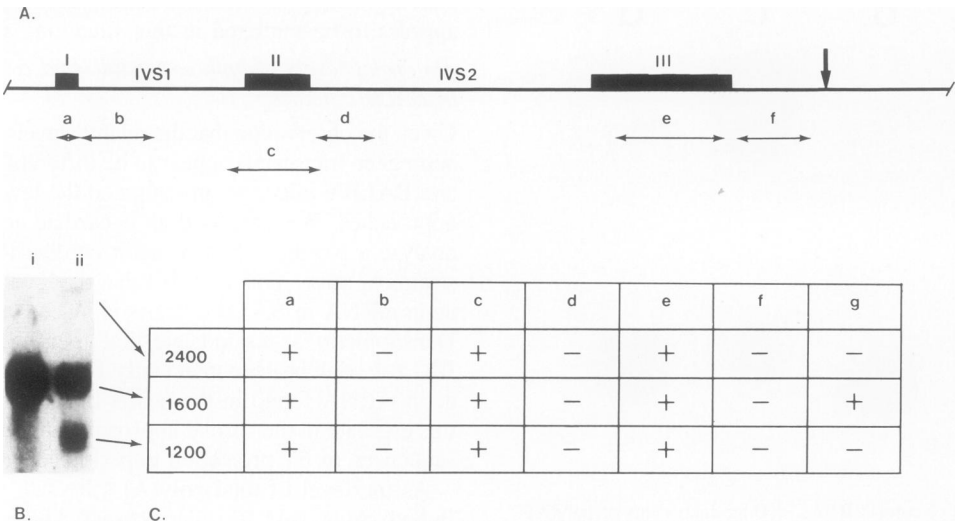
Abnormal levels of both α-cardiac and α-skeletal actin mRNA in BALB/c hearts

Given the observation that the relative levels of primary α-cardiac actin gene transcripts appear to be different in the hearts of C3H and BALB/c mice, we investigated the levels of mRNA by two approaches. We first used an α-cardiac actin-specific probe to analyse a Northern blot of adult cardiac RNA from these two strains of mice. This revealed that the level of mature α-cardiac actin mRNA in BALB/c hearts is indeed reduced (Figure 6.1). Densitometry scans indicate that, per μg of poly(A)$^+$ RNA, BALB/c adult hearts contain only 16.5% of the level of α-cardiac actin mRNA found in the hearts of C3H mice. We confirmed this estimate in the second approach using probes covering exon sequences in S1 protection experiments (data not shown).

As the level of total poly(A)$^+$ RNA/g of tissue is the same in both cases, BALB/c mice possess at least 6-fold less α-cardiac actin mRNA per g of adult cardiac tissue than do C3H mice. With this in mind, we investigated whether there was a 'reactivation' of the α-skeletal actin gene since these two genes are co-expressed during the development of striated muscle (Minty *et al.*, 1982; Gunning *et al.*, 1983; Mayer *et al.*, 1984) and α-skeletal actin gene transcripts are still present in adult cardiac tissue, albeit at low levels. We therefore asked the question 'is the level of α-skeletal actin mRNA stimulated in BALB/c hearts?' The same Northern blot (Figure 6.1) was dehybridised and hybridised to the 3′ non-coding sequence of mouse skeletal actin mRNA as a specific probe for these gene transcripts (Minty *et al.*, 1981). The result clearly demonstrates that there is more α-skeletal actin mRNA in BALB/c hearts than in C3H hearts (Figure 6.2). Densitometry scans indicate a 5:1 ratio between these two samples in favour of BALB/c preparations. This does not restore the overall actin mRNA level to that observed in C3H hearts. On the basis of the figure of 0.5% of total actin mRNA in normal adult hearts of small rodents being α-skeletal actin mRNA (Mayer *et al.*, 1984) we would estimate that the mRNA for this actin isoform rises to ~15% that of α-cardiac actin mRNA in the hearts of BALB/c mice.

