Zfy1 encodes a nuclear sequence-specific DNA binding protein

Pamela Taylor-Harris, Sally Swift, Alan Ashworth*

CRC Centre of Cell and Molecular Biology, Chester Beatty Laboratories, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK

Received 19 January 1995; revised version received 3 February 1995

Abstract ***Zfy1*** **is a mouse Y chromosomal gene encoding a zinc finger protein which is thought to have some function during spermatogenesis. Here we show that, when introduced into tissue culture cells, Zfy1 is targeted to the nucleus. Two independent signals are present within the protein for nuclear localization. This nuclear Zfy1 protein is able to bind strongly to DNA-cellulose and, using site-selection assays, we have identified specific Zfy1 DNA binding sites. Taken together these results suggest that Zfy1 is a nuclear-located sequence-specific DNA binding protein which functions during spermatogenesis.**

Key words: Spermatogenesis; Zinc finger protein; Y chromosome; Nucleus; DNA binding; Transcription factor

1. Introduction

The *ZFY* gene was originally identified as a human Y chromosomal gene that was a candidate for the testis-determining factor [1]. Although a primary role in sex determination has now been ruled out [2,3] the conservation of this gene on the Y chromosome of placental mammals suggests it has some important male-specific function, perhaps in spermatogenesis [4]. A gene closely related to *ZFY*, called *ZFX*, is also present on the X chromosome of all placental mammals so far analysed [4]. In mouse, however, this gene family is larger. Two genes, *Zfy1* and *Zfy2*, are present on the mouse Y chromosome [4], and an autosomal gene, *Zfa*, has been shown to be derived by retroposition of a transcript of *Zfx* [5]. Analysis of the sequence of several members of this gene family from several species [4,6] has revealed that they all encode proteins with thirteen repeats of the consensus sequence for the zinc finger motif [7,8].

The zinc finger is one of the structural motifs characteristic of DNA binding proteins [9]. It consists of a characteristic arrangement of cysteine and histidine residues which are thought to co-ordinate Zn^{2+} ions. This maintains the finger structure which can interact with the major groove of DNA in a sequence-specific manner [10,11]. Large numbers of genes encoding zinc finger proteins (ZFPs) are present in eukaryotic genomes [12,13]. Some of these have been identified as transcription factors such as TFIIIA [8] and Sp1 [14], developmental genes in *Drosophila* such as *Kruppel* [15] and genes implicated as oncogenes [16,17] or tumour suppressor genes [18]. Many other *ZFP* genes have been isolated by low stringency hybridization or hybridization with consensus oligonucleotide probes [9,12,13,19]. Although the presence of zinc fingers is often taken to be indicative of a role in transcription, some ZFPs may have a structural role, while others may bind exclusively to RNA and have functions in RNA storage or processes such as splicing.

In order to analyse the properties of the ZFY family of zinc finger proteins, we have raised specific antibodies to murine Zfy1. These have been used to demonstrate that Zfy1 protein localizes to the nucleus of transfected cells and to identify specific DNA binding sites for Zfy1, suggesting a role for this protein as a transcription factor regulating gene expression during spermatogenesis.

2. Materials and methods

2.1. Isolation of monoclonal antibodies to Zfy1

A 1.7 kb fragment encoding the C-terminal basic and zinc finger domains (residues 285–784) of *Zfy1* was amplified by PCR from the plasmid p955 [20] subcloned into the vector pRK171 [21] and introduced into *E. coli* bacterial strain BL21/DE3(pLysS). Electrophoresis of bacterial lysates on SDS-polyacrylamide gels revealed the induction by IPTG of a protein of 62 kDa, as predicted from the open reading frame of this part of *Zfy1* [20]. This C-terminal fragment of Zfy1 was named Zfy1C. The majority of the Zfy1C protein was found to be insoluble and present in inclusion bodies in the bacteria. These were purified by centrifugation at $12{,}000 \times g$ for 15 min at 4°C, solubilised in 1 M urea and the Zfy1C protein purified by preparative gel electrophoresis. The protein preparation was estimated to be greater than 95% pure.

