Development 112, 351–364 (1991)
Printed in Great Britain © The Company of Biologists Limited 1991

Gene regulatory factors of the sea urchin embryo

II. Two dissimilar proteins, P3A1 and P3A2, bind to the same target sites that are required for early territorial gene expression

CHRISTER HOOG*, FRANK J. CALZONE†, ANN E. CUTTING, ROY J. BRITTEN and ERIC H. DAVIDSON

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

* Present address Department of Molecular Genetics, Karolinska Institutet, S-10401, Stockholm, Sweden
† Present address Department of Developmental and Cell Biology, University of California, Irvine, CA 92717, USA

Summary

Previous work demonstrated that a negative regulatory interaction mediated by factor(s) termed 'P3A' is required for correct territory-specific gene expression in the sea urchin embryo. A probe derived from a P3A target site in the skeletogenic *SM50* gene of *Strongylocentrotus purpuratus* was used to isolate a cDNA clone coding for a factor that binds specifically to this site. This factor, called P3A1, contains two sequence elements that belong to the Zn finger class of DNA-binding motifs, and in these regions is most closely similar to the *Drosophila hunchback* factor. The P3A1 factor also binds to a similar target sequence in a second gene, *CyIIIa*, expressed in embryonic aboral ectoderm. Another sea urchin embryo protein factor, P3A2, has been isolated by affinity chromatography and cloned, as described in Calzone *et al. Development* 112, 335–350 (1991). P3A2 footprints the same target sites in the *SM50* and *CyIIIa* genes as does P3A1, but lacks the Zn finger sequence motifs and in amino acid sequence is almost entirely dissimilar to P3A1. A deletion analysis of P3A2 delimited the DNA-binding region, revealing that five specific amino acids in the first P3A1 finger region and four in the second P3A1 finger region are also present in equivalent positions in P3A2. The P3A1 and P3A2 factors could function as regulatory antagonists, having evolved similar target specificities from dissimilar DNA-binding domains.

Key words: DNA-binding protein, embryogenesis

Introduction

As described in the preceding paper (Calzone *et al.* 1991), DNA–protein interactions at *cis*-regulatory target sites called the P3A sites are required for spatial regulation of the *CyIIIa* actin gene. These interactions cause repression of the *CyIIIa* gene in territories other than the aboral ectoderm, to which *CyIIIa* expression is normally confined (Hough-Evans *et al.* 1990). Calzone *et al.* (1991) purified by affinity chromatography a regulatory factor, P3A2, which binds tightly and specifically to the P3A target sites. In this paper, we describe a second DNA-binding factor, P3A1, which recognizes the same target sites. P3A1 was isolated by direct ligand screening of a cDNA library made from cleavage stage embryo mRNA. On cloning and sequencing P3A1, we discovered, unexpectedly, that the early embryo in fact contains two very different DNA-binding proteins with high target specificity for the same P3A target site sequences. P3A1 is of the finger protein family, most closely related to the *Drosophila* Zn finger factor that is the product of the *hunchback* gene. P3A2 bears almost no resemblance to P3A1 in sequence, except for two short, adjacent sequence motifs that lie within the DNA-binding domain of P3A2 and also appear within the finger regions of P3A1. Though the P3A2 sequence does not possess the canonical features of known classes of finger proteins, these similarities in sequence may indicate the locus of its target site specificity.

There are several curious and interesting features of the P3A regulatory system. First, it appears that wholly dissimilar proteins have evolved to recognize the same regulatory target sequences. This in turn implies a functional significance, e.g. competitive interaction of P3A1 and P3A2 at these sites. In addition, the P3A factors must interact with other regulatory elements to account for the fact that they are required for specification of gene expression in nonoverlapping and functionally distinct lineage elements of the early

embryo. That is, in the aboral ectoderm the *CyIIIa* gene is on and the *SM50* gene is off, while the converse is true in skeletogenic mesenchyme, and yet both genes may require P3A interactions to prevent ectopic expression.

Materials and methods

Construction of cDNA libraries

Total RNA was isolated from unfertilized eggs or embryos, as described by Posakony *et al.* (1983) and Lee *et al.* (1986). Poly(A)$^+$ RNA was partially purified from the total RNA, as described by Posakony *et al.* (1983). 5 μg of poly(A)$^+$ RNA prepared from embryos of different stages were used to synthesize cDNA. The BRL cDNA synthesis kit was used with oligo(dT) primers. The cDNA was prepared for cloning as described by Watson and Jackson (1985) and ligated into λgt11 (Promega Biotec) which had been digested with *Eco*RI and dephosphorylated, or into λZAP DNA (Stratagene) Ligated DNA was packaged *in vitro* using a commercial extract ('Gigapack Gold', Stratagene). The cDNA library made in λgt11 from poly(A)$^+$ RNA prepared from 4 h (3rd–4th cleavage) embryos contained 5.2×10^6 primary plaques, and was not amplified before use (average insert size was 2 3 kb). The second cDNA library was constructed in λZAP from poly(A)$^+$ RNA prepared from 14 h (9th cleavage) embryos. This library contained $8\ 7\times10^5$ primary plaques and was amplified to a titer of 1×10^{11} pfu ml^{-1}.

Isolation of P3A1 cDNA clones

An aliquot of the 4 h embryo λgt11 cDNA library was probed by the direct ligand filter detection method of Vinson *et al* (1987) $1\ 3\times10^6$ plaques were screened at a density of 2×10^4 plaques per filter, after lytic infection of *E. coli* strain Y1090r$^-$ (Young and Davis, 1983) The probes used in screening the filters were prepared by nick translation (Rigby *et al* 1987) of size-selected, double-stranded concatenated oligonucleotides 25/26 (see Calzone *et al.* 1991 for sequence) using α-^{32}P-dATP and α-^{32}P-dCTP (800 Ci mmol^{-1}), according to Vinson *et al.* (1988) Typical labeling reactions yielded probe with a specific activity of about 5×10^8 cts min^{-1} μg^{-1} In the initial screen, the filters were exposed to 2×10^6 cts min^{-1} ml^{-1} of radiolabeled P3A probe DNA. 45 positive clones were identified and one λ38 was found to bind the P3A DNA probe in the following rescreens. Lambda 38 phage DNA was purified in large scale by standard methods A 1.5 kb cDNA insert was subcloned into the *Eco*RI site of pBluescript SK$^-$ (Stratagene) creating clone p38. A fragment of p38 was nick translated and used to screen 6×10^5 ZAP plaques from the 14 h library. 112 positives were found, 50 of which were plaque purified and converted into plasmids according to the manufacturer's protocol (Stratagene). The longest clone, p43, was isolated for further analysis. An oligonucleotide was synthesized complementary to the 5′ region of p43 (5′-AATGCAGAAACATCTCCAGGGTCATATAGATCGAG, position 127–161 in Fig. 4) and used to rescreen the same ZAP library. The longest clone isolated in this screen, p26, was selected. Another clone, p13, extended furthest in the 5′ direction, and was used to obtain the 5′ terminal sequence as indicated in Fig 4.

The mRNA coding for the P3A1 factor was later shown by single strand probe excess titration to belong to the rare class of maternal and early embryo mRNAs (Cutting *et al.* 1990). At 4 h postfertilization (the stage from which λgt11 cDNA library was made), there are only about 600 copies of this mRNA per embryo Since there is about 30 pg of total mRNA in the egg, of average length ~2.5 kb (Davidson, 1986, p.71), P3A1 message constitutes only 3×10^{-5} of the mRNA, and we should have expected on the average 37 P3A1 clones to have been present in the 1.3×10^6 clones initially screened. Of these about six should have been in the correct orientation and reading frame. That we found only one such clone could be due either to failure of many clones to include the 5′ half of the mRNA, which codes for the necesssary DNA-binding regions of the protein (see below); to insufficiently vigorous protein expression, or to a bias against cloning of this particular sequence. However, screening by DNA hybridization of other various embryo λ libraries with P3A1 DNA probes demonstrate that if anything this sequence is slightly over-represented. A subsequent direct comparison with a clone encoding a different DNA-binding protein isolated from the same library, of the number of clones that tested positive by direct DNA ligand screening with those positive by nucleic acid homology, yielded a ratio of 1:46 Taking into account the requirement for correct orientation and reading frame, the ratio of clones expected to be recovered by direct DNA ligand screening to clones actually detected was 7.7. We were thus fortunate to have recovered the single initial P3A1 isolate. The moral of this excursion is that for this method to be routinely useful in cases where the factor is encoded by an mRNA of such low prevalence, very large numbers of cDNA clones must be screened.

The P3A1 factor is present in the nuclear extracts from which P3A2 was isolated by affinity chromatography, according to reaction of the anti-P3A1 antibody. However, P3A1 did not bind to this column, perhaps because in the nuclear extracts the P3A1 binding site is occluded by another protein bound to it. Thus it was fortuitous that we applied *both* direct ligand screening and affinity chromatography to the problem of isolating the P3A factor(s)

Sequence of P3A1 cDNA clones

Clones p38, p43, p26, p13 and some additional overlapping cDNA clones were sequenced on both strands using subclones generated by an exonucleaseIII/S1 kit (Promega Biotec), using XL1 blue cells as host (Henikoff, 1984). Clones with deletions of the appropriate size were identified by agarose gel electrophoresis and used for dideoxy sequencing (Sanger *et al.* 1977) using Sequenase (USB) after alkali denaturation of the DNA (Mierendorf and Pfeffer, 1987). Gaps or ambiguities in the DNA sequence were resolved by using internal primers. The sequence data were compiled using the IBI program. The 5′ end of the insert in the original isolate, λ38, was sequenced using a λgt11 primer (New England Biolabs).