BALB/c hearts contain novel mRNAs homologous to actin sequences

Longer exposure of the filter presented in Figure 6.1 reveals the presence of novel mRNAs that hybridise to the α-cardiac actin mRNA-specific probe (Figure 7A, probe a; Figure 7B, lanes i and ii). Two bands are observed in BALB/c poly(A)$^+$ preparations which are not present in C3H RNA, one of 2400 nucleotides and a second of 1200 nucleotides. These are present at 4% and 25% the level of mature α-cardiac actin mRNA, respectively. They could arise from the duplication, the *bona fide* gene or combinations of the transcripts from both of these regions. To investigate this, we first rehybridised the blot with a nick-translated probe of the 3′ *Pvu*II−*Pst*I fragment from pAF81 (probe g, Figure 7). This contains sequences homologous to the C-terminal end of the protein beginning within exon V. It hybridises uniquely to the normal message of 1600 nucleotides (results are tabulated in Figure 7C). Thus, transcripts from the complete α-cardiac actin gene seem not to be involved in the production of these RNAs. Rather, they appear to be derived from transcripts of the duplication. The α-cardiac actin mRNA-specific probe used is homologous to the first, non-coding, 50 bp exon of the gene. Furthermore, we were unable to detect transcripts initiating 5′ to the proposed cap site (nucleotide +1, Figure 3A) in our S1



	a	b	c	d	e	f	g
2400	+	−	+	−	+	−	−
1600	+	−	+	−	+	−	+
1200	+	−	+	−	+	−	−

Fig. 7. Northern blot analysis of abnormal transcripts in BALB/c cardiac tissue. **(A)** The structure of the duplication is represented schematically (not to scale). Indicated are exons I, II and III (black boxes), intervening sequences 1 and 2 (IVS1 and 2) and approximate limits of the probes used (horizontal arrows). These probes are homologous to the following regions: probe a, nucleotides +1 to +50, Figure 3A; probe b, nucleotides +78 to +338, Figure 3A; probe c, nucleotides 1−270, Figure 3B; probe d, nucleotides 266−549, Figure 3B; probe e, nucleotides 129−370, Figure 3C; probe f, nucleotides 405−628, Figure 3C; probe g is the 3′ *Pvu*II−*Pst*I fragment from pAF81 containing sequences homologous to exons V−VII of the α-cardiac actin gene (Minty *et al.*, 1982). The vertical arrow defines the 3′-limit of the duplication. **(B)** Northern blot analysis of transcripts in BALB/c and C3H cardiac tissue with probe a. The filter was that described in Figure 6. **Lanes i** and **ii**, a long exposure (120 h) of lanes 1a and 1b, Figure 6. The sizes of bands are indicated in nucleotides. **(C)** Summary of the hybridisations performed. Probes a−g are as described above. Hybridisation is indicated by '+', no hybridisation is indicated by '−'.

and primer extension analysis of the promoter region. We conclude from this that these novel RNAs in fact initiate at the duplicated promoter region, probably at the duplicated cap site.

The sizes of the 2400- and 1200-nucleotide RNAs are compatible with a model in which transcripts initiate at the duplicated promoter, terminate between the repeat and the gene and are then abnormally processed. An RNA containing all three exons, both introns and little else would be ~2400 nucleotides. Excision of the second intron from this (1200 nucleotides) could then generate the smaller novel RNA. To investigate this possibility, we analysed these RNAs with probes from further M13mp8 recombinant phage homologous to the remaining exon and intervening sequences in the duplication (probes b−f, Figure 7A). Probes corresponding to the remaining exons (c and e) hybridise to both novel RNAs as well as to the 1600-nucleotide mRNA. Probes corresponding to introns (b, d and f) have no detectable homology to RNAs on this filter. Exons I, II and III consist of 526 nucleotides. The novel RNAs, by virtue of their size, must therefore contain other sequences derived from the region between the duplication and the gene. We conclude from these data that transcripts initiate at the duplicated promoter, terminate before they reach the *bona fide* gene and then are spliced and polyadenylated to generate the novel 'actin mRNAs' that we observe. In doing so, they appear to make use of the normal actin splicing signals but must also utilise cryptic 3′ splice and polyadenylation sites. Thus, correct splicing of the 5′ actin exons takes place in the absence of a complete actin gene primary transcript and does not require the 3′ portion of the pre-mRNA.

Discussion

We have characterised a mutant cardiac muscle actin locus present in certain *mus*1 inbred mouse lines. The tandem duplication of the promoter and 5′ exons of the α-cardiac actin gene in these mice appears to result in interference with the transcription of the gene itself. The only other documented example of a mutation affecting a major contractile protein in mammals is that of a mutant cytoplasmic β-actin gene in transformed human fibroblasts (Leavitt *et al.*, 1984). In this case, a single base change in the actin coding sequence results in a modified actin protein which is probably significant in relation to the transformed phenotype of these cells. In invertebrates, such as *Drosophila* or the nematode, mutations which affect movement have been mapped to within tropomyosin (Karlik and Fyrberg, 1985), myosin (MacLeod *et al.*, 1981; Mogami *et al.*, 1986) and actin genes (Karlik *et al.*, 1984). The mutation that we describe is not within the structural gene, but is a tandem duplication of the promoter and 5′ exons of the α-cardiac actin gene which is associated with abnormal levels of α-cardiac actin mRNA. This phenomenon has a number of structural and regulatory implications.