To raise monoclonal antibodies to Zfy1C, 70 μg of purified protein was used for each of two immunizations in CBH/Cbi rats. Spleen cells were fused to myeloma Y3 cells and hybridoma colonies grown under HAT selection [22]. Three positive hybridomas (MAb14a, MAb25a and MAb79a) were obtained by screening by immunoblotting against total cell lysates of bacteria expressing Zfy1C protein. These antibodies were typed as IgA (MAb14a), IgG2b (MAb25a) and IgM (MAb79a).

2.2. COS cell transfection and immunofluorescence

Full-length *Zfy1* cDNA and the various deletion mutants were cloned into the vector pcDNA I (Invitrogen) for expression in COS cells. *Zfy1* was tagged with the epitope (EQKLISEEDLN) of the Myc antibody 9E10 [23] using oligonucleotides inserted into the *Nco*I site corresponding to the initiating methionine. The protein A expression vector consisted of the human β-globin 5′ untranslated region and initiating methionine linked to the mature coding sequence of *Staphylococcus aureus* protein A from pRIT11 [24] cloned into the expression vector pCDM8 ([25] and S.S. and A.A., unpublished). COS cell transfection was as described by Aruffo et al. [25]. Transfection efficiency was in the range of 20–30%. For immunofluorescence the anti-Zfy1 monoclonal MAb25a was used at a 1:50 dilution. Myc-epitope tagged proteins were stained with an anti-Myc monoclonal 9E10 [23] at 1:4 dilution and protein A expression was detected by staining with an anti-protein A monoclonal antibody (Sigma) at a 1:100 dilution. A Nikon Optiphofluorescent microscope equipped with a Bio-Rad MRC 600 confocal imaging system was used for visualisation.

2.3. DNA cellulose column chromatography

Nuclear extracts from COS cells transfected as described above were prepared as described by Lassar et al. [26] in buffer W (20 mM HEPES pH 7.6; 1.5 mM $MgCl_2$; 0.1 mM EDTA; 0.5 mM $ZnCl_2$; 0.02% sodium azide; 1 mM DTT; 1 mM PMSF; 0.3 U/ml aprotinin; 5 μg/ml chymostatin; 10 μg/ml leupeptin; 5 μg/ml antipain; 5 μg/ml pepstatin A) containing 500 mM NaCl and 0.1% Triton X-100. These were diluted

*Corresponding author. Fax: (44) (171) 352 3299.

SSDI 0014-5793(95)00141-7

to 100 mM NaCl in buffer W and then loaded through a fine needle at a rate of 100 μl per minute onto a 5 cm long (0.5 cm diameter) column containing DNA cellulose in a SMART system (Pharmacia). The column was then washed with 1ml of buffer W containing 100 mM NaCl. A step gradient of NaCl was added up to 2 M NaCl. 50 μl fractions were assayed by Western blotting with the anti-Zfy1 monoclonal antibody MAb25a at 1:2,500 dilution followed by an anti-rat HRP conjugated antibody at 1:7,500 dilution. The blot was developed using the ECL detection kit (Amersham).

2.4. Identification of a DNA Binding Site for Zfy1

Recombinant Zfy1C protein was prepared for DNA binding by denaturation in urea followed by renaturation [27]. The protein was first denatured in 1.5 × buffer Z (150 mM KCl, 37.5 mM HEPES, pH 7.9, 1.5 mM DTT, 35 mM $ZnSO_4$, 0.15% NP-40; 30% glycerol, 1.5 mM PMSF) with 6 M guanidine HCl for 30 min at 4°C on a rotating wheel. The denaturant was slowly removed by dialysis against buffer Z with 1 M guanidine HCl for 3 h at 4°C. This was then repeated in buffer Z but without guanidine HCl to allow for gradual renaturation of the protein.

