Single-strand-DNA-binding factors specifically recognize the pyrimidine element in the chick $\alpha 2(I)$ collagen gene promoter

Dashzeveg BAYARSAIHAN* and Lewis N. LUKENS

Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, CT 06459-0175, U.S.A.

A pyrimidine element with mirror repeats centred at position –192 bp of the chick $\alpha 2(I)$ collagen promoter interacts with sequence-specific DNA-binding factors. These factors bind to only the pyrimidine strand of this region and have no affinity for the complementary purine strand. Binding activity is also seen with the double-stranded form of this element, but with less affinity than to the single-stranded pyrimidine species. South-western blot analyses have shown that proteins of 80 and 134 kDa in chick embryo fibroblast nuclear extracts bind to the pyrimidine

strand, whereas only a 134 kDa DNA-binding protein was found in chondrocyte nuclear extracts. The binding mechanism of these nuclear proteins with single-stranded DNA might be based on a non-B-DNA conformation of the pyrimidine element. The position of this binding site in the promoter region, its potential for adopting an unusual secondary structure and the presence of the 80 kDa binding factor in chick embryo fibroblasts, but not in chondrocytes, suggest a possible role for this factor in the expression of the $\alpha 2(I)$ collagen gene.

INTRODUCTION

The expression of the type I and type II collagen genes is tightly regulated during embryogenesis [1]. However, much remains to be learned about the mechanisms by which these genes are switched on and off during development. One example of a developmental change in the expression of these genes occurs during chondrogenesis where mesenchymal cells, which synthesize type I collagen, a heterotrimer of chain composition $[\alpha 1(I)]_2\alpha 2(I)$, differentiate into chondrocytes that make type II collagen, a homotrimer of a cartilage-specific type II α chain, $[\alpha 1(II)]_3$. Chondrocytes do not synthesize mRNAs encoding either type I α chain [2–4], although, unexpectedly, they do synthesize a transcript from the $\alpha 2(I)$ gene that begins 2.34 kb downstream of the usual start site in fibroblasts. This RNA does not encode a collagen polypeptide and its function is unknown [3]. When cultured chondrocytes are induced to change to a fibroblastic/mesenchymal phenotype by treatment with 5-bromo-2'-deoxyuridine, they stop transcribing the $\alpha 2(I)$ gene from the downstream start site, and shift to initiating transcription at the fibroblastic start site, with accompanying synthesis of $\alpha 2(I)$ collagen [2,5].

Analysis of the two different promoter regions responsible for transcription from these cell-type-specific start sites indicates that this developmental regulation of the $\alpha 2(I)$ collagen gene is mediated by sequence-specific DNA-binding proteins [6]. In the case of the fibroblast promoter the fragment from –712 to –85 bp upstream of the fibroblast transcription start site has been shown to be important for promoter activity, and mobility shift assays suggested that the region from –283 to –85 bp may be responsible for the major complex formed between this DNA fragment and proteins present in fibroblast nuclear extracts [6]. Within this region a pyrimidine element from –205 to –181 bp, containing mirror repeats of the sequence TCCCTCCCT, has been partially conserved between the chicken and mouse genes and has been proposed to be important in the regulation of $\alpha 2(I)$ collagen gene transcription [7]. Several nuclear factors bind

specifically to this site in the mouse promoter [8]. A major DNase I hypersensitive site that is associated with the transcriptional activity of this promoter also maps to this region in both chick and mouse chromatin [5,9]. This hypersensitive site is absent from the chromatin of chondrocytes that do not initiate transcription at the fibroblast start site, but it re-appears when bromodeoxyuridine-treated chondrocytes shift to utilization of the fibroblast start site [5].