Promoter features

Comparison of the promoter region of this gene with that in the chicken permits us to identify two 5′ regions of sequence homology in addition to perfectly conserved sequences around the cap site, TATA and CAAT boxes. Such localised sequence conservation may signify a functional role for these regions (Schmidt *et al.*, 1984; Daubas *et al.*, 1985). Upstream sequences have been implicated in the tissue-specific regulation of a number of muscle genes including that of the α-skeletal actin (Melloul *et al.*, 1984), troponin I (Konieczny and Emerson, 1985) and α-cardiac actin genes (Minty and Kedes, 1986; I.Garner and P.Daubas, unpublished results). Moreover, a number of muscle-specific genes are activated in heterokaryons by *trans*-acting factors (Wright, 1981; Blau *et al.*, 1983, 1985). It seems likely, therefore, that these conserved regions are the sites of action of such signal molecules affecting the expression of the α-cardiac actin gene. One of the two 5′ conserved regions is an E1A type enhancer core element as proposed by Hearing and Shenk (1983). We have previously noted such a sequence centred at −100 in the promoter region of the alkali myosin light chain $MLC1_F$ (Daubas *et al.*, 1985). Such enhancer sequences have been shown to be functional

in differentiated muscle cells (Felsani *et al.*, 1985). Immediately adjacent to this sequence is a 17-nucleotide sequence:

(CCCTGCCCTTGGCTCCA, −149 to −165)

This is 75% conserved between these two murine striated muscle genes. It may be significant in conferring a 'muscle context' to the enhancer element should it function as such.

Neither the chicken nor the human promoter sequences possess the dinucleotide repeat $(TG)_{24}$ seen in the mouse promoter between nucleotides −382 and −428 although such a sequence is present in the fifth intron of the human gene (Hamada *et al.*, 1982a). Such alternating co-polymers are highly conserved in a variety of eucaryotic genomes (Hamada and Kagunaga, 1982; Hamada *et al.*, 1982b). They can adopt the left handed Z-DNA conformation *in vitro* when subjected to negative torsional stress or high ionic strengths (Wang *et al.*, 1979; Arnott *et al.*, 1980; Haniford and Pulleybank, 1983; Nordheim and Rich, 1983a). They have been shown to occur in the regions surrounding a number of genes (Nordheim and Rich, 1983b) and to affect promoter efficiency in an enhancer-like fashion (Hamada *et al.*, 1984). The human sequence, however, contains not only a region homologous to the E1A enhancer core sequence, but also an SV40 enhancer core sequence at approximately the same position (−390) as the mouse $(TG)_{24}$ repeat (Minty and Kedes, 1986). This latter SV40-type sequence is embedded in an 8-nucleotide purine/pyrimidine repeat that also has Z-DNA potential. The human and mouse promoters, therefore, both contain two distinct non-overlapping enhancer-like sequences. This situation is reminiscent of that seen in polyoma virus, adenovirus E1A and mouse Ig_H genes (Herbomel *et al.*, 1984; Hearing and Shenk, 1983; Hen *et al.*, 1983). DNase I sensitivity studies of the situation in polyoma virus suggest that the two core sequences affect the chromatin structure of the virus (Herbomel *et al.*, 1981, 1984). The sequence pairs that we note above may behave similarly. Should this be the case, we may expect to find DNase I-sensitive sites around these regions of the α-cardiac actin gene.

The duplication and its transcripts

The 9.5 kbp of duplicated sequence contains exact repeats of the promoter region, the first three exons and all intervening sequences that we have examined. The 3′ limit of this region occurs within the third intervening sequence. The absence of mutations and the fact that it is detected in only some inbred *mus*1 strains (DBA2, BALB/c) suggest that it has arisen very recently. The duplicated promoter appears to be transcriptionally active giving rise to transcripts that terminate between the duplication and the *bona fide* gene. To our knowledge, a situation of this type has been observed in only one other case, the murine α-amylase multigene family (Bodary *et al.*, 1985). Here, four different members of the family are found to possess 5′ duplications containing the promoter region, the first exon and part of the first intron. Transcripts are not detectable from these regions and the duplicated ATG translation initiation signals are mutated to ATA. They differ, therefore, from the situation we observe in two critical aspects.

