A family of muscle gene promoter element (CArG) binding activities in *Xenopus* embryos: CArG/SRE discrimination and distribution during myogenesis

Michael V. Taylor*
Cancer Research Campaign Molecular Embryology Research Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Received February 21, 1991; Revised and Accepted April 18, 1991

**ABSTRACT**

The CArG box is an essential promoter sequence for cardiac muscle actin gene expression in *Xenopus* embryos. To assess the role of the CArG motif in promoter function during *Xenopus* development, the DNA-binding activities present in the embryo that interact with this sequence have been investigated. A family of four Embryo CArG box1 Factors (ECFs) was separated by a 2-step fractionation procedure. These factors were distinct from the previously described CArG box binding activity Serum Response Factor (SRF). ECF1 was the most prominent binding activity in cardiac actin-expressing tissues, and bound the CArG box in preference to a Serum Response Element (SRE). SRF was also detectable in muscle, but it bound preferentially to an SRE. The properties of ECF3 were similar to those of ECF1, but it was much less prominent in cardiac actin-expressing tissues. The properties of the other two factors were distinctive: ECF2 was of relatively low affinity and high abundance, whilst ECF4 bound non-specifically to ends of DNA. The binding activity (or activities) that interacted with the CArG box was found to be influenced by both the concentrations of the other CArG box binding activities and the sequence of the site. Although there was no evidence for a muscle-specific CArG box binding activity, the properties of ECF1 suggest that it could play a role in the expression of the cardiac actin gene during *Xenopus* development.

**INTRODUCTION**

In the frog *Xenopus laevis* the cardiac actin gene is first activated during gastrulation, and expression is maintained during somite formation and axial muscle differentiation. As in other vertebrate embryos the cardiac actin gene is co-expressed with the skeletal actin gene both in the myotomes of the somites, which give rise to the axial musculature, and in the developing heart (1, 2, 3). Our interest is the characterisation of the promoter sequences that control this pattern of cardiac actin gene expression and of the DNA binding activities that interact with them. Although most analyses of muscle cell differentiation and gene activity during this process have used established cell lines, we have chosen to analyse muscle differentiation in normal embryogenesis for two reasons. Firstly, it is the later stages of muscle differentiation that are best studied in these cell lines, which are already committed to myogenesis. Secondly, it is uncertain how far these lines reflect the events of normal differentiation, since many investigators have found anomalies in the muscle gene expression properties of these cells (4, 5, 6).

The *Xenopus* cardiac actin promoter, like those of all vertebrate muscle actins, contains multiple copies of a motif known as the CArG box (7, 8, 9). The most proximal of four CArG motifs, CArG box1 centred at -85, is necessary for embryonic expression of this gene after microinjection into fertilised *Xenopus* eggs (9). Because the three more distal CArG box motifs are not necessary for this expression (9), attention has focused on CArG box1 and the *Xenopus* embryo DNA binding proteins that interact with it. One CArG box1 binding activity has been characterised previously in extracts from late neurulae. The properties of this binding activity were indistinguishable from those of human Serum Response Factor (SRF) (10), and cloning of the frog homologue has now positively identified this embryo extract activity as *Xenopus* SRF (11).

SRF was first characterised through its interaction with the Serum Response Element (SRE) of the c-fos proto-oncogene. The SRE, which is a similar sequence to the CArG box, is a necessary and sufficient promoter element for transcriptional activation by serum growth factors (12). In contrast, the CArG box is a necessary element in the muscle-specific expression of cardiac actin genes in cell lines (13), as well as in the frog embryo (9). In the chicken skeletal actin gene the CArG box, and its flanking sequence, is sufficient for muscle-specific expression in primary cell cultures (14). These two very different patterns of gene expression could arise because there are CArG box/SRE binding proteins that are tissue-specific or can discriminate between CArG boxes and SREs. It is these possibilities that have been investigated in this paper.

* Present address: Wellcome/CRC Institute of Cancer and Developmental Biology, Tennis Court Road, Cambridge, CB2 1QR, UK
I have identified a family of four CArG box1 binding activities distinct from SRF in the *Xenopus* embryo. They have been characterised in terms of their binding sequence specificity and their distribution during embryonic development. One of them, ECF1, is prominent in embryonic tissues that express cardiac actin and binds to CArG box1 in preference to the c-fos SRE. It could therefore play a role in the control of cardiac actin gene expression in the developing *Xenopus* embryo.