Sequences binding to Zfy1C were selected from a pool of double-stranded 76-mer oligonucleotides which contained a central region of 26 random bp, flanked on either side by sequences of 25 bp each, which allowed recovery of small quantities of bound DNA by PCR amplification [28]. The sequences of these oligonucleotides were: Random sequence 76-mer oligonucleotide (R76), 5′-CAGGTCAGATCAGCGGATCCTGTCG$(N)_{26}$GAGGCGAATTCAGTGCATGTGCAGC-3′, 'Forward' primer, 5′-GCTGCACATGCACTGAATTCGCCTC-3′; 'Back' primer, 5′-CAGGTCAGATCAGCGGATCCTGTCG-3′. Random oligonucleotide binding assays were performed essentially as previously described [29,30]. Briefly, 1ug of purified renatured protein was electrophoresed on a 10% SDS-PAGE gel, electroblotted onto a nitrocellulose filter and the filter incubated at 4°C with the ^{32}P-labelled pool of random oligonucleotides (10^7–10^8 cpm/μg) prepared by primed synthesis [28]. After washing, the filters were autoradiographed and the area of the filter corresponding to bound DNA excised and the DNA eluted by heating to 100°C in water. Eluted DNA was amplified by 30 cycles of PCR using 'forward' and 'back' primers, and the 76 bp product purified by agarose gel electrophoresis. For subsequent rounds of binding, the 76 bp product was labelled by 18 cycles of PCR in the presence of ^{32}P-labelled nucleotide, as described [28]. DNA which remained bound to the Zfy1C protein after the fourth round of selection was eluted, amplified, labelled and used in subsequent rounds of enrichment by electrophoretic mobility shift assays (EMSAs) [29]. Briefly, 10 ng of purified, renatured Zfy1C protein was pre-incubated in 20 μl of binding buffer (10 mM HEPES, pH 7.9, 60 mM KCl , 1 mM DTT, 1 mM EDTA, 0.25mg/ml bovine serum albumin and 12% glycerol) with 1 μg poly(dI-dC) competitor DNA and 10 fmol (≈20,000 cpm) of ^{32}P-labelled, annealed oligonucleotides were added for 30 min at room temperature. The reactions were electrophoresed in non-denaturing 4% acrylamide gels in 0.5 × TBE buffer at room temperature at 10 V/cm for 90 min and the gels dried prior to autoradiography. The region of the dried gel containing DNA bound to the Zfy1C protein in the presence of 1 μg poly(dI-dC) was heated at 100°C in water and the eluted DNA amplified as for the ROB assays. After a further five rounds of enrichment selection DNA was amplified as above, cloned into pBluescript and the inserts from 23 clones sequenced [31].

3. Results

3.1. Zfy1 is located in the nucleus of transfected tissue culture cells

We expressed part of the coding sequence of murine Zfy1 in *E. coli* under the control of a T7 polymerase responsive promoter. The 62 kDa protein produced, which contains the basic domain and the 13 C-terminal zinc fingers of Zfy1, was insoluble and located in inclusion bodies in the bacteria. Using centrifugation, solubilization in urea and preparative gel electrophoresis we were able to isolate milligram quantities of the protein in essentially pure form (data not shown). Three monoclonal antibodies (Mab14a, Mab25a and Mab79a) were raised to this protein. These antibodies recognised Zfy1 in both western blots and ELISA assays (data not shown). We used these antibodies to investigate the subcellular location of Zfy1 introduced into COS cells by transfection. A full-length cDNA clone coding for murine *Zfy1* [20] was cloned into the mammalian expression vector pcDNA I in both sense and antisense orientations and introduced by transfection into COS cells. The subcellular location of the Zfy1 protein was determined by indirect immunofluorescence with the monoclonal antibody MAb25a. Clear nuclear staining was obtained in about 30% of cells transfected with an expression vector containing *Zfy1* (Fig. 1a (ii)) but not when the vector alone (Fig. 1a (iv)) was transfected. This reflects the approximately 30% transfection efficiency obtained. No staining was obtained if the Zfy1 antibody or the second antibody was omitted from the staining procedure (data not shown). Thus Zfy1 protein is localised to the nucleus when expressed in COS cells.