Preparation of bacterial extract containing P3A1 protein

An *Xho*I–*Bam*HI fragment was released from p43 (position 52–2473) and end filled. A *Bam*HI linker was used to create the correct frame for ligation into the expression vector pET3c (Studier and Moffat, 1986) generating clone pET3c-P3A1. A second fragment released by digestion with *Hin*cII (position 332–846) was ligated in a similar way into pET3a generating clone pET3a-P3A1. BL21(DE3)pLysS cells were infected with these constructs or, with pET3c vector alone, grown to mid-log phase, induced with IPTG for 2 h and harvested. For electrophoretic sizing and subsequent ligand gel blots, cells were lysed in SDS sample buffer and urea was added to a final concentration of 4 M. The samples were then heated for 5 min at 90°C and the solubilized protein was loaded onto SDS–polyacrylamide gels (Laemmli, 1970). For DNAase I footprint and gel shift analyses, the cells were harvested and resuspended in buffer C (20 mM Hepes

[pH.7.4], 40mM KC1, 0.1 HIM EDTA, lmM DTT 20% glycerol) with the addition of lysozyme [1mgmF^{1}]). Cells were incubated on ice for 1 h and then briefly sonicated. The insoluble material was spun out and the supernatant was stored at -70°C For DNAase I footprinting and gel retardation experiments bacterial extracts of P3A1 expression clones were prepared as described by Calzone *et al.* (1991). The recombinant protein produced by pET3c-P3Al was purified by affinity chromatogTaphy using a concatenated DNA-binding site derived from the *Cyllla* gene (ohgonucleotides 11/12 of *Th6z6 et al* 1990).

Ligand-specific protein gel blots

Bacterial proteins produced by the pET expression construct were resolved on SDS-polyacrylamide gels and electrophoretically transferred to nitrocellulose paper, using standard blotting conditions. The filters were then probed according to the protocol of Miskimins *et al* (1985) using the same buffer system as in the ligand clone screening method of Vinson *et al* (1988). The four double-stranded DNA probes used represented binding sites found in the upstream region of the *Cyllla* gene (Th6z6 *et al* 1990) and were constructed from complementary ohgonucleotides with offset termini to generate 'sticky ends' that would allow concatenation. The sequences of some of these probes are given in Table 1 of Th6z6 *et al.* (1990). Those utilized in this study were ohgonucleotides 23/24 (the PI site); 5/6 (the P6 site); 19/20 (the P7I site), all given by Th6z6 *et al.* (1990) Ohgonucleotides 25/26 are given in the accompanying paper (Calzone *et al* 1991), and the relevant native P3A site of the *SM50* gene is shown in Fig. 3C of this paper (see Table 2 of accompanying paper for other P3A target sites in these genes)

Gel shift reactions and DNAase I footprints

The quantitative gel shift reactions used for estimation of $K_{\wedge}$, were carried out as described by Calzone *et al.* (1988) and in legend to Fig. 2. DNAase I footprints were carried out by standard methods as described in detail by Calzone *et al.* (1991).

Locating the DNA-binding region of P3A2

A Bluescnpt cDNA clone containing the complete coding sequence of the P3A2 protein (Calzone *et al.* 1991) was used as the starting material. The 3' series of deletions was generated from a slightly shortened version of this clone, in which 30 nt were removed from the 5' end by exonuclease digestion (to eliminate a foldback that caused premature stops in transcription). The shortened clone, pP3A2A, was linearized by digestion with different restriction enzymes, at the following positions in the sequence (Calzone *et al.* 1991):

P3A2A	1	1335	*Espl*
	2	1070	*Sphl*
	3	1014	*Hpal*
	4	854	*Stul*
	5	745	*Xmnl*
	6	695	*Ncol*
	1	623	*Seal*
	8	571	*Hindi*
	9	480	*Styl*

To construct 5' deletions PCR was used to generate copies of selected regions of the P3A2 coding sequence (Sarkar and Sommer, 1989). The template was a *Hindlll-Htndlll* fragment from the P3A2 clone [position 1-2883 in the P3A2 DNA sequence (Calzone *et al.* 1991)]. The ohgonucleotides that were used contained, in addition to a sequence complementary to the desired locations in the P3A2 cDNA, a 23-base T7 promoter sequence followed by a 9-base translation initiation signal that fused a methionine in frame with the P3A2 protein sequence. These oligonucleotide primers were as follows (the digits in the name for each oligonucleotide indicate the start of the region of the cDNA clone to which the oligonucleotide is complementary):

561 5'-TAATAO^CTCACTATAC«WCACCGCCATCGCTACATCGGTCAACAAGATG

505 5'-TAATA«^CTCACTATACKX^GAC«KXATGCCGTCTAATGAGAATTCAGAG

453

351 5'-TAATAC<^CTCACTATAGGGAGACC«X^TGCGGCAGCAA<XNTRRCGTTCTA

153 5'-TAATACX^CTT:ACTATAGGGACACCGCCATGCCAAGTATGACGGATGATGTC

1163 5'-CAATGTGGTTACTGTTTGCCC (includes no T7 promoter sequence)

PCR was performed with oligonucleotide 1163 (3'-end), and one of five additional ohgonucleotides (5'-end) using a Perkin-Elmer 'Gene-Amp' kit according to the manufacturer's recommendation in a 100pi\ reaction. After lO min at 94 °C, 35 cycles of PCR were performed with the Perkin-Elmer Cetus automated thermal cycler. *1 jA* (—0.5 fig) of the amplified DNA were added to 43 *jA* of RNA transcription mixture and was transcribed using T7 RNA polymerase (Stratagene)

The linear DNA obtained from the PCR reactions was used as template for T7 RNA polymerase to transcribe uncapped RNA suitable for *in vitro* translation (Promega Biotec). The RNA templates (about 2fig) were added to a rabbit reticulocyte lysate (Promega Biotec) containing [^{35}S]methionine according to the manufacturer's specifications, and the protein products were confirmed by analysis of the reaction products on SDS-polyacrylamide gels.

Results

P3A1 cDNA clones

Target binding sites for the P3A factors in the *SM50* and *Cyllla* genes show a core sequence element, $^{C}/TX\,^{C}/TGCGC^{A}/_{T}$ (Calzone *et al.* 1988, and preceding paper; Th6z6 *et al.* 1990; Thiebaud *et al.* 1990). The P3A1 factor described herein was isolated by screening an expression cDNA clone library with concatenated oligonucleotides, each monomer of which contains two of these core sequence elements. This probe represents a naturally occurring, high affinity double P3A site located at position -128 to -102 of the *SM50* gene. The exact sequence of this site is shown in Table 2 of the accompanying paper of Calzone *et al.* (1991). Oligonucleotides 25/26, which were used to produce the probe, consist of precisely this sequence with a few additional flanking nucleotides from the *SM50* gene, as also given in the accompanying paper. The concatenated probe was labeled and used to screen a set of early embryo cDNA libraries by the method of Vinson *et al.* (1988). The initial P3A1 isolate was obtained from a Agtll library containing inserts of 4h embryo cDNA. One reproducibly positive isolate, called A38, was obtained from 1.3x10 plaques produced by a portion of the unamplified packaging mixture. After purification the insert of this recombinant, which was about 1.5 kb in length, was converted into a subclone (p38) and a fragment of p38 was used to screen a second cDNA library, produced from 14 h embryo cDNA, and constructed in AZAP. The longest of the positive clones contained a 2.6 kb insert (p43), and this clone was used for much of the following study.

Interactions in vitro *between P3A1 protein and DNA target sites*

An insert from clone p43 containing sequence coding for amino acids 9–387 (the -COOH terminus) was transferred to the pET3c expression vector (Studier and Moffat, 1986). This clone is denoted pET3c-P3A1. After IPTG induction the proteins were extracted (see Materials and methods, and legend to Fig. 1), and loaded onto an SDS–polyacrylamide gel. Following electrophoresis they were transferred to nitrocellulose and reacted with the probe containing the P3A site, and with three other probes that bind tightly to *other* sea urchin embryo nuclear factors involved in the regulation of the *CyIIIa* gene, *viz* the P1, P6 and P7I sites (Calzone *et al.* 1988; Thézé *et al.* 1990). Fig. 1 shows that only the P3A probe is bound by the bacterial products of the expression clone pET3c-P3A1. This result is dependent on IPTG induction, and is not obtained with extracts from bacteria containing the vector alone (data not shown). The apparent relative molecular mass of the active fusion protein is about 60×10^3 (in this vector the fusion product includes only 12 amino acids encoded by the host sequence).

The fusion protein produced by the pET3c-P3A1 expression construct was recovered and used for quantitative gel shift reactions in order to measure its relative affinity for the *SM50* and *CyIIIa* target sites. The protein product of pET3c-P3A1 was purified using a site-specific DNA affinity column (Calzone *et al.* 1991). We utilized quantitative gel shift competition series (cf. Calzone *et al.* 1988) to determine values for the equilibrium dissociation constant (K_D) by least squares analysis. Representative gel shift competitions are shown in Fig. 2A, for reaction of the recombinant P3A1 protein with the *CyIIIa* target sequence; and in Fig. 2B for its reaction with the *SM50* target sequence. Beneath each panel the data are plotted in a form which allows K_D to be calculated directly from the slope (which, as given in the legend, is $-K_D^{-1}$). Fig. 2 shows that the bacterial protein produced by pET3c-P3A1 indeed reacts specifically with both the *SM50* P3A and the *CyIIIa* P3A sites. The K_D values obtained in these particular reactions are about the same, i.e. 6.0×10^{-8} M for the *CyIIIa* target fragment and 3.5×10^{-8} M for the *SM50* target fragment. These values are not significantly different, and experiments in which the *SM50* oligonucleotide competitor was used with the *CyIIIa* probe, show that the recombinant P3A1 protein partitions equally between these target sites. It is useful to compare these values with the K_D measurements obtained in similar experiments using the P3A2 factor, and presented in the accompanying paper of Calzone *et al.* In contrast to the *equivalent* binding affinities for the *SM50* and *CyIIIa* sites displayed by P3A1, P3A2 binds to the *SM50* probe about 50-fold better than it does to a probe containing a single *CyIIIa* site, though it binds to the *CyIIIa* site with about the same affinity as does P3A1. Thus the apparent K_D values extracted from the experiments of Calzone *et al.* were $K_D=7.4\times10^{-8}$ M for the reaction of P3A2 with the *CyIIIa* site, but $K_D=1.7\times10^{-9}$ M for its reaction with the *SM50* probe. However, there are two reasons why interpretation of this comparison in itself is not straightforward. (*i*) The *CyIIIa* fragment used for the gel shifts as probe contains