**MATERIALS AND METHODS**

**Xenopus extracts**

The whole cell extract was prepared at 4°C with a cocktail of protease inhibitors to minimise proteolytic attack. There was no nuclear isolation step and so no risk of proteins leaking out from the nuclei. *Xenopus* embryos were cultured and staged as described (9). 1000 stage 18 neurula embryos (a stage when the cardiac actin gene is expressed and muscle has begun to differentiate) were extracted in 8 ml of homogenisation buffer (50 mM Tris.HCl (pH7.9), 25% glycerol, 50 mM KCl, 0.1 mM EDTA, buffer C in ref. 35) plus a cocktail of inhibitors (2 mM DTT, 0.5 mM PMSF, 25 µg/ml aprotinin, 25 µg/ml leupeptin and 10 µg/ml bestatin) in a 2-step process: (i) disruption with a motorised homogeniser, (ii) cell lysis in a Dounce homogeniser. The homogenate was centrifuged (3000g, 20 min), the supernatant recentrifuged (100,000g, 90 min) and stored at -80°C. The small scale extracts from staged or dissected embryos were made by homogenisation in the above buffer as described (28). The stage 18 nuclear extract was as described (9).

Adult frog hearts were perfused extensively in Frog Ringer's solution with 0.2 mg/ml heparin and finally with no heparin to remove the contaminating erythrocytes. The ventricular muscle cut into small pieces was disrupted in an IKA Ultra-Turrax homogeniser in homogenisation buffer for the whole cell extract or, for the nuclear extract, in 1.4M sucrose, 10% glycerol, 15M KCl, 10mM HEPES (pH 7.6), 1 mM EDTA, 0.5 mM spermedine, 0.15 mM spermine plus the cocktail of protease inhibitors. The whole cell extract was then made by homogenisation in a Dounce homogeniser, centrifugation (3000g, 20 min), and re-centrifugation of the supernatant (100,000g, 90 min), which was stored at -80°C. The nuclear extract was made by layering the homogenate onto a cushion of homogenisation buffer with 1.65M sucrose, and centrifugation (75,000g, 60 min). The pelleted nuclei were resuspended in 100 mM KCl, 10% glycerol, 10 mM Hepes (pH 7.6), 3 mM MgCl₂, 0.1 mM EDTA and homogenised. KCl was added to 0.55M whilst stirring for 30 min, the lysate centrifuged (100,000g, 40 min), and stored at -80°C.

**Extract fractionation**

After ammonium sulphate precipitation, the precipitated protein and final supernatant were dialysed against homogenisation buffer without aprotinin or leupeptin and stored at -80°C. The precipitated protein was loaded onto a Bio-Rex 70 cation exchange column, which was washed with running buffer (50 mM Tris (pH 7.9), 10% glycerol, 50 mM KCl, 0.1 mM EDTA, 0.05% NP-40) plus inhibitors (2 mM DTT, 0.5 mM PMSF, 25 µg/ml aprotinin, 25 µg/ml leupeptin). Protein was eluted with successive steps of running buffer plus inhibitors with 0.15, 0.3 and finally 0.6 M KCl. 1.0 M KCl did not elute any additional CArG box1 binding activities. Fractions were dialysed against running buffer minus NP-40 with 20% PEG and stored at -80°C.

**Electrophoretic mobility shift assay (EMSA)**

An inverted repeat of the left-half of the CArG box1 motif was synthesised as a 22-mer with KpnI ends 5'-GCTACC-AAATTGGTACGGTAC-3'. The self-anneled oligo was cloned into the pUC18 KpnI site. An inverted repeat of the right-half, 5'-CTGCCCTTATTAAAGGCACGTCATC-3', was similarly synthesised and cloned. The CArG box1 and Xenopus c-fos SRE plasmids were as described (11). Probes were made by excising the insert with EcoRI and AccI for the left- and right-half CArG box1, and EcoRI and BamHI for the CArG box1 and SRE, and end-labeling with [α-32P] dATP and Klenow fragment.