How might such a phenomenon arise? The 3′ limit of the actin duplication is a stretch of seven alternating purine/pyrimidine residues (ACACACA, Figure 3C, nucleotides 665−671) that is homologous to a stretch of eight alternating residues in the gene itself (ACACACAT, Figure 3C, nucleotides 665−672). Sequences of this type have the potential to adopt alternative DNA conformations and have been implicated in other gene rearrangements (Slightom *et al.*, 1980; Flanagan *et al.*, 1984). Such alternative DNA conformations have been identified as intermediates in eucaryotic recombination events (Kmiec *et al.*, 1985). With this in mind, we suggest that the sequence of eight alternating purine and pyrimidine residues in the third intron of the gene may have promoted the duplication event. Unequal cross-over is the simplest mechanism by which to explain its origin. The structural features of the duplication are consistent with such a hypothesis. Such an event may provide a substrate for the potential evolution of a new promoter specificity in this and other multigene families. Events of this type may have been important in the evolution of the developmental regulation of actin genes.

Transcripts from the duplication are processed using the normal splice signals flanking the actin exons and cryptic splice and polyadenylation sites 3′ to the duplication. These events give rise to RNAs of 2400 and 1200 nucleotides which accumulate in BALB/c hearts. These two RNAs contain all three exons (a total of 526 nucleotides) but lack both intervening sequences 1 and 2 and, in view of their size, must contain other sequences derived from the 3′ end of the primary transcript. Splicing from the end of the third exon must utilise cryptic acceptor sites in the transcript. Such chimaeric splicing events have been described previously (Chu and Sharpe, 1981). Splicing of the actin portion, therefore, proceeds normally in the absence of the 3′ end of the primary gene transcript *in vivo*. Similar truncated transcripts have been shown to function faithfully in *in vitro* splicing reactions (Frendewey and Keller, 1985). The results presented here demonstrate that, in addition, they may do so *in vivo*.

Inhibition of proximal promoters

The situation we describe comprises two identical promoters separated by ~9.5 kbp of genomic DNA. The phenomenon is reminiscent of retroviral integration where the provirus is flanked by two long terminal repeats (LTRs). These contain sequences required for transcriptional initiation and polyadenylation of viral transcripts and fulfil these roles at opposite ends of the proviral genome. Cullen *et al.* (1984) showed that early after infection, the 3′ LTR of an avian retrovirus was incapable of transcribing an adjacent gene when the 5′ LTR was transcriptionally active. Transcripts extending through the 3′ LTR completely inhibit its promoter activity. Premature termination of these transcripts relieves inhibition of the 3′ promoter. They called this effect 'transcriptional interference'. Such an effect cannot explain the phenomenon which we observe as transcripts from the 5′ promoter do not appear to extend through the downstream promoter.

Emerman and Temin (1984) constructed retroviral vectors containing two genes, each with its own promoter. They found that the 3′ gene is suppressed when there is selection for the expression of the 5′ gene. Interestingly, the converse situation applied as well, where 'transcriptional interference' could not operate. They proposed a model in which transcription from one promoter can inhibit that from a nearby promoter by mutually exclusive chromatin structures. They noted that such an inhibition is not absolute and that, in some cases, transcription can occur from both promoters. The situation that we describe is compatible with such a model although we cannot eliminate the possibility that splicing events also play a role here. Chromatin structures could affect the accessibility of each promoter to transcriptional factors or simply render them transcriptionally inactive. This could result in either promoter being completely inhibited in a given cell or promoter competition. Our data do not distinguish between these possibilities as our RNA preparations were prepared from a mixed cellular population. The situation we observe could, therefore, represent the global result of such an effect. Whatever the mechanism, these data suggest that such effects may regulate gene expression in eucaryotic cells. A number of loci have been described that can be transcribed from two proxi-

mal promoters including a muscle myosin light chain gene (see Robert *et al.*, 1984). Perhaps such mechanisms operate in the transcriptional regulation of these genes.

Actin gene expression

During striated muscle development there is significant expression of both sarcomeric actins whereas in adult cardiac or skeletal muscle tissue one isoform predominates. Thus, in mice, skeletal actin represents <5% of the total actin in adult hearts (Vandekerckhove *et al.*, 1986). Cardiac and skeletal actin proteins differ by only four out of 375 amino acids. Although the functional significance of these differences is not clear, they may represent an optimalisation to the physiological requirements of each adult tissue. The preferential expression of one member of the gene pair may also reflect regulatory constraints at the genome level.