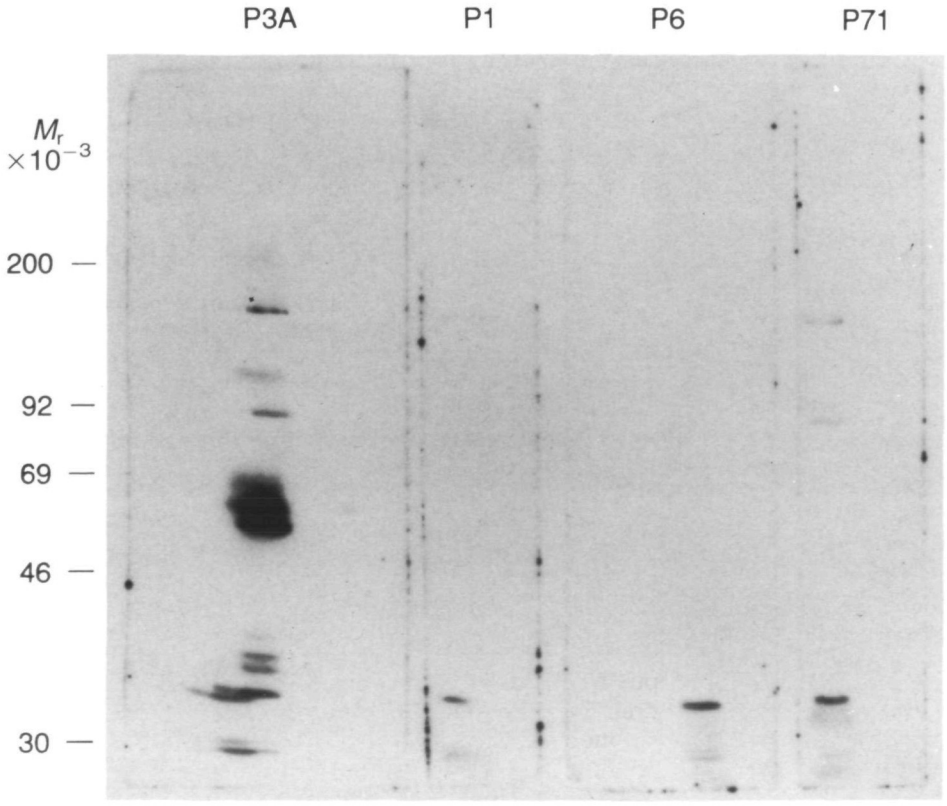

Fig. 1. Ligand gel blot of fusion proteins generated from pET3c-P3A1 expression clone. This clone lacks only the initial 12 amino acids of the complete open reading frame (see Materials and methods). After induction with IPTG for 2 h the cells were collected and resuspended in SDS–lysis buffer with 4 M urea An identical amount of protein was loaded in four separate wells of an SDS–PAGE gel. Following electrophoresis the proteins were transferred electrophoretically to a nitrocellulose filter, which was cut into four strips. Each strip was reacted with a unique double-stranded concatenated DNA probe; P3A (A), P1 (B), P6 (C), and P7I (D) (for probes see Materials and methods). Molecular weight standards (Amersham) are shown on the left. Minor bands representing nonspecific DNA-binding proteins can be observed between $30–40\times10^3$ and at $>90\times10^3$. Two probable proteolytic P3A2 derivatives migrate at $\sim38\times10^3$.

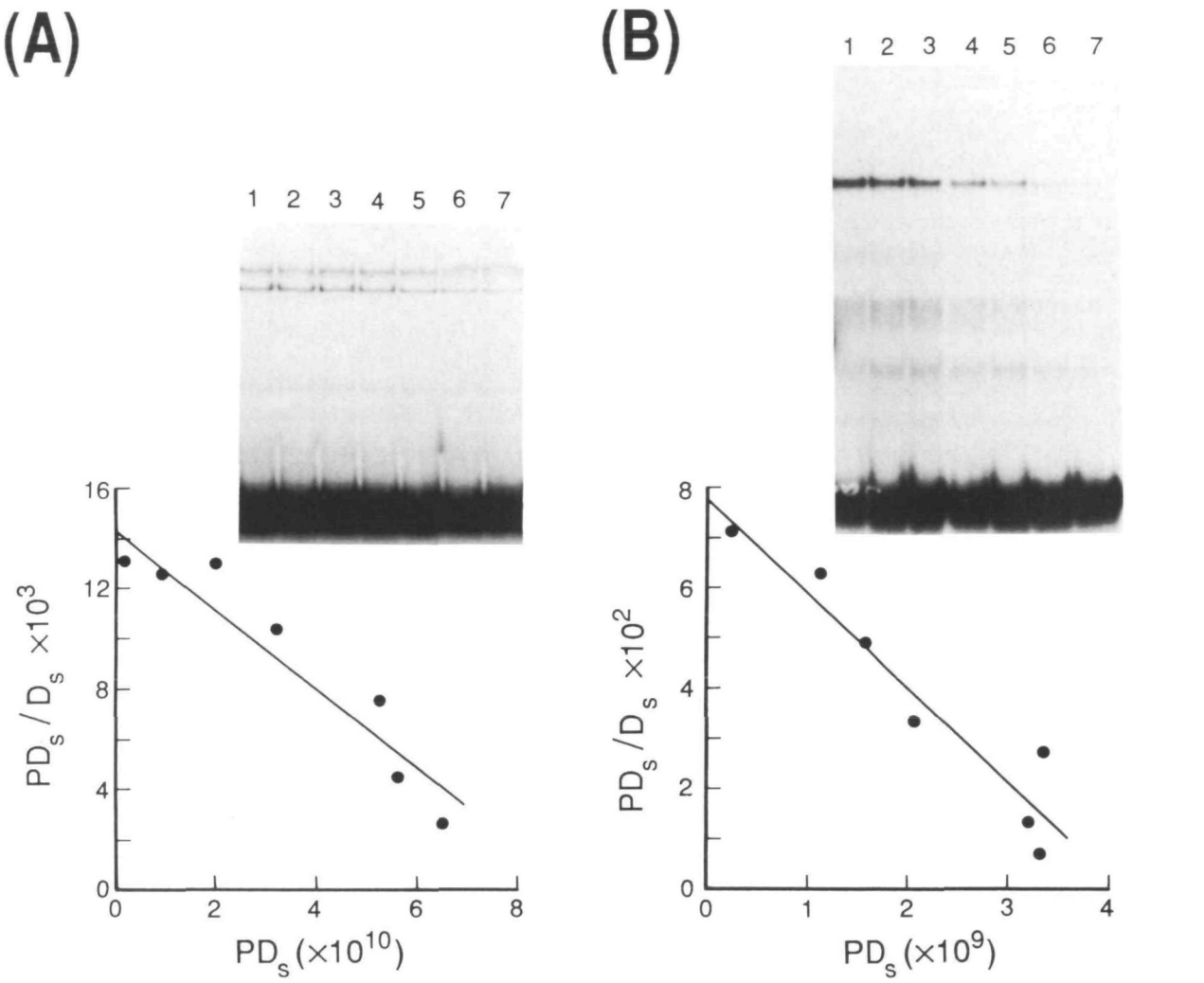

Fig. 2. Specific gel shift reactions mediated by recombinant P3A1 protein. The fusion protein encoded by pET3c-P3A1 was recovered and purified as described in Materials and methods, and utilized for gel shift competition series with (A) *CyIIIa* and (B) *SM50* target sites as probes. In A, the probe was the 86 nt Z fragment of the *CyIIIa* regulatory domain, which contains a single P3A target site (see map in Thézé *et al.* 1990), and the competitor was constructed by concatenating the 25 nt complementary oligonucleotides 11/12 (see Materials and methods; the sequence included in oligonucleotides 11/12 is indicated in Fig. 4C of this paper). Lanes 2–7 contain increasing amounts of specific competitor. Probe was present at 2.5×10^{-10} M. The significance of the two complexes detected when P3A1 was reacted with the *CyIIIa* Z probe is unknown and, as seen in B, only a single complex was observed when the probe contained the tandem P3A target sites The graphs in the lower portion of the figure show determinations of the equilibrium dissociation constant (K_D) from this reaction, determined by least squares analysis using the function

$$\frac{PD_s}{D_s} = -\frac{1}{K_D}PDs + \frac{1}{K_D}P_0,$$

where PD_s/D_s is the ratio of complex to free specific probe (measured from the gel shift reaction), and P_0 is the concentration of active factor in the reaction K_D was 6.0×10^{-8} M for this reaction, P_0 was about 9×10^{-10} M, and poly(dI)·poly(dC) was present at about 8.0×10^{-4} M. In B, the *SM50* target sequence probe was a monomeric oligonucleotide containing the tandem P3A sites, and the competitor was concatenated oligonucleotide 25/26. Lane assignments are as in A. The K_D value extracted from this experiment was $3\,5\times10^{-8}$ M, P_0 was about 4.1×10^{-9} M and poly(dI)·poly(dC) was present at 8.0×10^{-5} M. Probe was present at about 2.0×10^{-9} M. K_D for the nonspecific reaction of P3A1 with poly(dI)·poly(dC) should be about $1\,8\times10^{-3}$ M (cf. Calzone *et al* 1988). Thus, in A ~40% of the protein is expected to be bound to the nonspecific DNA, and in B less than 5% of the protein was in nonspecific complex.

a single target site while the *SM50* probe fragment contains two nearby but nonidentical sites (see below), and (*ii*) the P3A1 protein was a recombinant product of a bacterial expression system that was affinity purified, while the P3A2 protein utilized for these studies was purified directly by affinity chromatography from sea urchin embryo nuclear extract, and this factor is subject to modifications that may affect its DNA-binding activity. Nevertheless, the low molar values for K_D measured in Fig. 2 of this paper, which fall well within

the range reported for known DNA-binding regulatory proteins, demonstrate that P3A1 is indeed a specific DNA-binding protein that recognizes the P3A target sites of both *SM50* and *CyIIIa* genes. Furthermore, although proof of the point must await direct purification of the native P3A1 protein, the comparison with P3A2 raises the possibility that P3A1 does not discriminate between the *CyIIIa* and *SM50* sites, while P3A2 does. The same distinction obtains in the case of closely spaced P3A sites in a third gene, *Spec*I. Here again P3A2 binds these sites relatively tightly, but P3A1 responds to them the same as it does to the single *CyIIIa* site utilized for the experiment of Fig. 2A (data not shown). Were this difference to obtain *in vivo*. differential functional significance of the two factors would be implied, with respect to interactions with the *CyIIIa*, *SM50* and other genes that bear regulatory P3A target sites.