The standard binding reactions and subsequent gels were as described (28). The amount of protein in each assay was the following unless otherwise stated: unfractionated stage 18 whole cell extract -12 µg, or nuclear extract -1.3 µg; fractions enriched in ECF1 -0.95 µg, in ECF2 -0.7 µg, in ECF3 -0.35 µg, and in ECF4/SRF -1.1 µg; adult heart and dissected embryo extracts -approximately 6 µg; embryonic heart extract -1 µg. Four different non-specific DNAs were used: 1 µg poly(dIl-dc), poly(dI-dc) with gel-purified oligos as competitor when necessary; 0.53 µg sheared salmon sperm DNA plus 0.22 µg Mspl/Sau3A digested pUC18 plasmid for the comparison with previous conditions (10); 0.5 µg uncut pUC18 plus 0.5 µg of specifically digested pUC18 for the ECF4 assays; or 1 µg Mspl/Sau3A digested pUC18 for the assays with proteinase K. Competitor double-strand oligos were made by heating gel-purified oligos to 65°C in 10 mM Tris.HCl (pH8), 5 mM MgCl₂, followed by slow cooling. The specific 32-mer CArG box1 and 26-mer c-fos SRE oligos have been described (9, 36). The non-specific oligos were 5'-GATGAGGGAGCTCCGCAG-3' and 5'-C-TGCGGAGCTCCCTCATC-3', two 18-mers from the coding sequence of Xenopus vimentin (37, 38). For the detection of EMF3 in unfractionated extracts MgCl₂ was omitted from the binding reaction when indicated. These assay conditions are altered in many respects from those used previously (10). The most important changes which allowed the detection of CArG box1 binding activities other than SRF were in the non-specific DNA and the type of protein extract used (see Results).

**RESULTS**

**Fractionation of a Xenopus embryo whole cell extract**

A whole cell extract was made from neurula embryos by a rapid and simple procedure designed to maximise the recovery of intact DNA-binding proteins (see Methods). The extract was used in an electrophoretic mobility shift assay (EMSA) with a CArG box1 DNA probe to detect a number of Embryo CArG box1 Factors (ECFs)(Fig.1A). It was compared with a neurula nuclear extract used previously to characterise a single, SRF-like CArG box1 binding activity (10). Each extract was assayed with either poly(dI-dc), poly(dI-dc) or a mixture of sheared salmon sperm DNA and pUC plasmid as non-specific DNA. Binding activities were revealed using the whole cell extract and poly(dI-dc), poly(dI-dc) that were undetectable using the nuclear extract and salmon sperm/plasmid DNA. The whole cell extract was then fractionated by a two step procedure. Firstly, a 40-70% ammonium sulphate precipitation separated ECF2 from the other CArG box1 binding activities. ECF2 remained in solution (Fig.1B, lane 4). Secondly, the 40-70% cut was fractionated by Bio-Rex 70 ion-exchange chromatography by step-elution with 0.15, 0.3 and 0.6 M KCl. Each fraction was enriched for one or more CArG box1 binding activities (Fig.1B).
A number of experiments indicated that the binding activities that were detectable on CArG box1 were influenced by the concentration of the other CArG box binding activities in the assay. ECF2 was readily detectable in the unfractionated extract when a low concentration of extract was assayed (Fig. 1B, lane 1). However, when the concentration was increased ECF2 binding was greatly diminished, and other CArG box1 binding activities were more prominent (lane 2). The fraction that showed ECF2 binding, when separated from the other CArG box1 binding activities (lane 4), corresponded to approximately 1/10 of an embryo equivalent, whilst the unfractionated extract that had no detectable ECF2 binding (lane 2) corresponded to approximately one embryo equivalent. Similarly, ECF3 was barely detectable in the unfractionated extract, and yet was readily detectable in the fraction depleted of SRF, ECF1, 2 and 4 (lane 5). Alternatively, by not including MgCl₂ in the binding assay ECF1 and SRF bound poorly and ECF3 was again detectable (Fig. 1C). This effect of MgCl₂ was exploited to determine whether or not ECF3 was present in unfractionated, small-scale extracts from staged embryos or dissected embryo pieces (see below).