We have shown that, in BALB/c mice, when the level of α-cardiac actin mRNA is reduced to 16.5% of its normal level in adult cardiac tissue, the level of α-skeletal actin mRNA is increased 5-fold. A similar increase is seen in the level of α-skeletal actin mRNA in adult rat hearts under cardiac overload induced by aortic constriction (Schwartz *et al.*, 1986). This situation differs from that which we observe in that the level of α-cardiac actin mRNA in these hypertrophic hearts remains high. How might such a phenomenon arise? Mayer *et al.* (1984) demonstrated that in adult rat hearts, the α-skeletal actin gene is DNase I sensitive whereas this is not the case in non-muscle cells. Moreover, there is already a low level of mRNA derived from this gene in normal adult rodent hearts (Minty *et al.*, 1982; Mayer *et al.*, 1984). Thus, the α-skeletal actin gene in adult hearts appears to be in an active transcriptional state. The increase in the level of this mRNA in BALB/c adult hearts may be the result of an mRNA stabilisation effect, although there is no precedent for this in actin gene regulation in muscle. Alternatively, it may reflect a transcriptional effect on the α-skeletal gene. It is also possible that individuals have been selected which carry a compensatory mutation in the α-skeletal actin locus resulting in higher level of this transcript. Given the result of Schwartz *et al.* (1986) where a similar effect is pathologically induced in individual animals, such a selection phenomenon seems improbable. It therefore seems more likely that there is a direct effect on the level of transcription of the α-skeletal actin gene in these mice. In the situation we observe, the contribution of the α-skeletal actin mRNA to the total actin levels in BALB/c hearts is not sufficient to restore normal concentrations. We estimate that it boosts these levels to ~20% of that observed in C3H mice. To maintain actin protein levels, it may be necessary to increase translation rate or protein half-life. Such effects would constitute levels of post-transcriptional control in actin gene expression. Whatever the extent of such adaptations, no hypertrophy is evident (I.Garner and J.L.Guénet, unpublished results) and it is clear that cardiac function and viability are not deleteriously affected in BALB/c mice. Furthermore, the two additional RNA species of 1200 and 2400 nucleotides generated from the duplicated region of the cardiac actin gene in BALB/c mice do not appear to interfere with cardiac function. They contain α-cardiac actin coding sequence up to amino acid 149 and could be translated to give rise to novel proteins (≥149 amino acids) beginning with the first domain of the α-cardiac actin protein (Kabsch *et al.*, 1985). The mechanical force of muscular contraction is generated at the interface between myosin heavy chain and actin in the sarcomere. Cross-linking experiments reveal that the actin residues involved in this interaction are numbers 1, 2, 3, 4 and 11 (Sutoh, 1982). These would all be present in proteins translated from the 1200- and 2400-nucleotide RNAs. However, it seems probable that such hybrid actin molecules, were they produced, would not be capable of polymerisation into the fine filaments of the sarcomere.

The increase in α-skeletal transcripts that we observe is important in two respects. Firstly, it suggests that the repression of α-skeletal actin mRNA synthesis in adult cardiac tissue is reversible. Secondly, it implies that the mechanisms controlling the expression of these two genes are sensitive to the levels of α-actin mRNA or protein. These observations are compatible with a model of regulation for these genes including multiple levels of control such that a basal level of regulation may be provided by the chromatin structure in which a gene is situated. Alterations in this during muscle development may allow transcription of this gene pair. During maturation of the muscle, the interaction with modulatory factors (positive or negative) may result in tiers of regulation allowing greater flexibility in the expression of cardiac and skeletal muscle actin genes.

Materials and methods

Isolation of recombinant phage

This was performed essentially as described in Robert *et al.* (1984). Two recombinant λ phage libraries (each 10^6 phage in 50 dishes) were screened *in situ*. One was a partial *Hae*III/*Alu*I digest of BALB/c mouse sperm DNA cloned in λ Charon 4A with *Eco*RI linkers (referred to in Robert *et al.*, 1984) and gave rise to λgA8. The second was a partial *Sau*3A digest of BALB/c DNA cloned between the *Bam*HI site of λ47.1 by standard procedures and gave rise to λIG10. Both were screened with the 5′ *Pst*I−*Bgl*II fragment of pAF81. All positive phage were plaque purified and further screened with the 3′ *Pvu*II−*Pst*I fragment of pAF81. Phage of interest were purified on CsCl gradients and DNA extracted by phenol extraction.