The footprints and hypersensitive sites shown in Fig. 3A,B demonstrate explicitly the sequences engaged by the DNA-binding domains of the P3A1 factor in the *SM50* and *CyIIIa* target sites. For these experiments, a second P3A1 expression clone was used (pET3a-P3A1D), which contains a shorter insert coding for amino acids 104–272. The protein encoded by this construct includes the complete DNA-binding domain of the P3A1 protein. The truncated pET3a-P3A1Δ protein was employed for the experiments of Fig. 3 because its accumulation in the bacterial expression system and stability in extracts greatly exceeds that of the nearly complete protein encoded by pET3c-P3A1.

In Fig. 3C the target regulatory sequences of the *CyIIIa* and *SM50* genes that are protected by P3A1 are compared with those protected by P3A2. P3A2 footprints are shown in Calzone *et al.* (1991) and in B of Fig. 3. Both proteins clearly protect the core $^{T}/_{C}X$ $^{T}/_{C}GCGC^{A}/_{T}$ elements of these sites, which are marked by boxes in Fig. 3C (their relative orientations are indicated by arrows). Methylation interference experiments (not shown) also indicate that both proteins make contacts with G residues on both strands of the central GCGC element. These data, the footprint studies summarized in Fig. 3C and the oligonucleotide competition results shown in Fig. 2 confirm that P3A1 and P3A2 react *with the same DNA target sites*. On the *CyIIIa* site the DNAase I footprint patterns overlap exactly on the 5′→3′ (top) strand, and within one or two nucleotides at each end on the 3′←5′ (bottom) strand. However, the P3A1 protein binds very differently from the P3A2 protein on the double *SM50* site. P3A1 reacts strongly with only the proximal site, where it covers 13 nucleotides just as on the lower strand of the *CyIIIa* site, while P3A2 equally protects *both* sites (this almost certainly requires two P3A2 molecules). At higher concentrations P3A1 weakly interacts with the upstream *SM50* site as well. This impressive difference between the P3A1 and P3A2 footprints obviously could account for the striking difference in the measured K_D values for the two proteins. Thus the footprints provide a reasonable qualitative interpretation for the observation that P3A2 prefers the *SM50* probe about 50 times more than it does the *CyIIIa* probe, while P3A1 regards them more or less indiscriminately, and binds to both with about the same affinity as P3A2 does to the *CyIIIa* site. However, the P3A1 protein may respond specifically to a different feature of the proximal *SM50* site, *viz* that if considered by itself, this site is symmetrical (vertical arrow). Thus the P3A1 factor positions itself directly over its axis of symmetry, so that both termini of the protected region lie in equivalent positions, within the sequence elements $C^{A}/_{G}CCCC$. This feature of the *SM50* site could be functionally important, since this symmetrical sequence motif, to which P3A1 is sensitive, is not found in the *CyIIIa* site.

P3A1 is a finger protein

Fig. 4 displays the sequence of the P3A1 cDNA and the protein for which it codes. The actual mRNA probably begins a short distance upstream of the 5′ terminus shown. However, it is unlikely that this mRNA has a very large 5′ untranslated leader because, though the clone includes only 25 nucleotides prior to the translation initiation site, the size of the sequenced transcript, 2473 nt, is indistinguishable from the size of the transcript revealed by RNA gel blots (Cutting *et al.* 1990) measured by reference to standards as about 2500±50 nt. Use of the first ATG is supported by the adjacent sequence CÄAÄṪCATGG, which at all except the starred positions is identical with the 10 nt consensus translation start site of Kozak (1987). There follows an 1161 nt open reading frame, a stop codon at position 1186 in the sequence, and a 1287 nt long 3′ untranslated region, which in the recovered clones lacks a poly(A) tail, though the mRNA is indeed polyadenylated in the embryo (Cutting *et al.* 1990). The open reading frame of the P3A1 sequence specifies a protein with a relative molecular mass of 42×10^3 (387 amino acids), although, as noted above, the recombinant expressed protein migrates on SDS–polyacrylamide gels as if its mass were 60×10^3 (Fig. 1, lane 1). Similar overestimates of size have been noted for other transcription factors, including GCN4 (Hope and Struhl, 1986), RAP1 (Shore and Nasmyth, 1987) and HSF (Sorger and Pelham, 1988). Since in the present case the recombinant protein was expressed in bacteria, its anomalous migration is unlikely to be due to significant post-translational modification, and its cause might be sought in its unusual amino acid composition. The P3A1 protein is predicted to have a high content of serine, threonine and proline, together ~29 %, as also noted for some other DNA-binding proteins, e.g. GF1 (Tsai *et al.* 1989), Sp1 (Kadonaga *et al.* 1987) and SRF (Norman *et al.* 1988). The amino acid terminal region of the P3A1 protein (positions 1–53) is strongly acidic, with a net charge of −11 and perhaps could function as a transcriptional activator region (reviewed by Mitchell and Tjian, 1989). A short strongly basic region (7/11 amino acids) appears after position 237, similar to the nuclear localization signals described in other proteins (Dingwall and Laskey, 1986). No significant similarities are detected, comparing the derived P3A1 protein sequence with those in the available sequence data