An *in vitro* S1-nuclease-sensitive site maps precisely to this pyrimidine element [10,11]. This sensitivity could be associated with the formation of a non-B-DNA structure [12]. Because of the possible importance of this DNA region in the $\alpha 2(I)$ collagen promoter for regulation of the $\alpha 2(I)$ collagen gene, we undertook a search for nuclear proteins that might specifically interact with this pyrimidine element. We now report the finding of DNA-binding activities specific for the single-stranded form of this pyrimidine element.

MATERIALS AND METHODS

Isolation of cells and preparation of nuclear extract

Chick embryo fibroblasts were isolated from decapitated 10-day embryos (purchased from Spafas, Norwich, CT, U.S.A.) by a standard procedure [2] and incubated on tissue culture plates in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (Gibco) and 10% CO₂. Chondrocytes were isolated from the sterna of 14-day chick embryos by the floater selection method [5]. Nuclei were prepared from cells chilled to 0 °C and nuclear extracts were then prepared by a standard procedure [6] and stored in aliquots at –80 °C. Protein concentration was assayed using the BCA kit (Pierce) with a BSA standard.

Oligonucleotides

Oligonucleotides CT and GA were purchased from Integrated DNA Technologies. M1, M2, M3, M4, Sp1 and their comp-

* To whom correspondence should be addressed.

Table 1 Oligonucleotides used in the present work

Sequences are given for each oligonucleotide, with mutated nucleotides underlined.

Oligonucleotide	Sequence
CT	TCTTCGCTCCCTTCTCCCTCCCTC
GA	GGGCGAGGGAGGGAGGAAGGGAGGG
dsWT	Double-stranded CT/GA oligonucleotide
ssM1	TCTTCGCTGGCTTCTCCCTCCCTC
dsM1	Double-stranded species of M1
ssM2	TCTTCGCTCCCTTCTCCGCTCACTC
dsM2	Double-stranded species of M2
ssM3	TCTTCGCTGGCTTCTCCGCTCACTC
ssM4	TCTTCGCTCCCTAGGTCCCTCCCTC
ssSp1	TTCAGACGTTCCCGCCCTC
dsSp1	Double-stranded species of Sp1
ssM13	CACAATTCCACACAAC

plementary strands were obtained from Eppendorf. M13 represents the Sequencing Primer, Hybridization Probe M13, catalogue no. 71202, from USB. The sequences of the oligonucleotides are shown in Table 1.

Double-stranded oligonucleotides were prepared by annealing the complementary strands, and were purified from single-stranded oligonucleotides on a 20% polyacrylamide gel. The double-stranded CT/GA oligonucleotides had a 4-base overhang at each 5' end which was filled in with the Klenow enzyme prior to use. The dsSp1 oligonucleotide refers to the Sp1-like sequence present in the chick $\alpha 2(I)$ collagen promoter, from -336 to -312 inclusive [13]. The ssSp1 oligonucleotide contains the pyrimidine-rich sequence on the non-coding strand, from -332 to -312 inclusive.

Electrophoretic mobility shift assay

Single-stranded oligonucleotides were radiolabelled using T4 polynucleotide kinase (New England Biolabs) and [γ - 32 P]ATP (Amersham). Double-stranded oligonucleotides were labelled by end-filling with Klenow enzyme. Labelled nucleic acids were purified using a Microcon-3 spin column (Amicon). Binding reactions were carried out in 50 mM Hepes/KOH, pH 7.9, 50 mM KCl, 4% (v/v) glycerol, 1 mM $MgCl_2$ and 0.5 mM dithiothreitol in a final volume of 20 μ l with 2–4 μ g of a nuclear extract, 5×10^4 c.p.m. of a labelled probe and 1 μ g of poly(dI-dC) (Pharmacia). After a 20 min incubation at room temperature, samples were resolved on a 5% native polyacrylamide gel in $0.5 \times$ TBE [14] for 2 h at 200 V at 4 °C.