Restriction mapping

Physical mapping of recombinant phage was performed by a combination of single, double and partial restrictions. The latter were selectively labelled at either Cos site using synthetic oligonucleotides complementary to the right or left Cos site as described by Rackwitz *et al.* (1984). The order of restriction fragments was easily determined following gel electrophoresis and autoradiography. Restriction mapping of the α-cardiac actin locus in BALB/c DNA was performed similarly but digests were analysed by Southern blot using fragments derived from phages λgA8 and λIG10 as probes.

Generation of probes

Purified DNA restriction fragments were nick-translated as described in Minty *et al.* (1983). Probes were generated from M13 recombinants by strand synthesis *in vitro*. Following hybridisation to the M13 17-mer (-20) primer (NEB), this was extended *in vitro* in the presence of labelled dNTPs using DNA polymerase I Klenow fragment. Suitable restriction digests, denaturing gel electrophoresis and electro-elution yielded region-specific single-stranded probes containing M13 sequences up to and including the primer.

Nucleotide sequencing

The sequences of regions hybridising to the 5′ *Pst*I−*Bgl*II fragment of pAF81 were determined. The 5′ end of the 9-kbp *Eco*RI fragment (3.5 kbp) present in λgA8 was subcloned into pBR322 to generate plasmid $pAE8_1$. This plasmid (duplication) and phage λIG10 (gene) were sonicated (10 μg each) to generate random fragments. These were end-repaired with DNA polymerase I Klenow fragment (Boehringer Mannheim) and separated on agarose gels. Fragments of 400−600 bp were purified by electro-elution and ligated into the *Sma*I site of M13mp8. Following transfection (Hanahan, 1983) desirable recombinants were identified by plaque hybridisation using the insert in $pAE8_1$ and the 5′ *Pst*I−*Bgl*II fragment from the pAF81 as probes. These were sequenced by the dideoxy method (Sanger *et al.*, 1977; Biggin *et al.*, 1983).

Southern and Northern blots

Mouse DNA was restricted with excess enzyme, run on a 1% agarose gel and transferred to a nitrocellulose filter as described in Minty *et al.* (1982). Blots were revealed by autoradiography following overnight hybridisation with probes in 50% formamide, 0.9 M NaCl, 50 mM sodium phosphate (pH 7.0), 5 mM EDTA, 0.1% SDS, 0.2% Ficoll 400, 0.2% polyvinylpyrrolidine, 0.2% bovine serum albumin and 100 μg/ml denatured sonicated salmon sperm DNA at 42°C. Final washings were in 0.1 × SSC and 0.1% SDS at 65−70°C.

RNA samples were prepared, denatured with glyoxal, separated on an agarose

gel and blotted onto a diazotised Biodyne A membrane (Pall, US) as described in Caravatti *et al.* (1982). Hybridisations were as above but final washings were in 0.1 × SSC and 0.1% SDS at 60°C. Exposure times for Northerns and Southerns varied between 2 and 120 h.

S1 and primer extension experiments

Probes were generated for both of these techniques from M13mp8 recombinants. For S1 mapping, 100 000 c.p.m. (Cerenkov) of probe were precipitated with 10–50 μg total or 0.1–0.5 μg poly(A)$^+$ mouse cardiac RNA and resuspended in 10 μl of 50% formamide, 0.9 M NaCl, 50 mM sodium phosphate buffer (pH 7.0), 1 mM EDTA. Hybrids formed after a 15 h incubation at the desired temperature were expelled into 200 μl of S1 buffer and digested with 5000 U of enzyme (Boehringer Mannheim) for 1 h at 30°C. The products were analysed on 6% acrylamide sequencing gels with M13mp8 sequencing reaction products as markers.

For primer extension experiments, 5000 c.p.m. (Cerenkov) of probe was precipitated with 20 μg of total mouse cardiac RNA and resuspended in 10 μl of 100 mM NaCl, 20 mM Tris-HCl (pH 8.3), 100 mM EDTA. After 3 h incubation at 60°C, primers were extended by dilution with 10 μl 2× concentrated buffer (80 mM Tris-HCl, pH 8.3, 10 mM $MgCl_2$, 4 mM dithiothreitol, 4 mM of each cold dNTP) and the addition of 10 U of AMV reverse transcriptase. Elongated products were analysed on denaturing 6% acrylamide gels next to the products of sequencing reactions performed on the template used to generate the primer.

Acknowledgements

We are grateful to Dr G.Hatfull for his advice on DNA sequencing, to J.L.Guénet for assistance with inbred mice lines, to Dr S.Wain-Hobson, Dr C.Bishop and Dr A.Weydert for help with the λ47.1 library and to Professor F.Gros for encouragement and support. We acknowledge the comments of our colleagues on the manuscript. The laboratory is funded by the Institut Pasteur, the Centre Nationale de la Recherche Scientifique, the Ministére de l'Industrie et de la Recherche, and the Muscular Dystrophy Association of America. During the course of this work, I.G. was the recipient of fellowships from the British Royal Society and the Pasteur/Weizmann foundation, P.J.B. was the recipient of an EMBO fellowship.