3.2. Two nuclear localization signals are present in Zfy1

To determine which regions of Zfy1 were responsible for nuclear targeting of the protein, we examined the sub-cellular location of deletion mutants of the protein. The Zfy family proteins appear to have a common modular structure composed of a large N-terminal acidic domain, a short basic region and the C-terminal zinc finger region [4]. By inspection the basic region located between amino acids 372 and 382 (PKQKSKKKKR) shows some similarity to the SV40 nuclear localization signal PKKKRKV [32]. Part of Zfy1 containing this sequence (amino acid residues 372–384) was fused to the C-terminus of protein A (construct e in Fig. 1b) and the subcellular location of the protein determined by immunofluorescence with an anti-protein A antibody. This demonstrated that this sequence could indeed behave as a functional nuclear localization signal. To determine if this sequence was necessary for nuclear localization of Zfy1, we introduced a 700 bp deletion into Zfy1 (nucleotides 794–1497, amino acids 214–415). This construct (b in Fig. 1b) was introduced into COS cells and the protein localized by immunofluoresence with the MAb25a. Surprisingly this construct also localized to the nucleus. Thus Zfy1 contains two potentially functional nuclear localization signals. The cytoplasmic location of an N-terminal myc-tagged Zfy1 fragment (construct d, Fig. 1b) (amino acid residues 1–214) demonstrated that the second nuclear localization is located in the zinc finger domain of Zfy1.

3.3. Zfy1 is a DNA binding protein

The presence of thirteen copies of the zinc finger motif within Zfy1 suggested that this protein is capable of binding to DNA. To test this possibility, nuclear extracts were prepared from COS cells transfected with the full length Zfy1 expression vector described above. These extracts were chromatographed on double-stranded DNA (dsDNA) cellulose columns in a SMART system (Pharmacia) (Fig. 2). The SMART system allows chromatography of small amounts of protein with minimal losses. After loading, the column was washed extensively with buffer W containing 100 mM NaCl. The column was developed by the application of a stepwise NaCl gradient and column fractions analysed by electrophoresis on SDS/polyacrylamide gels and immunoblotting with antibodies against Zfy1. The Zfy1 protein bound to the dsDNA column and was eluted as tight peak at about 350 mM NaCl (Fig. 2). This