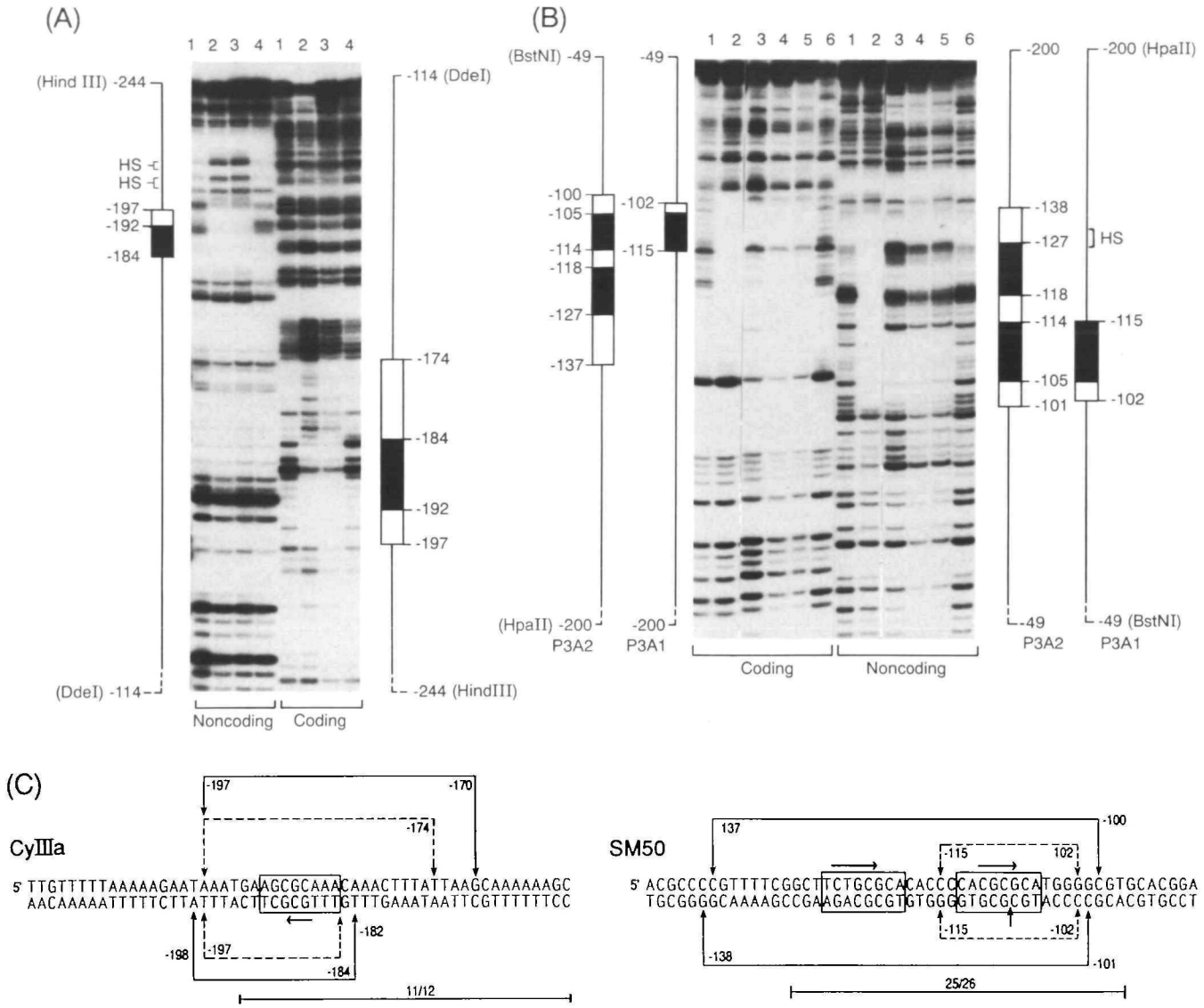

Fig. 3. DNAase I footprint and sequences protected by recombinant P3A1, compared to sequences protected by P3A2. The recombinant P3A1 used in these experiments was the 24×10^3 fragment produced by the pET3a-P3A1 construct (see text) P3A2 was affinity-purified from nuclear extract as described in the accompanying paper. The sequences of the *CyIIIa* and *SM50* regulatory DNA present in each probe are marked in the maps aligned vertically with the reactions. The labeled strand is noted below each set of reactions. The boxes represent the regions of each probe that were protected from nuclease attack by P3A1 or P3A2 as indicated The filled in boxes locate the P3A target site noted in the text. (A) *CyIIIa* footprints. The reactions shown in lanes 1 and 4 for each probe did not contain recombinant protein. In lanes 2 and 3, the probe was incubated with 2 μl and 4 μl of bacterial extract containing recombinant P3A1 before treatment with DNAase I Control extract generated a DNAase I profile identical to that shown in lanes 1 and 2. Marker lanes are not shown. HS indicates sites which became hypersensitive to DNAase I when probe was incubated with P3A1 (darker bands than in no protein control) P3A2 footprints obtained with the same probes are shown in the first paper of this series. (B) *SM50* footprints. Control reactions that did not contain P3A1 or P3A2 are shown in lanes 1 and 6. The footprint obtained with P3A2 is shown in lane 2. The reactions shown in lanes 3, 4, and 5 were incubated with 2, 4, and 6 μl of P3A1 bacterial extract before DNAase I treatment. (C) Sequences protected by P3A1 and P3A2 proteins. Only strong footprints are shown. Solid brackets represent nucleotides protected by P3A2, dashed lines represent nucleotides protected by P3A1 DNA-binding fragment produced by pET3a-P3A1 (or by the nearly complete protein produced by pET3c-P3A1; not shown here). Left, *CyIIIa* target site; right, *SM50* target site. The boxes indicate the canonical core of the P3A target site recognized by both proteins (Thézé *et al* 1990; Calzone *et al* 1988), *viz* 5'-$^{T}/_{C}X^{T}/_{C}GCGC^{A}/_{T}$. Arrows (5'→3') indicate the orientation of this site with respect to the sequences shown (transcription starts downstream to the right). The sequences included in oligonucleotides 11/12 and 25/26 referred to in text are indicated The vertical arrow shows a dyad axis of symmetry in the *SM50* site covered by the P3A1 protein.

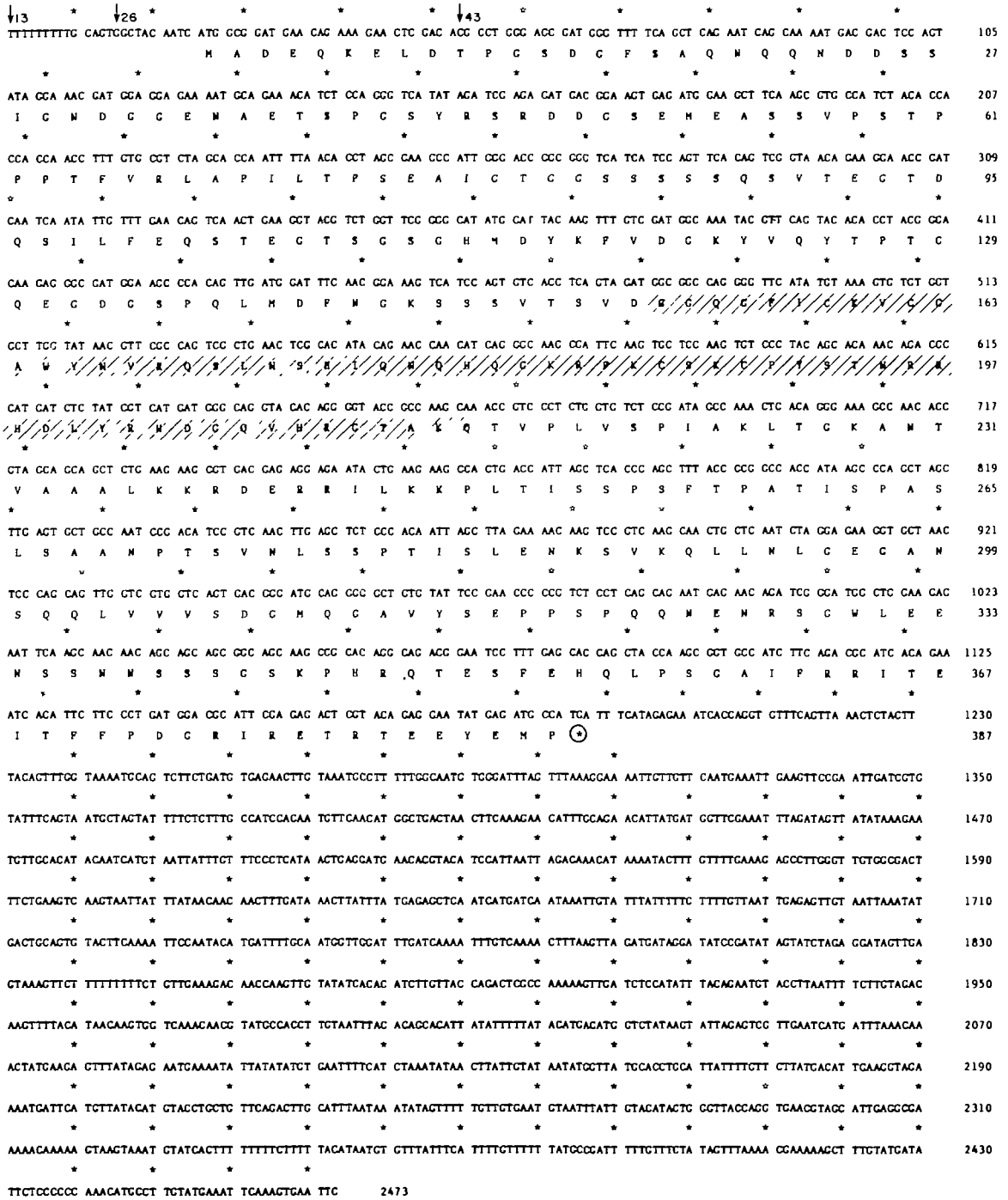

Fig. 4. DNA and deduced amino acid sequence for P3A1. The complete P3A1 cDNA sequence was derived from fragments of cDNA clones as indicated in Materials and methods. 5′ endpoints of inserts of cDNA clones p43, p26 and p13 are indicated by arrows. The two proposed finger sequence elements are cross hatched. The circled asterisk indicates the TGA stop codon. The sequence has been entered in the Genbank/EMBL database, accession no. X57013.

banks, except for two regions that are obviously similar to the Zn finger sequences of many known DNA-binding proteins. These regions are indicated by shading in Fig. 4. They conform to the general form of Zn finger motifs, Phe-X-Cys-$X_{(2-4)}$-Cys-X_3-Phe-X_5-Leu-X_2-His-$X_{(3-5)}$-His [as described by Miller *et al.* (1985) and Berg (1988)]. We showed in the experiment of Fig. 3 that a truncated version of the P3A1 protein including only amino acids 104–274 and thus these two finger sequence elements, is capable of binding the P3A target sites specifically.

In Fig. 5 the P3A1 finger region is shown aligned with a number of other finger proteins, all of which are known directly to serve as DNA-binding regulatory

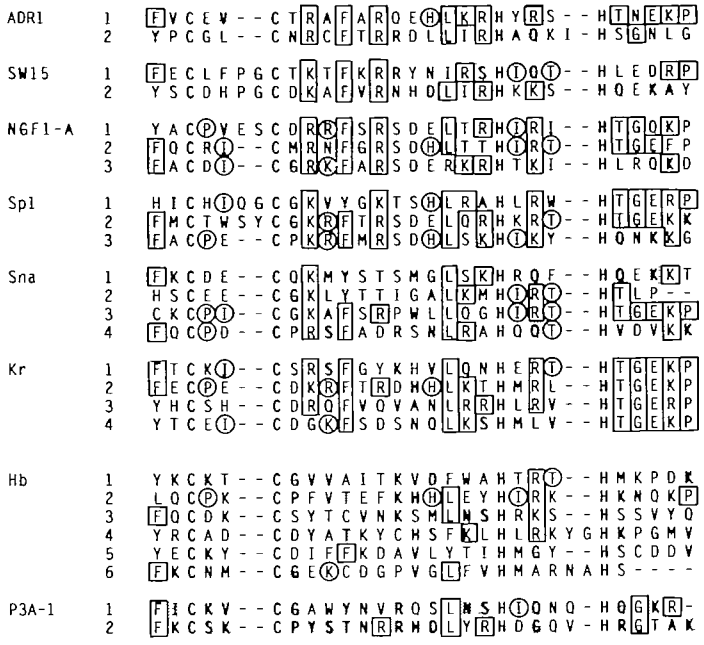

Fig. 5. Comparison of selected finger sequences with those of P3A1. References for the sequences shown are as follows: ADR1; Hartshorne *et al.* 1986; SWI5, Stillman *et al.* 1988; NGF1-A, Milbrandt, 1987; Sp1, Kadonaga *et al.* 1987, snail, Boulay *et al* 1987; Kruppel, Rosenberg *et al.* 1986; hunchback, Tautz *et al.* 1987). At two locations, dashes have been introduced to produce the maximum alignments The horizontal space separates the *hunchback*-P3A1 group of 8 fingers from the remaining 18 finger elements. The canonical C-C and H-H residues shared by the known Zn finger structures and with P3A1 are shown in bold face. Residues other than these that occur in given positions in ≥9 of the 18 fingers of the top comparison group are boxed In this comparison, arginine (R) and lysine (K) are considered equivalent. Residues that occur commonly but in less than half of the 18 finger elements in the top comparison group are circled (arbitrarily considered as ≥5/18 occurrences). Residues that occur in given positions in the P3A1 fingers and *also* in any of the other finger elements are shaded.

factors, so as to display the subfamilial relationships among the examples included (see legend). The invariant C-C and H-H residues (bold face in Fig. 5), as well as several other residues that frequently (i.e. in over half the examples) also occur in specific locations in Zn finger domains (boxed residues) are also present in the P3A1 finger sequences. There is no evidence for P3A1 that Zn^{2+} or other metals are required for DNA binding, however. Fig. 5 displays a specially close relationship between the sequence of the P3A1 fingers and those of the *Drosophila hunchback* factor. Amino acids shared (in register) between either of the P3A1 finger sequences and any of the other finger sequences in Fig. 5 are shaded (excluding those that are common to more than half of all the Zn fingers shown, i.e. the invariant C and H residues and the boxed residues). It can be seen that of the six *hunchback* fingers two share six such residues and two five with the P3A1 finger sequences. One finger sequence of the four *snail* fingers also shares five such residues, while all other fingers in Fig. 5 are less similar. Furthermore, other amino acids that are common (circled) in the 18 finger regions excluding the *hunchback* and P3A1 fingers are rarely seen in the latter two sets of fingers. For example, five of the six *hunchback* fingers lack the aromatic tryptophan or phenylalanine residue that occurs four amino acids beyond the second conserved cysteine and that is present in all 18 of the other fingers shown in Fig. 5. Berg (1988) suggested that this aromatic amino acid is part of a hydrophobic core important for the folding of zinc fingers and that this amino acid can be replaced by another aromatic residue positioned two amino acids following the second cysteine; indeed 50 % of both the *hunchback* fingers and the P3A1 fingers display tryptophan or phenylalanine at this position, while these amino acids are found in none of the other 18 finger structures included in Fig. 5.