References

Alonso,S., Minty,A., Bourlet,Y. and Buckingham,M.E. (1986) *J. Mol. Evol.*, **23**, 11–22.

Arnott,S., Chandrasekaran,R., Birdsall,D.L., Leslie,A.G.W. and Ratcliff,R.L. (1980) *Nature*, **283**, 743–745.

Bains,W., Ponte,P., Blau,H. and Kedes,L. (1984) *Mol. Cell. Biol.*, **4**, 1449–1453.

Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.

Blau,H., Chiu,C-P. and Webster,C. (1983) *Cell*, **32**, 1171–1180.

Blau,H., Pavlath,G.K., Hardeman,E.C., Chiu,C-P., Silberstein,L., Webster,S.G., Miller,S.C. and Webster,C. (1985) *Science*, **230**, 758–766.

Bodary,S., Grossi,G., Hagenbuchle,O. and Wellauer,P.K. (1985) *J. Mol. Biol.*, **182**, 1–10.

Buckingham,M.E. (1985) *Essays Biochem.*, **20**, 77–109.

Caravatti,M., Minty,A.J., Robert,B., Monterass,D., Weydert,A., Cohen,A., Daubas,P. and Buckingham,M.E. (1982) *J. Mol. Biol.*, **160**, 59–76.

Carmon,Y., Czosnek,H., Nudel,U., Shani,M. and Yaffe,D. (1982) *Nucleic Acids Res.*, **10**, 3085–3097.

Chang,K.S., Rothblum,K.N. and Schwartz,R.J. (1985) *Nucleic Acids Res.*, **13**, 1223–1249.

Chu,G. and Sharp,P.A. (1981) *Nature*, **289**, 378–382.

Cullen,B.R., Lomedico,P.T. and Ju,G. (1984) *Nature*, **307**, 241–245.

Czosnek,H., Nudel,U., Shani,M., Barker,P.E., Pravtcheva,D.D., Ruddle,F.H. and Yaffe,D. (1982) *EMBO J.*, **1**, 1299–1305.

Czosnek,H., Nudel,U., Mayer,Y., Barker,P.E., Pravtcheva,D.D., Ruddle,F.H. and Yaffe,D. (1983) *EMBO J.*, **2**, 1977–1979.

Daubas,P., Robert,B., Garner,I. and Buckingham,M.E. (1985) *Nucleic Acids Res.*, **13**, 4623–4643.

Eldridge,J., Zehner,Z. and Paterson,B.M. (1985) *Gene*, **36**, 55–63.

Emerman,M. and Temin,H.M. (1984) *Cell*, **39**, 459–467.

Felsani,A., Maione,R., Ricci,L. and Amati,P. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 753–757.

Flanagan,J.G., Lefranc,M-P. and Rabbitts,T.H. (1984) *Cell*, **36**, 681–688.

Frendewey,D. and Keller,W. (1985) *Cell*, **42**, 355–367.

Gunning,P., Ponte,P., Blau,H. and Kedes,L. (1983) *Mol. Cell. Biol.*, **3**, 1985–1995.

Gunning,P., Ponte,P., Kedes,L., Eddy,R. and Shows,T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1813–1817.

Hamada,H. and Kakunaga,T. (1982) *Nature*, **298**, 396–398.

Hamada,H., Petrino,M.G. and Kagunaga,T. (1982a) *Proc. Natl. Acad. Sci. USA*, **79**, 5901–5905.

Hamada,H., Petrino,M.G. and Kagunaga,T. (1982b) *Proc. Natl. Acad. Sci. USA*, **79**, 6465–6469.

Hamada,H., Seidman,M., Howard,B.H. and Gorman,C.M. (1984) *Mol. Cell. Biol.*, **4**, 2622–2630.

Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557–580.

Haniford,D.B. and Pulleybank,D.E. (1983) *Nature*, **302**, 632–634.

Hearing,P. and Shenk,T. (1983) *Cell*, **33**, 695–703.

Hen,R., Borelli,E., Sassone-Corsi,P. and Chambon,P. (1983) *Nucleic Acids Res.*, **11**, 8747–8749.

Herbomel,P., Saragosti,S., Blangy,D. and Yaniv,M. (1981) *Cell*, **25**, 651–658.