Southwestern blot analysis

The nuclear proteins were resolved on a discontinuous 10% SDS/polyacrylamide gel [14]. Proteins (20 μ g/lane) were electrophoretically transferred from unstained gels to nitrocellulose filters for 5 h at 140 mA. The filters after transfer were treated with a blocking/renaturation solution (50 mM Tris/HCl, pH 8.0, 50 mM NaCl, 2 mM EDTA, 2 mM β -mercaptoethanol, 0.1% Nonidet P-40 and 5% non-fat dried milk) for 1 h at room temperature with gentle shaking. Then filters were incubated in a binding solution [10 mM Tris/HCl, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM β -mercaptoethanol, 0.5 mM PMSF and 10 μ g of poly(dI-dC)] overnight at 4 °C with 3×10^5 c.p.m./ml labelled probe. The filters were washed for 3×15 min each at room temperature with the same binding solution without PMSF

and poly(dI-dC), and were dried and exposed to Kodak X-AR film for an appropriate time at -80 °C [14].

RESULTS AND DISCUSSION

Single-stranded oligonucleotides (see Table 1) corresponding to the conserved mirror repeat sequence in the pyrimidine/purine asymmetrical region of the $\alpha 2(I)$ collagen gene promoter, from -205 to -181 relative to the transcription start site [13], were used in gel mobility shift assays. When these oligonucleotides were incubated with nuclear extract from 10-day chick embryo fibroblasts, a retarded complex was observed with the labelled CT strand which was competed away by a 10–20-fold molar excess of unlabelled single-stranded CT oligonucleotide (Figure 1). However, no complex-formation was observed with the labelled purine (GA) strand (Figure 1), indicating that this binding activity is specific for the pyrimidine strand.

To determine the specificity of the interaction between the single-strand-binding nuclear factor(s) and the pyrimidine sequence, we tested increasing amounts of a series of mutant oligonucleotides in competition experiments using the mobility shift assay. The mutant oligonucleotides ssM1 and ssM2, which contain a disruption in the 5' and 3' elements of the mirror repeat respectively, competed poorly for binding to the nuclear factors compared with the wild-type CT sequence (Figure 2A). When both of these disruptions were present in the same sequence (ssM3), very little competition was seen even at 300-fold excess (Figure 2B). However, a change of three nucleotides in the middle of the CT oligonucleotide (ssM4), without disruption of either element of the mirror repeat, did not decrease the binding affinity, since this ssM4 oligonucleotide competed with the retarded complex as well as did the wild-type CT oligonucleotide (Figure 2B). These results suggest that the nuclear factor(s) specifically recognize the pyrimidine sequences present in the mirror repeats. The conclusion that the single-stranded pyrimidine-binding activity is sequence-specific was further re-

Figure 1 Fibroblast nuclear factors bind to the single-stranded pyrimidine sequence in the $\alpha 2(I)$ collagen promoter but not to the complementary purine strand

Gel mobility shift assays were performed with nuclear extracts from 10-day chick embryo fibroblasts using 5' end-labelled single-stranded CT and GA oligonucleotides as probes (see the Materials and methods section and Table 1). Lanes 1 and 8, no extract; lanes 2, 4, 6 and 9 contain probes without competitors; lanes 3, 5, 7 and 10–12 contain excess unlabelled competitors.

Figure 2 Mutations in the mirror repeat elements of the single-stranded pyrimidine sequence decrease binding activity

Competition with the binding of the ^{32}P -labelled wild-type CT sequence by the mutant sequences, as noted, was assayed by the mobility shift assay. All lanes contained the ^{32}P -labelled CT oligonucleotide (probe). Lane 1 contained no extract and lane 2 contained no unlabelled competitor. A molar excess, as indicated, of the unlabelled mutant single-stranded oligonucleotides was added to all other lanes. Mutations in three bases of the left-hand and right-hand elements of the mirror repeats respectively are present in ssM1 and ssM2, and both mutated elements are present in ssM3, whereas mutations in the three bases that separate these elements are present in ssM4 (see the Materials and methods section and Table 1).

inforced by the finding that the single-stranded Sp1 and M13 oligonucleotides, containing 75% and 56% CT respectively, did not compete for binding when present at a 20-fold molar excess (Figure 3A). Increasing the amount of the Sp1 competitor up to a 300-fold excess did not decrease the band shift (results not shown).