Sequence similarities in the DNA-binding domains of P3A1 and P3A2

Standard comparison of the P3A1 and P3A2 sequences indicated that the proteins are almost entirely dissimilar (only about 11 % identical amino acids). This result poses something of a paradox, since as shown in Fig. 3 the two factors bind to essentially the same target DNA sequences. To obtain a more focussed insight, we produced N- and C-terminal deletions of the P3A2 protein, and determined their effect on its DNA-binding activity, since the nature of the P3A2 DNA-binding domain was not obvious *a priori* from any features of its primary sequence. A cDNA clone containing a complete coding sequence. for P3A2 (Calzone *et al.* 1991) was truncated at its 3′ end at conveniently located restriction sites, or at its 5′ end by oligonucleotide primed PCR mutagenesis, as described in Materials and methods. The truncated clones were then transcribed using the vector T7 polymerase start site, and the RNAs were translated *in vitro*. The series of truncated products resulting from the N-terminal and C-terminal deletions are shown respectively in Fig. 6a1 and b1. Gel shift reactions with the *SM50* target site were then carried out with these polypeptides as shown in parts a2 and b2 of Fig. 6. At the N terminus, the P3A2 region required for DNA binding evidently includes a sequence that lies between amino acids 25 and 90, since of the N-terminal deletions, only that terminating at position 25 still recognizes the P3A probe [Fig. 6a2]. Deletions to amino acid 90 and all further deletions completely abolish binding. Truncation from the C-terminal end showed that a P3A2 polypeptide including residues 1–258 exhibits normal binding [Fig. 6b2], while further deletion to position 222 abolishes specific DNA binding. The DNA-binding domain of P3A2 is thus <233 and >130 amino acids in length and lies between positions 25–90 and 222–258. Though of low resolution, this delimitation of the P3A2 DNA-binding region afforded an interesting observation, shown in Fig. 7A. When carefully examined for sequence similarities with P3A1, five amino acids that lie within or adjacent to the first finger region of P3A1

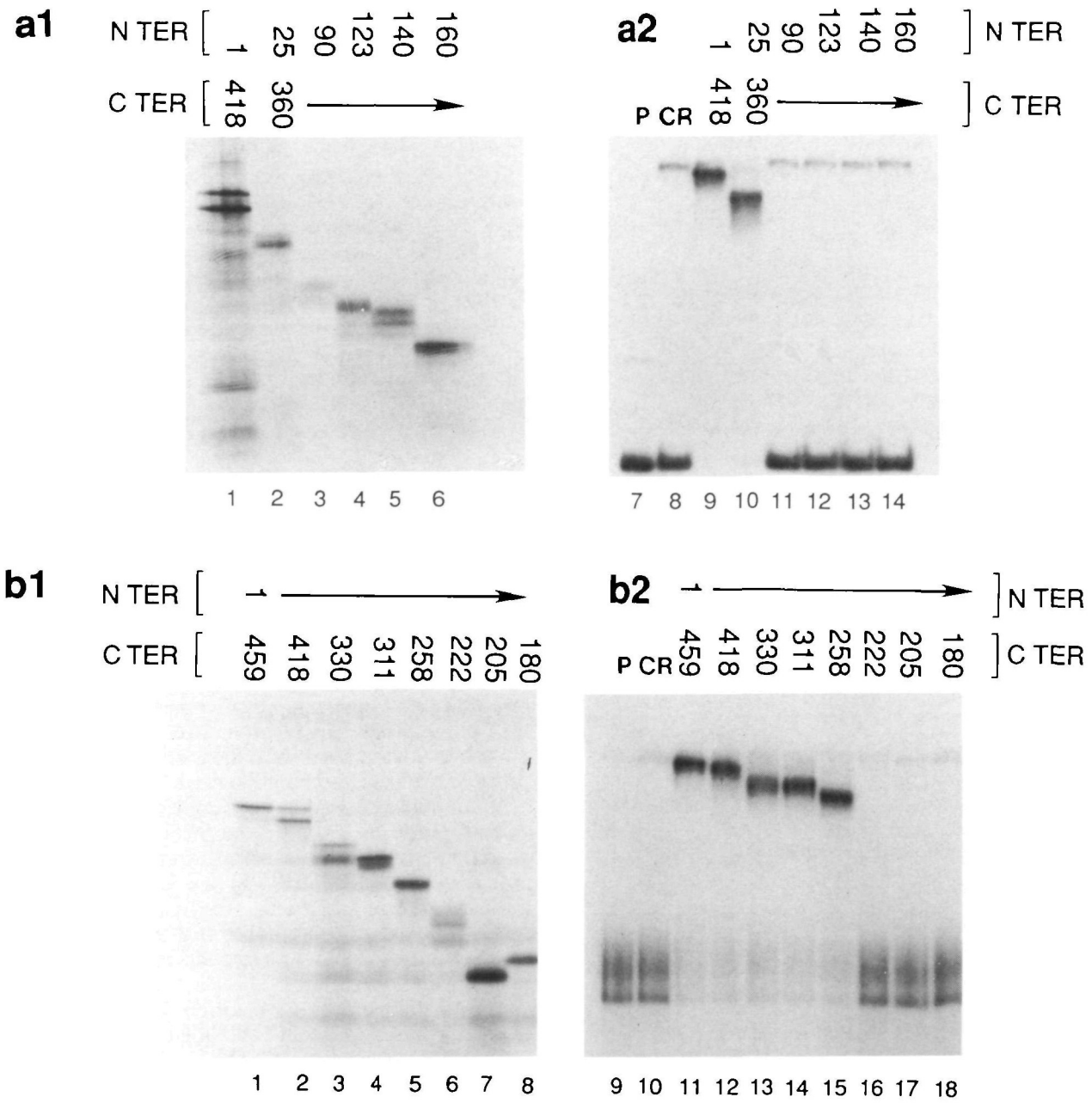

Fig. 6. Delimitation of DNA-binding domain of P3A2 using recombinant deletion mutants. 5′ and 3′ deletions of the P3A2 cDNA were constructed as described in Materials and methods, and transcribed *in vitro* using the vector T7 polymerase site. These transcripts were then translated in a cell-free reticulocyte system, and labeled with [^{35}S]methionine. In a1 and b1, the translation products are displayed on an SDS–polyacrylamide gel by autoradiography The respective N termini and C termini of the various products are indicated at the top of each panel In a1 is shown products of the N-terminal deletion series. Lane 1 shows the translation products of a clone that extends from position 1 to 418. The remainder of the N-terminal deletions were created from a starting clone truncated at the 3′ end at position 360 In b1, illustrating products of the C-terminal deletion series, lane 1 shows the product of a clone that includes the complete coding sequence, i.e. positions 1–459. In a2 and b2 are shown gel shift assays that demonstrate retention or loss of specific DNA-binding activity of the polypeptides shown respectively in a1 and b1 The probe was the *SM50* binding site shown in Fig 3C Lanes 7 in a2 and 9 in b2 marked P contain only probe, and the adjacent lanes, marked CR contain probe and reticulocyte extract but no synthetic P3A2 polypeptide. Free probe migrates to the bottom of the gels and the specific complexes, formed in very large excess of poly(dI)·poly(dC), migrate more slowly. As expected, their mobility is observed to increase as polypeptide size decreases

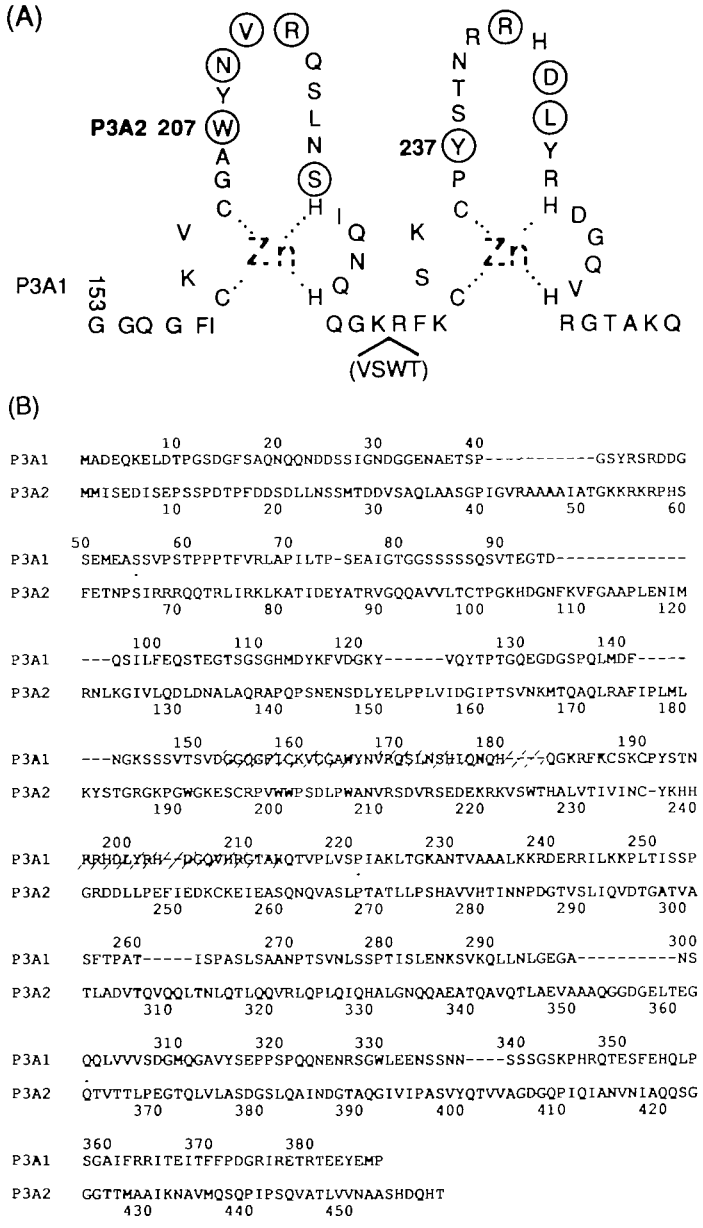

Fig. 7. Comparison of P3A1 and P3A2 sequences. (A) The finger regions of P3A1 are shown conventionally aligned as two Zn-coordinated finger structures (note that we show no evidence for the actual involvement of Zn in this paper). Circled amino acids occur, in the same registration within the DNA-binding domain of P3A2 as well, as shown in B To obtain the registration shown by the circled amino acids in the downstream (right) P3A1 finger, four amino acids in the region between the fingers must be inserted in the P3A1 sequence where indicated by the parenthesis, which includes the P3A2 residues at this point. (B) Maximized alignment of P3A1 and P3A2 sequences (ALIGN program: Myers and Miller, 1988). This alignment yields only about 11% identity, even with free insertion of gaps in the shorter (P3A1) sequence. The alignment shown was chosen by registering the two clusters of identical amino acid residues that occur in the finger regions of P3A1, as shown in A.