The binding specificity of the CArG box1 factors
Each of SRF, ECF1, ECF2 and ECF3 in the enriched fractions bound specifically to CArG box1, as demonstrated by competition with CArG box1 oligo, but not non-specific oligo (Fig. 2). SRF binding was competed more readily than ECF1-3. The efficiency of competition is a measure of relative binding affinity (15), and so of these activities SRF had the highest binding affinity for CArG box1. Whilst, the ECF1-, ECF3- and SRF-enriched fractions used in the assay corresponded to approximately one embryo equivalent, the ECF2-enriched fraction corresponded to only 1/10 an embryo. Taken together with its relatively low binding affinity and the proportion of probe bound, this indicates that ECF2 is much the most abundant activity in the embryo extract.

In contrast to these other binding activities, ECF4 was competed by both the specific and the non-specific oligo. One explanation is that ECF4 has an affinity for DNA ends. This idea was supported by a number of experiments. Firstly, ECF4 bound to four different probes, prepared from pUC18, with a similar
affinity to the CArG box1 probe (not shown): it showed no sequence specificity. Secondly, binding assays were carried out with CArG box1 probe and equal amounts of pUC plasmid as non-specific DNA (Fig.3). The plasmid was either uncut, linearised or digested with restriction enzymes to give the indicated number of 3' overhang, 5' overhang or blunt ends. The amount of ECF4 detected decreased according to the number, rather than the type, of ends.

It has been suggested that a CArG box binding activity that migrates in a position similar to ECF1, is a protease-resistant core of SRF (16). A band of similar mobility can be produced by proteinase K treatment of a mix of SRF-containing extract and a CArG/SRE probe (16, 17). This observation was reproduced (Fig.4, lanes 1, 2). However, ECF1 does not appear to be a protease-resistant core of SRF for three reasons: (i) ECF1 did not co-migrate with the SRF core (lane 3); (ii) ECF1 disappeared after proteinase K treatment of the ECF1/CArG complex (using the same amount of enzyme as used to generate the resistant core from SRF)(lane 4); (iii) ECF1, unlike the SRF core, was not detectable when pUC plasmid, rather than poly(dI-dC),poly(dI-dC), was used as the non-specific DNA (lane 5).

**Asymmetric recognition of CArG box1**

The interaction of the various binding activities with DNA probes of an inverted repeat of the right-half of CArG box1 (R/R), an inverted repeat of the left-half (L/L), or the wild-type motif (L/R), was assessed by the EMSA (Fig.5). There was a marked difference in the profile of binding activities detectable in the unfractionated embryo extract on the different probes (lanes 1–3). This was investigated further using the fractions enriched for the various CArG box1 binding activities. SRF bound the three probes with the following preference: R/R > L/R > L/L (lanes 4–6). This is consistent with previous work which indicated that *Xenopus* SRF interacts more closely with the right half of CArG box1 than the left half (10), and that human SRF bound more strongly to an inverted repeat of an SRE half-site than to the wild-type SRE (18). ECF4 bound to each of the three probes with similar affinity, which is consistent with its assignation as a DNA-ends binding protein.

In contrast to SRF, ECF1 bound the three probes with the following preference: L/R > R/R >> L/L (lanes 7, 11, 12). This binding preference was confirmed by using oligos of the three sites as competitors for complex formation on L/L (lanes 7–10). In vitro binding of ECF1 was detectable on L/L only under conditions that strongly favour an interaction. If there were other CArG box1 binding activities present (lane 3) or if there was other CArG box DNA (lane 10), then no binding to L/L was detectable. Both ECF2 and ECF3 also bound the three probes with the same order of preference: L/R > R/R >> L/L (lanes 13–18). Therefore, each of ECF1, 2 and 3 showed a preference for binding the right half of CArG box1. This is consistent with a binding assay using a CArG box1 variant with two point mutations in the left half to which SRF cannot bind, but each of ECF1, 2 and 3 can (not shown).