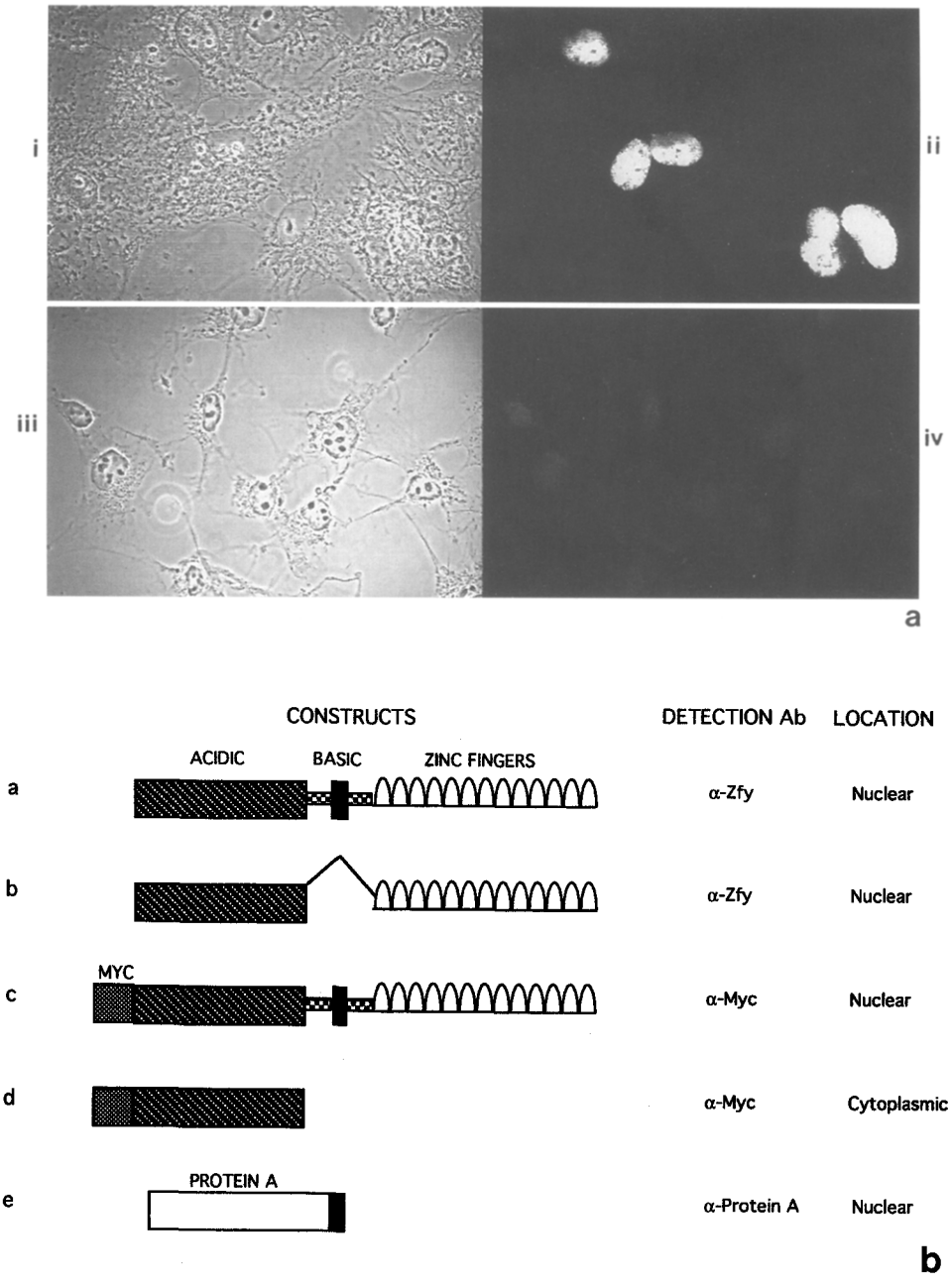

Fig. 1. Nuclear Localisation of Zfy1 expressed in COS cells. (a) Indirect immunofluorescence staining of COS cells some of which are expressing Zfy1 protein. COS cells were grown on coverslips, transfected with a Zfy1 expression vector (i and ii) or the expression vector alone (iii and iv) and after 3 days were stained with anti-Zfy1 monoclonal antibody Mab25a (ii and iv). Parts i and iii show the stained cells as viewed by phase contrast microscopy. (b) Structure and subcellular localization of Zfy1 expression constructs. Expression vectors carrying various parts of Zfy1 either unfused, fused with protein A or tagged with the myc 9E10 antibody epitope were transfected into COS cells. The cells were fixed and the expressed protein localised by indirect immunofluorescence with the appropriate antibody. Constructs are as described in sections 2 and 3. Zfy1 has been divided into three domains as described previously [4].

indicates that Zfy1 expressed in vivo has bona fide DNA binding activity.

3.4. *Identification of a DNA binding site for Zfy1*

The demonstration that Zfy1 had dsDNA binding properties typical of a DNA binding protein led us to investigate whether this protein bound DNA sequence specifically. Recombinant Zfy1C protein was purified by preparative gel electrophoresis, then denatured and re-folded in the presence of zinc cations and used in random oligonucleotide binding assays [30]. Oligonucleotides bound to the Zfy1 protein were eluted and amplified by PCR, then used in further rounds of binding, elution and amplification, to select for sequences that bound specifically. After four rounds of selection, the DNA was used in a further five rounds of enrichment by the electrophoretic mobility shift assay (EMSA). After the final round of selection the eluted