are also found, in registration, within the DNA-binding region of P3A2. Furthermore, if a gap of four amino acids is allowed between the fingers of P3A1, a similar feature is revealed in the second finger of P3A1, where there are four amino acids in the DNA-binding region of P3A2 that are identical, in register, with respect to this P3A1 finger. In both fingers, the identical amino acids begin with the aromatic amino acid following the second cysteine. However, as Fig. 7B shows, the overall sequences of P3A1 and P3A2 show no other significant similarities whatsoever, even when the alignment is chosen on the basis of these nine identical finger region residues. Since the two factors recognize almost the same target sequences, these observations invite a direct test by mutagenesis of the significance of these nine amino acids for the target specificity of the P3A2 protein.

Discussion

The P3A1 target site appears in the regulatory domains of at least three territory-specific genes that are activated regionally in the early sea urchin embryo and, in one of these genes, interactions at this site are known to be required for the correct spatial pattern of expression. In this paper, we describe the P3A1 DNA-binding factor, which specifically interacts with these target sites. Unexpectedly, a second factor, P3A2, that recognizes the same target sites, has also been discovered, as described in the first paper of this series. As we demonstrate in Fig. 7, except for small sequence elements within the DNA-binding regions of these two proteins, the P3A1 and P3A2 mRNAs code for wholly dissimilar polypeptides. P3A1 displays canonical features of the Zn-finger class of DNA-binding protein, while P3A2 resembles no known DNA-binding proteins, and thus appears to represent a novel class of regulatory factors (Calzone *et al.* 1991).

DNA-binding structures of P3A1 and P3A2 factors

P3A1 evidently contains two adjacent finger sequence elements of the C-C/H-H class, as shown explicitly in Fig. 5. This structure was initially recognized in the repetitive DNA-binding domain of the TFIIIA transcription factor (Miller *et al* 1985), in which each of the repeat elements is organized into a finger-like structure by a tetrahedrally coordinated Zn ion at its base. Many DNA-binding proteins utilizing this canonical structure are now known, and a variety of structural and mutational evidence demonstrates that the finger structure is indeed that which mediates the sequence-specific interactions with DNA target sequences in which these proteins engage functionally (e.g. Diakun *et al.* 1986; Parraga *et al.* 1988; Nagai *et al.* 1988; Kadonaga *et al.* 1987; Berg, 1988; Redemann *et al.* 1988). Thus it is not surprising that a truncated 24×10^3 version of the P3A1 protein that includes its two finger sequence elements displays the same site specificity as does the complete 43×10^3 protein.

An interesting subdivision of the class of finger

proteins is shown in Fig. 5, where it can be seen that the P3A1 fingers share a number of sequence features with the fingers of the *hunchback* factor and, like the *hunchback* factor, lack certain residues that are otherwise common amongst DNA-binding Zn finger protein sequences. However, the target site for the *hunchback* factor, $^{G}/_{C}{}^{A}/_{C}$ATAAAAAA (Stanojević *et al.* 1989) displays no similarity whatsoever with the target site for the P3A1 factor, $^{C}/_{T}X^{C}/_{T}$GCGC$^{A}/_{T}$. Thus the specific sequences that are responsible for target site recognition within the finger region must include amino acids besides those that are shared specifically between the *hunchback* and P3A1 fingers. This result is consistent with the observation that none of the *hunchback* fingers displays the *sequence* of either of the P3A1 fingers, including the five and four amino acids respectively present in both the P3A1 fingers and the P3A2 DNA-binding domains. Many though not all of these amino acids are located near the tip of the P3A1 fingers, and between the tip and the histidine residues, i.e. in the region believed to be configured into an α-helix that lies in the major groove of the target site DNA, where sequence-specific contacts probably occur (Lee *et al.* 1989; Berg, 1988; Gibson *et al.* 1988; Brown and Argos, 1986). These local sequences might indeed contribute to the target site recognition elements of P3A2 as well as of P3A1, accounting for their almost identical site specificity. Since P3A2 lacks all the standard features of Zn fingers, it would then follow that there must be alternative ways to organize a three-dimensional structure in which almost the same DNA–protein contacts are formed, that do not require the recognition elements to be mounted on a canonical C-C/H-H Zn finger. However, it is important to keep in mind that a much larger domain of the P3A2 factor is apparently required for DNA binding than is included in the two finger structures of P3A1. The deletion experiments shown in Fig. 6 indicate the minimum necessary DNA-binding domain of P3A2 to encompass $\geqslant$140 amino acids. Of course this could include sequences required for other functions interspersed amongst the elements that immediately contact the DNA target site. Since no DNA-binding proteins displaying significant sequence similarities to P3A2 have been reported, much more information will be required to develop an image of structure–function relationships within this large region of the P3A2 factor. The interesting fact remains that a finger protein and a non-finger protein recognize the same regulatory sites with nearly the same affinity.

Possible biological significance

Specification of both the aboral ectoderm and the skeletogenic mesenchyme, the territories of the embryo in which the *CyIIIa* and *Spec*I, and the *SM50* genes are respectively expressed, occurs during cleavage (Davidson, 1986, 1989). Both P3A1 and P3A2 are encoded by low abundance maternal mRNAs; there are about 300 copies of P3A1 mRNA per egg, rising to about 2000 per embryo at 24 h (at that stage ~4 molecules/cell), and about 1100 copies of P3A2 mRNA per egg, a level that remains approximately constant (Cutting *et al.* 1990). Calculations show that the levels of these mRNAs present in the very early embryo would just suffice to provide sufficient levels of newly synthesized P3A factors in the nuclei of the mid-cleavage stage founder cells, when active P3A factors should first be required (Cutting *et al.* 1990). Since the amounts of P3A1 and P3A2 mRNA are approximately the same by mid-cleavage the amounts of the two factors are probably comparable. However, we do not yet know where in the very early embryo the P3A1 and P3A2 factors are located. By late, even mid-embryogenesis, aboral ectoderm is a wholly committed, terminally differentiated cell type, in contrast to oral ectoderm and the vegetal plate derivatives, which continue to differentiate in new directions (Davidson, 1986, 1989; Cameron *et al.* 1987). We suspect that in so far as it pertains to aboral ectoderm gene expression the P3A regulatory system is a *specification* system that functions very early in development. Thus the factor is present even before the *CyIIIa* gene is activated (Calzone *et al.* 1988), and competitive interference with P3A2 interactions *in vivo* results in ectopic specification. These factors could, however, have other functions in the still differentiating embryonic cell types of the oral ectoderm and gut-mesenchyme territories. But it seems most unlikely that these maternally encoded factors are physically absent from the aboral ectoderm founder cell lineages *in early cleavage*. Initially the egg is radially symmetrical in terms of *potential* to generate oral *versus* aboral ectoderm. That is, any quadrant of the 4-cell embryo can give rise to a complete larva (Davidson, 1989). An hypothesis (Davidson, 1989) is that the cytological polarization that is the initial indication of this axis causes a *modification* of radially distributed factors such as P3A1 and P3A2 that would affect their activity along this dimension of the embryonic space.

In considering the P3A regulatory system, we begin with the argument that two different factors that interact with the same or overlapping sites are likely to perform opposite functions, i.e. to act competitively. Since we know that at least one of these factors behaves *negatively* the other might be required for transcriptional *activation*. The literature contains several examples of negative factors that probably function by competitively interfering or modifying positive interactions, in which both factors utilize the same or overlapping binding sites. Among such examples are the positive and negative interactions in the interferon gene (Keller and Maniatis, 1988); the CCAAT and CCAAT-displacement factor interactions in a sea urchin histone gene (Barberis *et al.* 1987); and the overlapping and oppositely acting K_r and *bcd* interactions, and *bcd* and *giant* interactions identified in the 'stripe 2' regulatory element of the *Drosophila eve* gene (Small *et al.* 1990). For our case an additional clue is provided by the *in vitro* interaction studies we present in Figs 2 and 3 of this paper. The most striking difference in DNA-binding activities between P3A1 and P3A2 that we have detected is their response to the target sites in the *SM50* regulatory domain. Table 2 of Calzone

et al. (1991) shows that P3A2 binds about 50× more tightly to the double site in this gene than to the single *CyIIIa* site, and about 10× more tightly to a third target site of the *SM50* gene located proximally. In contrast, as shown in this paper, P3A1 recognizes all of the target sites with about the same affinity (Fig. 2), which is also about the same affinity that P3A2 displays for the *CyIIIa* site. Thus, at relatively *low* factor concentration, the *SM50* target site would be expected to be occupied by P3A2, while the *CyIIIa* site would not. The two factors could both be involved in regulating both aboral ectoderm and skeletogenic genes if they function competitively. Thus, for example, we may suppose that P3A2 is the factor that functions negatively, as a repressor of transcriptional expression, and that P3A1 competitively antagonizes this repression. Then, were the ratio of P3A1 to P3A2 to increase during development in favor of P3A1 by a modest factor, say over a 10-fold range about equivalence, the transcription of genes such as *CyIIIa* would be allowed, assuming the remainder of the required transcription factors were available (Thézé *et al.* 1990; Calzone *et al.* 1988). However, at such ratios repression of *SM50* expression by P3A2 would everywhere be dominant. To account for *SM50* activation exclusively in the skeletogenic mesenchyme lineage it is likely, as postulated earlier (Davidson, 1989), that other positive skeletogenic gene activators should be present *ab initio* in this autonomously specified lineage. A testable prediction is that there should be a much smaller concentration of the P3A2 repressor relative to P3A1 in the confined cellular domain from which the skeletogenic founder cells descend, i.e. the 5th cleavage micromeres, than in any other lineage elements of the cleavage stage embryo.