Competition experiments were also performed with double-stranded oligonucleotides. A 20-fold molar excess of the homologous double-stranded CT/GA oligonucleotide (WT) competed much less efficiently than the single-stranded CT oligonucleotide, as shown by comparison of lanes 3 and 4 in Figure 3(A), and of lanes 1–7 in Figure 3(B) with lanes 1–7 of Figure 1. The double-stranded versions of the M1 and M2 oligonucleotides, with mutations in the elements of the mirror repeat, competed as poorly as their single-stranded counterparts (Figure 3A, lanes 5–8). At higher levels of competitor virtually complete competition was seen with a 160-fold excess of the double-stranded WT oligonucleotide and with a 300-fold excess of the double-stranded M1 oligonucleotide (Figure 3B). These experiments demonstrate that the factors that recognize the single-

Figure 3 The single-stranded wild-type pyrimidine sequence is bound with higher affinity than the double-stranded wild-type sequence

Competition with the binding of the wild-type CT sequence (probe) by an excess of the unlabelled oligonucleotide, as indicated, was assayed as in Figure 2. Lane 1 contained no extract, lane 2 contained no competitor, and lanes 3–12 contained a molar excess of the unlabelled competitor, as indicated.

stranded wild-type pyrimidine sequence have 5–10-fold less affinity for the double-stranded wild-type sequence. In summary, the nuclear factors detected in these assays bind with high affinity only to a pyrimidine single strand and exhibit a sequence specificity for the intact mirror repeat present in the pyrimidine-rich sequence of the $\alpha 2(I)$ collagen gene promoter.

Southwestern blots were performed in order to estimate the sizes of the pyrimidine-specific binding factors. Two major polypeptides of 80 and 134 kDa that bound the CT oligonucleotide were present in fibroblast nuclear extracts, whereas only a 134 kDa protein was observed in chondrocyte nuclear extracts (Figure 4A). To obtain additional evidence that the same binding factors were being detected in the mobility shift and Southwestern assays, the mutated CT oligonucleotide containing a disruption in both elements of the mirror repeat (ssM3) was also tested in the Southwestern blot. As predicted from the gel mobility shift assays (Figure 2B), this oligonucleotide was not bound by the proteins that bound the wild-type CT oligonucleotide in the Southwestern blots (Figure 4B). Further analysis is necessary to determine if the 134 kDa polypeptides present in chondrocytes and fibroblasts are identical. The presence of the 80 kDa protein in fibroblasts, but not in chondrocytes, raises the possibility that it plays a role in the tissue-specific switch in transcription start sites in the $\alpha 2(I)$ collagen gene.

Figure 4 Detection by Southwestern blots of 80 and 134 kDa polypeptides in fibroblast nuclear extracts that bind the single-stranded pyrimidine mirror repeat

The blots were probed with 5' end-labelled single-stranded CT (ssCT; **A**) or mutant M3 (ssM3; **B**) oligonucleotide. Each lane contained 20 μ g of nuclear extract from chondrocytes (lane 1) or fibroblasts (lane 2). Protein standard molecular masses ('Rainbow'; Amersham) are indicated on the left (in kDa).

Several specific single-strand-DNA-binding proteins have been described (for example, see [15–25]). However, in contrast to the many double-stranded-DNA-binding proteins that have been characterized and cloned, only a limited number of sequence-specific single-strand-DNA-binding proteins have been described. The proteins described in the present paper represent a new example, but differ from those previously described in that they specifically recognize a pyrimidine mirror repeat and have an unusually high molecular mass. The function(s) of these proteins in chick embryonic cells, their role, if any, in the regulation of the $\alpha 2(I)$ collagen gene, and their relationship with other previously described single-strand-DNA-binding proteins remains to be elucidated.