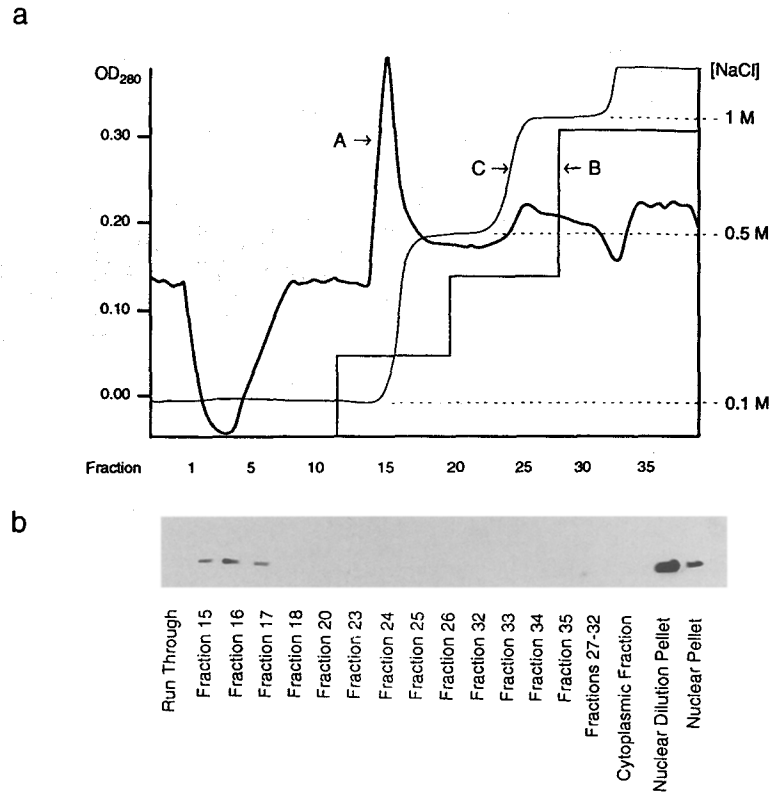

Fig. 2. DNA binding Properties of Zfy-1 Expressed in COS cells. (a) Fractionation of COS cell extracts on DNA cellulose. Trace A: spectrophotometric readout at OD_{280} of fractions eluted from the DNA cellulose column. Trace B: NaCl concentration applied to the DNA cellulose column in a step gradient fashion. Trace C: The salt concentration of fractions eluted from the column. (b) Immunoblot of the fractions eluted from the DNA cellulose column. These were electrophoresed through a SDS-polyacrylamide gel, blotted onto nitrocellulose and probed with anti-Zfy-1 antibody. Fractions are as in part a. Cytoplasmic Fraction and Nuclear Pellet are sub-cellular fractions prepared as described in section 2. Nuclear Dilution Pellet corresponds to nuclear extract diluted to 100 mM NaCl prior to loading on the gradient.

DNA was amplified, cloned and individual sub-clones sequenced. The sequences of individual clones and consensus sequences are shown in Fig. 3. Of the 23 clones analysed 20 contained a sequence corresponding to one of three categories of DNA binding sequence. The other 3 contained no detectable homology. This large enrichment for the motifs underlined in Fig. 3 demonstrates that we have isolated at least part of a specific binding site for Zfy1. All the selected sequences in Fig. 3 contain at least one and in some cases (class I) two copies of the core sequence AGGCC, suggesting that that this may be crucial for DNA binding.

4. Discussion

Zfy1 is a zinc finger protein encoding gene located on the mouse Y chromosome. Although previously thought to have a function in testis determination, such a role has now been excluded [1–3]. It has been suggested that the *Zfy1* gene may have a function in spermatogenesis, perhaps as a transcription factor [4]. Here we examine the subcellular localisation and DNA binding potential of this protein. In initial studies we attempted to determine the subcellular location of endogenous Zfy1 but we were unable to detect the protein in male mouse germ cells by immunofluorescence. We therefore expressed Zfy1 in COS cells and demonstrated that the protein was located in the nucleus of the transfected cells. We went on to show that a previously recognised region of Zfy1 homologous to the SV40 T antigen nuclear localisation signal [20,33] is indeed capable of targeting a heterologous protein to the nucleus. However this was shown to be dispensable for Zfy1 nuclear localisation as a deletion mutant lacking this sequence was also located in the nucleus. It is possible that this second signal is cryptic and revealed only when the normally utilised sequence is removed. This redundancy may be common as several other proteins have been described which contain multiple nuclear localisation signals; the yeast repressor $\alpha 2$ has two non-homologous signals, one located at the N terminus as well as one in the homeodomain [34].