**Relative affinities for the CArG box and the SRE**

To determine whether any of the CArG box1 binding activities discriminated between CArG box1 and the *Xenopus* c-fos SRE, probes of each sequence were incubated with either the unfractonated neurula extract or the fractions enriched for the different factors. In the unfractionated extract the profile of binding activities on CArG box1 is strikingly different from that on the SRE. On CArG box1, ECF1 was the most prominent binding activity, whereas on the SRE, ECF2 and SRF were the most prominent (Fig.6A and B, lanes 1–5). In the assays with enriched fractions, the amount of complex formed reflected the affinity of each binding activity for the two DNAs, because the amounts of extract and probe were the same in each case (Fig.6A and B, lanes 6–19). The most marked difference was with ECF1,
which bound more strongly to the CArG box than to the SRE. ECF3 also bound more strongly to the CArG box, whereas binding of ECF2 and ECF4 was similar to both motifs. SRF bound a few-fold more strongly to the SRE than the CArG box, confirming previous work with embryo extracts and cloned Xenopus SRF (10, 11). These observations were strongly supported by competing complex formation with each protein fraction on each probe with either CArG box or SRE oligo. On each probe CArG box oligo was the stronger competitor of ECF1 and ECF3, SRE was the stronger competitor of SRF, and CArG box and SRE were similarly effective competitors of ECF2 and ECF4.

The distribution of CArG box binding activities during embryonic muscle development

The CArG box binding activities present in embryonic somites and heart, the two tissues that express the cardiac actin gene during development, and also those in the adult heart, were
compared with those in the whole embryo extract. In an extract from dissected somites, which are almost entirely myotome in *Xenopus*, ECF1 and SRF were the major binding activities detected, with substantially more of the former (Fig. 7A, lane 1). Small amounts of ECF2 and ECF3 were detectable, especially in the absence of MgCl₂ (not shown), and there was little detectable ECF4. With embryonic heart extract, ECF1 was the binding activity most readily detected (lane 2). Small amounts of ECF2 and SRF were also present. There was a little ECF3, again clearer in the absence of MgCl₂ (not shown), and no detectable ECF4. The most readily detectable CArG box1 binding activity in these two embryonic muscle tissues was ECF1, which suggests that it is likely to be interacting with the CArG box when the cardiac actin gene is expressed in embryonic muscle.

In unfractionated adult heart extract ECF1 was again the most readily detectable binding activity (Fig. 7A, lanes 4–6). Ammonium sulphate precipitation made the binding activity clearer than in the unprecipitated protein (lanes 7, 8). These factors were less readily detectable than in the neuroblast extract. ECF3 was not detected in the adult heart extract, either with or without MgCl₂, or after fractionation (not shown).

A series of experiments was undertaken to determine whether changes in the concentration of any of the CArG box1 binding activities correlates with cardiac actin gene expression. Figure 7B shows an EMSA with a series of extracts from successive stages of early development (lanes 1–7). No dramatic differences in the relative levels of ECF 1, 2, 4 and SRF were detected during this period when the cardiac actin gene is activated. There was also no change in ECF3, detectable in these extracts by omitting MgCl₂ (not shown).

The regional localisation of the CArG box1 binding activities was investigated at three stages of development: before activation of cardiac actin expression (stage 9); when expression is established and somites are first developing (stage 18), and when axial muscle development is substantial (stage 32). Stage 9 blastulae were dissected into animal, equatorial and vegetal pieces: the embryonic axial muscle normally develops from the equatorial region. Somites and the adjoining neurectoderm, which does not express cardiac actin, were each dissected from stage 18 neurulae. Stage 32 tadpoles were dissected into belly pieces that contained no striated muscle, and axial/tail pieces where the major tissue was axial muscle. Extracts from each dissected piece were tested in the EMSA (Fig. 7B, lanes 8–14). There was no preferential localisation of SRF, ECF1 and ECF2, nor of ECF3 in the absence of MgCl₂ (not shown), although compared to somite extract, the neurectoderm extract contained somewhat less SRF with respect to ECF1. The only major difference was that ECF4 was undetectable in the axial/tail extracts, but prominent in belly extracts. There was, therefore, no evidence for a CArG box1 binding activity in the early *Xenopus* embryo whose presence was positively correlated with expression of cardiac actin and the development of embryonic muscle.