We thank Catherine Farrell for helpful discussions and Badam Enhmandah for technical assistance. This work was supported by grant 93-37206-9495 from the United States Department of Agriculture.

REFERENCES

- 1 Vuorio, E. and de Crombrughe, B. (1990) *Annu. Rev. Biochem.* **59**, 837–872
- 2 Askew, G. R., Wang, S. and Lukens, L. N. (1991) *J. Biol. Chem.* **266**, 16834–16841
- 3 Bennett, V. D. and Adams, S. L. (1990) *J. Biol. Chem.* **265**, 2223–2230
- 4 Farrell, C. M. and Lukens, L. N. (1995) *J. Biol. Chem.* **270**, 3400–3408
- 5 Beck, K. M., Seekamp, A. H., Askew, R. G., Mei, Z., Farrell, C. M., Wang, S. and Lukens, L. N. (1991) *Nucleic Acids Res.* **19**, 4975–4982
- 6 Wang, S. and Lukens, L. N. (1994) *Matrix Biol.* **14**, 313–321
- 7 Schmidt, A., Yamada, Y. and de Crombrughe, B. (1984) *J. Biol. Chem.* **259**, 7411–7415
- 8 Karsenty, G. and de Crombrughe, B. (1990) *J. Biol. Chem.* **265**, 9934–9942
- 9 McKeon, C., Pastan, I. and de Crombrughe, B. (1984) *Nucleic Acids Res.* **12**, 3491–3502
- 10 Finer, M. H., Fodor, E. J. B., Boedtker, H. and Doty, P. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1659–1663
- 11 McKeon, C., Schmidt, A. and de Crombrughe, B. (1984) *J. Biol. Chem.* **259**, 6636–6640
- 12 Larsen, A. and Weintraub, H. (1982) *Cell* **29**, 609–622
- 13 Boedtker, H., Finer, M. and Aho, S. (1985) *Ann. N. Y. Acad. Sci.* **460**, 85–116
- 14 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Press, Cold Spring Harbor, NY
- 15 Yee, H. A., Wong, A. K. C., Van de Sande, J. H. and Rattner, J. B. (1991) *Nucleic Acids Res.* **19**, 949–953
- 16 Santoro, I. M., Yi, T.-M. and Walsh, K. (1991) *Mol. Cell. Biol.* **11**, 1944–1953
- 17 Quinn, J. P. and McAllister, J. (1993) *Nucleic Acids Res.* **21**, 1637–1641
- 18 Altio, S. and Groner, B. (1994) *Mol. Cell. Biol.* **14**, 6004–6012
- 19 Haas, S., Stepkowski, A., Siracusa, L. D., Amini, S. and Khalili, K. (1995) *J. Biol. Chem.* **270**, 12503–12510
- 20 Smidt, M. P., Russchen, B., Snippe, L., Wijnholds, J. and Ab, G. (1995) *Nucleic Acids Res.* **23**, 2389–2395
- 21 Rajavashisth, T. B., Taylor, A. K., Andalibi, A., Svenson, K. L. and Lusk, A. J. (1989) *Science* **245**, 640–643
- 22 Hollingsworth, M. A., Closken, C., Harris, A., McDonald, C. D., Pahwa, G. S. and Maher, L. J. (1994) *Nucleic Acids Res.* **22**, 1138–1146
- 23 Ito, K., Sato, K. and Endo, H. (1994) *Nucleic Acids Res.* **22**, 53–58
- 24 Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Arigan, M. and Levens, D. (1994) *Genes Dev.* **8**, 465–480
- 25 Kolluri, R., Torrey, T. A. and Kinniburgh, A. J. (1992) *Nucleic Acids Res.* **20**, 111–116