As Zfy1 appears to be a nuclear protein and contains zinc finger motifs characteristic of nucleic acid binding proteins, we decided to examine the DNA binding potential of the protein. We used transient transfection and DNA-cellulose chromatography of nuclear extracts to demonstrate that Zfy1 expressed in vivo was capable of binding to DNA with binding characteristics typical of bona fide DNA binding proteins. In order to determine whether this binding was specific or non-specific we utilised site-selection assays to define DNA binding sites for Zfy1. The similarity of the selected sequences suggested that we had indeed identified specific binding sites. The specificity has been confirmed by the use of band shift assays (data not shown). The identification of distinct but related groups of sequences may reflect the binding of zinc fingers as independent modules [10,11].

The possibility that Zfy1 interacts with a longer DNA bind-

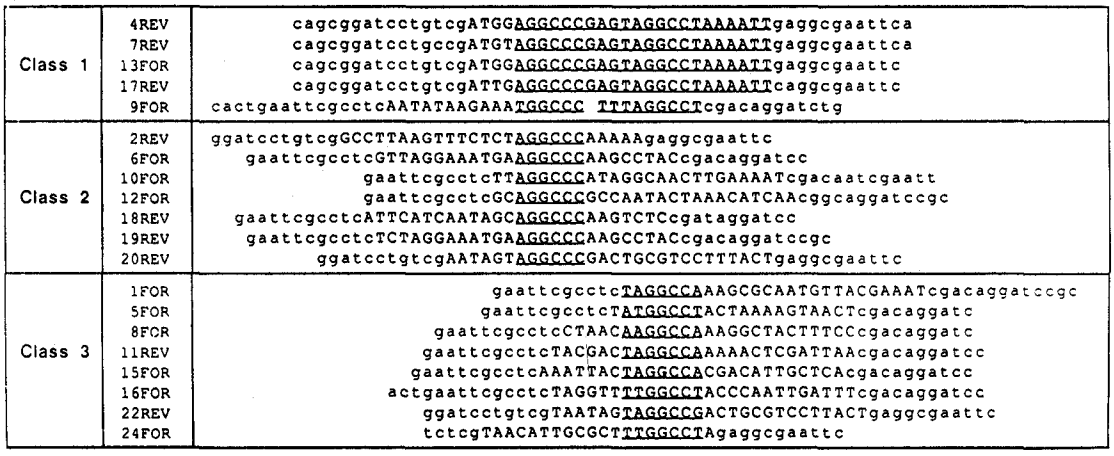

Fig. 3. Sequence of oligonucleotides selected by Zfy1. Recombinant Zfy1C was used in site-selection assays to isolate specific DNA binding sites from a pool of random oligonucleotides. These were cloned and sequenced. The sequences shown have been divided into three classes based on sequence homology and the homologous regions underlined.

ing site of which we have identified part could be addessed by utilising the binding site already identified in combination with a further round of site selection to 'bootstrap' to a longer and presumably higher affinity site. Alternatively the identification of a bona fide in vivo binding site for Zfy1 should allow characterisation of an extended binding site. Interestingly, a sequence related to the site selected sequences identified here is found at the 5′ end of the *Zfx* gene [35]. Recently we have been able to show that Zfy1 can bind to this sequence in gel mobility shift assays (data not shown) and this suggests the intriguing possibility that Zfy1 may be able to trans-regulate the expression of the *Zfx* gene in vivo.

Acknowledgements: This work was supported by the Cancer Research Campaign and the Medical Research Council. We are grateful to C. Dean and J. Cordell for generating monoclonal antibodies, H. Patterson for expertise on the confocal microscope and P. Denny, R. Nicholas and F. Connor for helpful discussions.