We thank Mariann Brunkener for construction of the λgt11 library. This work was supported by an NIH grant HD-05753. Support for the initial phases were provided by a Developmental Biology grant from the Lucille P. Markey Charitable Trust C.H. was supported by a postdoctoral fellowship from EMBO.

References

Barberis, A , Superti-Furga, G and Busslinger, M (1987) Mutually exclusive interaction of the CCAAT-binding factor and a displacement protein with overlapping sequences of a histone gene promoter *Cell* **50**, 347–359

Berg, J. M (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins *Proc natn Acad Sci U S A* **85**, 99–102

Boulay, J L , Dennefeld, C and Alberga, A (1987) The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers *Nature* **330**, 395–398

Brown, R S and Argos, P. (1986) Fingers and helices *Nature* **324**, 215

Calzone, F J , Höög, C., Teplow, D B , Cutting, A E., Zeller, R W., Britten, R J and Davidson, E. H (1991) Gene regulatory factors of the sea urchin embryo I. Purification by affinity chromatography and cloning of P3A2, a novel DNA binding protein *Development* **112**, 335–350

Calzone, F J., Thézé, N , Thiebaud, P., Hill, R L , Britten, R J and Davidson, E. H (1988) Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo *Genes Dev* **2**, 1074–1088

Cameron, R. A., Hough-Evans, B R , Britten, R J and Davidson, E H (1987) Lineage and fate of each blastomere of the eight-cell sea urchin embryo *Genes Dev* **1**, 75–85

Cutting, A E , Höög, C , Calzone, F J , Britten, R J and Davidson, E H (1990) Rare maternal mRNAs code for regulatory proteins that control lineage specific gene expression in the sea urchin embryo *Proc natn Acad Sci U S A* **87**, 7953–7959

Davidson, E H (1986) *Gene Activity in Early Development* Academic Press Orlando, Florida

Davidson, E H (1989) Lineage-specific gene expression and the regulative capacities of the sea urchin embryo A proposed mechanism *Development* **105**, 421–445

Diakun, G P , Fairall, L and Klug, A. (1986) EXAFS study of the zinc-binding sites in the protein transcription factor IIIA *Nature* **324**, 688–689

Dingwall, C and Laskey, R. A (1986) Protein import into the cell nucleus *Ann Rev Cell Biol* **2**, 367–390

Gibson, T. J , Postma, J P M , Brown, R S and Argos, P (1988) A model for the tertiary structure of the 28 residue DNA-binding motif ('zinc finger') common to many eukaryotic transcriptional regulatory proteins *Protein Eng* **2**, 209–218

Hartshorne, A T , Blumberg, H and Young, E T (1986) Sequence homology of the yeast regulatory protein ADR1 with transcription factor TFIIIA *Nature* **320**, 283–287

Henikoff, S (1984) Unidirectional digestion with Exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.

Hope, I A and Struhl, K (1986) Functional dissection of a eucaryotic transcriptional activator, GCN4 of yeast *Cell* **46**, 885–894

Hough-Evans, B R., Franks, R R , Zeller, R W , Britten, R J and Davidson, E H (1990). Negative spatial regulation of the lineage specific CyIIIa actin gene in the sea urchin embryo *Development* **108**, 41–50.

Kadonaga, J T , Carner, K R , Masiarz, F. R and Tjian, R (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain *Cell* **51**, 1079–1090

Keller, A D and Maniatis, T. (1988) Identification of an inducible factor that binds to a positive regulatory element of the human b-interferon gene. *Proc natn Acad Sci U S A* **85**, 3309–3313

Kozak, M. (1987). Analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucl Acids Res* **15**, 8125–8143.

Laemmli, U K (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4 *Nature* **227**, 680–685

Lee, J J , Calzone, F J , Britten, R J , Angerer, R C. and Davidson, E H (1986) Activation of sea urchin genes during embryogenesis *J. molec. Biol* **188**, 173–183

Lee, M S , Gippert, G P , Soman, K. V , Case, D A and Wright, P. E. (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245**, 635–637

Mierendorf, R C and Pfeffer, D (1987) Direct sequencing of denatured plasmid DNA *Meth Enzym* **152**, 556–562

Milbrandt, J (1987) A nerve growth factor-induced gene encodes a possible transcriptional regulatory factor *Science* **238**, 797–799

Miller, J., McLachlar, A D. and Klug, A. (1985). Repetitive zinc binding domains in the protein transcription factor IIIA from *Xenopus* oocytes *EMBO J* **4**, 1609–1614.

Miskimins, W , Roberts, M , McClelland, A and Ruddle, F (1985) Use of a protein blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter specific region of the transferrin receptor gene *Proc natn Acad Sci U S A* **82**, 6741–6744

Mitchell, P J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins *Science* **245**, 371–378

Myers, E and Miller, W (1988) Optimal alignments in linear space *CABIOS* **4**, 11–17

NAGAI, K , NAKASEKO, Y , NASMYTH, K. AND RHODES, D (1988) Zinc-finger motifs expressed in *E coli* and folded *in vitro* direct specific binding to DNA *Nature* **332**, 284–286

NORMAN, C , RUNSWICK, M., POLLOCK, R AND TREISMAN, R (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element *Cell* **55**, 989–1003

PARRAGA, G , HORVATH, S J., EISEN, A , TAYLOR, W E , HOOD, L , YOUNG, E T AND KLEVIT, R. E. (1988) Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* **241**, 1489–1492

POSAKONY, J W , FLYTZANIS, C N , BRITTEN, R J AND DAVIDSON, E H. (1983). Interspersed sequence organization and developmental representation of cloned poly(A) RNAs from sea urchin eggs *J molec Biol* **167**, 361–389.

REDEMANN, N , GAUL, U AND JACKLE, H (1988) Disruption of a putative Cys-zinc interaction eliminates the biological activity of the *Krüppel* finger protein *Nature* **332**, 90–92

RIGBY, P W. J , DIECKMAN, M , RHODES, C AND BERG, P (1987). Labeling deoxyribonucleic acid to high specific activity *invitro* by nick translation with DNA polymerase I *J molec Biol* **237**, 237–251

ROSENBERG, U B , SCHRÖDER, C., PREISS, A , KIENLIN, A , COTE, S , RIEDE, I AND JACKLE, H (1986). Structural homology of the product of the *Drosophila Kruppel* gene with transcription factor IIIA *Nature* **319**, 336–338.

SANGER, F , NICKLEN, S AND COULSON, R. A (1977) DNA sequencing with chain-terminating inhibitors *Proc natn Acad Sci U S A* **74**, 5463–5467

SARKAR, G AND SOMMER, S. S (1989) Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity *Science* **244**, 331–334

SHORE, D AND NASMYTH, K. (1987) Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements *Cell* **51**, 721–732.

SMALL, S , KRAUT, R , HOEY, T. AND LEVINE, M. (1991) Transcriptional regulation of a pair-rule stripe in *Drosophila Genes Dev* (in press)

SORGER, P K AND PELHAM, H R B (1988) Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation *Cell* **54**, 855–864

STANOJEVIĆ, S , HOEY, T AND LEVINE, M (1989) Sequence-specific DNA-binding activities of the gap proteins encoded by *hunchback* and *Kruppel* in *Drosophila Nature* **341**, 331–335

STILLMAN, D J , BANKIER, A T., SEDDON, A , GROENHOUT, E G AND NASMYTH, K A (1988) Characterization of a transcription factor involved in mother cell specific transcription of the yeast *HO* gene *EMBO J* **7**, 485–494

STUDIER, F. W. AND MOFFATT, B A (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes *J molec Biol* **189**, 113–130.

TAUTZ, D , LEHMANN, R , SCHNURCH, R , SEIFERT, E , KIENLIN, A., JONES, K AND JACKLE, H (1987) Finger protein of a novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes *Nature* **327**, 383–389.

THÉZÉ, N , CALZONE, F J , THIEBAUD, P , HILL, R L., BRITTEN, R J AND DAVIDSON, E H (1990). Sequences of the CyIIIa actin gene regulatory domain bound specifically by sea urchin embryo nuclear proteins *Molec Reprod Dev* **25**, 110–122

THIEBAUD, P , GOODSTEIN, M , CALZONE, F J , THÉZÉ, N , BRITTEN, R J AND DAVIDSON, E H (1990) Intersecting batteries of differentially expressed genes in the sea urchin embryo. *Genes Dev* **4**, 1999–2010

TSAI, S -F , MARTIN, D I K , ZON, L I , D'ANDREA, A D , WONG, G G AND ORKIN, S H (1989) Cloning of the cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells *Nature* **339**, 446–451

VINSON, C. R., LAMARCO, K L , JOHNSON, P F , LANDSCHULTZ, W H AND MCKNIGHT, S L (1988) *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage *Genes Dev* **2**, 802–806

WATSON, C AND JACKSON, J (1985) An alternative procedure for synthesizing double-stranded cDNA for cloning in phage and plasmid vectors In *DNA Cloning, A Practical Approach*, vol 1 (ed D Glover), pp 79–88 Oxford IRL Press.

YOUNG, R AND DAVIS, R (1983) Efficient isolation of genes by using antibody probes *Proc natn Acad Sci U S A* **80**, 1194–1198

(Accepted 6 February 1991)