**DISCUSSION**

Four specific CArG box1 binding activities were identified in *Xenopus* embryonic tissues that express the cardiac actin gene. The activity (or activities) that bound the CArG box was influenced both by the concentration of other CArG box binding activities and by the sequence of the site; a very different profile of activities was detectable on CArG box1 than on the SRE. These results suggest subtle transcriptional regulation at CArG boxes in all cell types in *Xenopus* embryos that is determined by the relative activities of the different CArG box binding proteins present and the sequence of the target site. Although no CArG box1 binding activity present only in embryonic muscle was identified, my results indicate that ECF1 and SRF are the factors most likely to be bound to CArG box1 in developing muscle. Since CArG box1 is required for expression of the cardiac actin gene, these binding activities are likely to have a role in the transcription of this gene in embryonic muscle cells.

ECF1 was the most prominent CArG box1 binding activity in extracts from somites, and embryonic and adult heart. It bound the CArG box in preference to an SRE, and bound preferentially to the right half of CArG box1. ECF1 may be related to MAPF1 and 2, which interact with CArG box1 of the chicken skeletal actin promoter and bind to one half of the CArG box, but have not yet been characterised in detail (19, 14). MAPF1 and 2 are distinguishable by size, and the mobility of ECF1 in the EMSA is similar to that of MAPF2. Whilst it has been suggested that MAPF2 is a protease-resistant core of SRF (16), the results in this paper indicate that ECF1 is not this SRF core. The properties of MAPF1 are similar to those of the c-fos SRE binding activity p62 (20), although MAPF1 binds poorly to the SRE (14). Many of the properties of ECF3 suggest that it is related to MAPF1 p62: its preferential binding to the right half of CArG box1, and to a CArG box over an SRE, and its mobility in the EMSA. Moreover, it is eluted from a cation exchange column at a lower salt concentration than SRF, and is detectable using poly(dI-dC),poly(dI-dC), but not plasmid as non-specific DNA (20).

SRF was prominent in the somite extract and also present, but at a lower level, in both embryonic and adult heart extracts. This is consistent with a role for it in muscle gene expression. Such a role is suggested by preliminary experiments that have identified a CArG box variant with two point mutations that is bound by ECF1–4, but not by SRF. Whilst the wild-type CArG box can replace all four CArG boxes in a cardiac actin promoter directing embryonic expression (9), this variant CArG box cannot (M.V.T., T.Mohun and J.Gurdon—unpublished). However, like ECF1, SRF was detectable at stages of early *Xenopus* development and in cell types where the cardiac actin gene is not expressed. Moreover, there is no evidence for a tissue-specific variant of SRF that might go undetected in this type of analysis (11). Unlike ECF1, SRF showed a preference for the SRE over the CArG box in binding assays.

ECF2 was identified as a high abundance, relatively low affinity CArG box1/SRE binding activity. A function might be to occupy CArG/SRE sequences in the absence of other factors. In general there may be abundant lower affinity, site-specific factors that keep promoter sequences available for binding of other factors. This idea is supported by the hierarchy of binding activities observed: ECF 2 and 3 binding was greatly diminished in the presence of other CArG box1 binding activities. Competition between binding activities at a single promoter motif is also observed on the CCAAT-box in the sea urchin histone H2B-1 and mouse a-globin promoters (21, 22).

ECF4 was shown to be a DNA ends binding protein. HeLa NFIV is one of the few proteins shown to require DNA ends for binding (23). NFIV and ECF4 share the property of not distinguishing between 5', 3' and blunt ends, but NFIV binds single-strand M13 efficiently, whereas ECF4 does not (not shown). Although an ends-binding protein has not been described in *Xenopus* embryos, they do contain an activity that joins 5'.
REFERENCES

I also gratefully acknowledge the financial support of the Cancer Research Campaign.

ACKNOWLEDGEMENTS

I wish to thank John Gurdon, in whose lab this work was carried out, for the somite and neurectoderm dissections, Tim Mohun for the CARG box 1 L/L and R/R plasmids, Jam Tata for help with the adult heart extract preparation, Micaela Fairman for advice on chromatography, Colin Sharpe and Janet Heasman for the vimentin oligos, and Tony Kouzarides, Nick Hopwood, Jeremy Rashbass, Colin Sharpe for comments on the manuscript. I also gratefully acknowledge the financial support of the Cancer Research Campaign.

REFERENCES