References

[1] Page, D., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L.G. (1987) Cell 51, 1091–1104.
[2] Koopman, P., Gubbay, J., Collignon, J. and Lovell-Badge, R. (1989) Nature 342, 940–942.
[3] Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. and Lovell, B.R. (1991) Nature 351, 117–121.
[4] Koopman, P., Ashworth, A. and Lovell-Badge, R. (1991) Trends Genet. 7, 132–136.
[5] Ashworth, A., Skene, B., Swift, S. and Lovell-Badge, R. (1990) EMBO J. 9, 1529–1534.
[6] Connor, F. and Ashworth, A. (1992) Nucl. Acids Res. 20, 5845.
[7] Brown, R.S., Sander, C. and Argos, P. (1985) FEBS Lett. 186, 271–274.
[8] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609–1614.
[9] El Baradi, T. and Pieler, T. (1991) Mech. Dev. 35, 155–169.
[10] Pavletich, N.P. and Pabo, C.O. (1991) Science 252, 809–817.
[11] Pavletich, N.P. and Pabo, C.O. (1993) Science 261, 1701–1707.
[12] Bray, P. and Thiesen, H.-J. (1990) The New Biologist 2, 325–327.
[13] Ashworth, A. and Denny, P. (1991) Mammalian Genome 1, 196–200.
[14] Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) Cell 51, 1079–1090.
[15] Rosenberg, U.B., Schroeder, C., Priess, A., Kienlin, A., Cote, S., Riede, I. and Jackle, H. (1986) Nature 319, 336–339.
[16] Kinzler, K.W., Ruppert, J.M., Bigner, S.H. and Vogelstein, B. (1988) Nature 332, 371–374.
[17] Morishita, K., Parker, D.S., Mucenski, M.L., Jenkins, N.A., Copeland, N.G. and Ihle, J.N. (1988) Cell 54, 831–840.
[18] Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H., Jones, C. and Housman, D.E. (1990) Cell 60, 509–520.
[19] Bellefroid, E., Lecocq, P.J., Benhida, A., Poncelet, D.A., Belayew, A. and Martial, J.A. (1989) DNA 8, 377–387.
[20] Ashworth, A., Swift, S. and Affara, N. (1989) Nucleic Acids Res. 17, 2864.
[21] Studier, F.W. and Moffat, B.A. (1986) J. Mol. Biol. 189, 113–130.
[22] Köhler, G. and Milstein, C. (1975) Nature 256, 495–497.
[23] Evan, G.L., Lewis, G.K., Ramsay, G. and Bishop, M. (1985) Mol. Cell. Biol. 5, 3610–3616.
[24] Lowenadler, B., Nilsson, B., Abrahmsen, L., Moks, T., Ljungqvist, L., Holmgren, E., Paleus, S., Josephson, S., Philipson, L. and Uhlen, M. (1986) EMBO J. 5, 2393–2398.
[25] Aruffo, A. and Seed, B. (1987) Proc. Natl. Acad. Sci. USA 84, 8573–8577.
[26] Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintraub, H. (1991) Cell 66, 305–315.
[27] Hoey, T. and Levine, M. (1988) Nature 332, 858–861.
[28] Pollock, R. and Treisman, R. (1990) Nucl. Acids Res. 18, 6197–6204.
[29] Denny, P., Swift, S., Connor, F. and Ashworth, A. (1992) EMBO J. 11, 3705–3712.
[30] Treacy, M.N., He, X. and Rosenfeld, M.G. (1991) Nature 350, 577–584.
[31] Ashworth, A. (1993) in: Transcription Factors: A Practical Approach (Latchman, D. Ed.) pp. 125–142, IRL Press, Oxford, UK.
[32] Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984) Cell 39, 499–509.
[33] Mardon, G. and Page, D.C. (1989) Cell 56, 765–770.
[34] Hall, M.N., Craik, C. and Hiraoka, Y. (1990) Proc. Natl. Acad. Sci. 87, 6954–6958.
[35] Mardon, G., Luoh, S.W., Simpson, E.M., Gill, G., Brown, L.G. and Page, D.C. (1990) Mol. Cell Biol. 10, 681–688.