Herbomel,P., Bourachot,B. and Yaniv,M. (1984) *Cell*, **39**, 653–662.

Kabsch,W., Mannherz,H.G. and Suck,D. (1985) *EMBO J.*, **4**, 2113–2118.

Karlik,C.C. and Fyrberg,E.A. (1985) *Cell*, **41**, 57–66.

Karlik,C.C., Coutu,M.D. and Fyrberg,E.A. (1984) *Cell*, **38**, 711–719.

Kmiec,E.B., Angelides,K.J. and Holloman,W.K. (1985) *Cell*, **40**, 139–145.

Konieczny,S.F. and Emerson,C.P. (1985) *Mol. Cell. Biol.*, **5**, 2423–2432.

Leavitt,J., Gunning,P., Porreca,P., Ng,S-Y., Lin,C-S. and Kedes,L. (1984) *Mol. Cell. Biol.*, **4**, 1961–1969.

MacLeod,A.R., Karn,J. and Brenner,S. (1981) *Nature*, **291**, 386–390.

Mayer,Y., Czosnek,H., Zeelon,P.E., Yaffe,D. and Nudel,U. (1984) *Nucleic Acids Res.*, **12**, 1087–1100.

Melloul,D., Aloni,B., Calvo,J., Yaffe,D. and Nudel,U. (1984) *EMBO J.*, **3**, 983–990.

Minty,A.J. and Kedes,L. (1986) *Mol. Cell. Biol.*, **6**, 2125–2136.

Minty,A.J., Caravatti,M., Robert,B., Cohen,A., Daubas,P., Weydert,A., Gros,F. and Buckingham,M.E. (1981) *J. Biol. Chem.*, **256**, 1008–1014.

Minty,A.J., Alonso,A., Caravatti,M. and Buckingham,M.E. (1982) *Cell*, **30**, 185–192.

Minty,A.J., Alonso,A., Guénet,J.-L. and Buckingham,M.E. (1983) *J. Mol. Biol.*, **167**, 77–101.

Mogami,K., O'Donnell,P.T., Bernstein,S.I., Wright,T.R.F. and Emerson,C.P., Jr. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1393–1397.

Mount,S. (1982) *Nucleic Acids Res.*, **10**, 459–472.

Nordheim,A. and Rich,A. (1983a) *Nature*, **303**, 674–679.

Nordheim,A. and Rich,A. (1983b) *Proc. Natl. Acad. Sci. USA*, **80**, 1821–1825.

Ponte,P., Gunning,P., Blau,H. and Kedes,L. (1983) *Mol. Cell. Biol.*, **3**, 1783–1791.

Rackwitz,H-R., Zehetner,G., Frischauf,A-M. and Lehrach,H. (1984) *Gene*, **30**, 195–200.

Robert,B., Daubas,P., Akimenko,M-A., Cohen,A., Garner,I., Guénet,J.-L. and Buckingham,M.E. (1984) *Cell*, **39**, 129–140.

Robert,B., Barton,P., Minty,A.J., Daubas,P., Weydert,A., Bonhomme,F., Catalan,J., Chazottes,D., Guénet,J.-L. and Buckingham,M.E. (1985) *Nature*, **314**, 181–183.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schmidt,A., Yamada,Y. and de Crombrugghe,B. (1984) *J. Biol. Chem.*, **259**, 7411–7415.

Schwartz,K., de la Bastie,D., Bouveret,P., Oliviero,P., Alonso, S. and Buckingham,M. (1986) *Circ. Res.*, in press.

Slightom,J.L., Blechl,A.E. and Smithies,O. (1980) *Cell*, **21**, 627–638.

Sutoh,K. (1982) *Biochemistry*, **21**, 3654–3661.

Vandekerckhove,J. and Weber,K. (1979) *Differentiation*, **14**, 123–133.

Vandekerckhove,J. and Weber,K. (1981) *Eur. J. Biochem.*, **113**, 595–603.

Vandekerckhove,J. and Weber,K. (1984) *J. Mol. Biol.*, **179**, 391–413.

Vandekerckhove,J., Bugaisky,G. and Buckingham,M.E. (1986) *J. Biol. Chem.*, **241**, 1836–1843.

Wang,A.H-J., Quigley,G.J., Kolpak,F.J., Crawford,J.L., van Boom,J.H., van der Marel,G. and Rich,A. (1979) *Nature*, **282**, 680–686.

Wright,W.E. (1981) *J. Cell. Biol.*, **91**, 11–16.

Received on 9 July 1986