Human HLA-DRα gene: a rare oligonucleotide (GTATA) identifies an upstream sequence required for nuclear protein binding

Rafaella Barbieri[1], Patrizio Giacomini[2], Stefano Volinia[1], Claudio Nastruzzi[1], Anna Maria Mileo[3], Umberto Ferrini[3], Marco Soria[4], Italo Barrai[5], Pier Giorgio Natali[2] and Roberto Gambari[1]

[1]*Istituto di Chimica Biologica, Università di Ferrara, Ferrara,* [2]*Laboratorio di Immunologia, Istituto Regina Elena, Roma,* [3]*Laboratorio di Biofisica, Istituto Regina Elena, Roma,* [4]*Farmitalia, Carlo Erba, Milano* and [5]*Istituto di Zoologia, Università di Ferrara, Ferrara, Italy*

Received 22 May 1990

Synthetic oligonucleotides containing putative regulatory sequences are currently employed to identify and isolate genes coding for nuclear binding factors. Upstream DNA sequences of eukaryotic genes required for transcriptional activity and tissue specificity can be identified by means of biochemical techniques as well as computer analysis using homology searching. An alternative approach has been recently proposed by our research group. Scanning DNA sequences 1.8 megabases in length from a Genetic Sequence Data Bank, we have identified rare oligonucleotides 5 base pairs (bp) long, which are localized within or close to regulatory segments in mammalian promoters. In this paper we demonstrate that the rare GTATA sequence identifies an upstream region of the HLA-DRα gene which operates in conjunction with the sequence AGAAGTCAG, homologous to a box found in many interferon-inducible genes, in binding nuclear proteins.

Oligonucleotide; DNA-binding factor; HLA-DRα

1. INTRODUCTION

Conserved upstream DNA sequences (CUS), playing biological functions have been identified in eukaryotic genes by means of classical computer approaches (homology searching of 'consensus' sequences) [1-3] as well as by biochemical analyses, including DNA-binding protein data [4], footprint analysis [5], and CAT assay [6].

We have recently proposed a computational approach alternative to homology searching, in order to identify putative 'signal' sequences possibly involved in regulating the expression of eukaryotic genes [7,8]. We have designed and tested computer programs to calculate the frequencies of short oligonucleotides in genomic regions [7]. Scanning a 1.8 megabases (Mb) genomic sample from the BBN GenBank Genetic Sequence Data Bank [9], rare 5-mers have been identified that are preferentially located in the non-coding portions of eukaryotic genes, including the 5′ upstream regions [8].

The aim of the present investigation was to test whether the analysis of the location of rare 5-mers could be useful to identify DNA regions capable to bind nuclear proteins. The model system used in our study is the human HLA-DRα gene [10], that displays a rare 5-mer (GTATA) 278 nucleotides upstream of the start of transcription.

Correspondence address: R. Gambari, Istituto di Chimica Biologica, Università di Ferrara, Via L. Borsari n.46, 44100 Ferrara, Italy

Abbreviations: mer, oligonucleotide; Mb, megabase; bp, base pairs; CUS, conserved upstream DNA sequences

2. MATERIALS AND METHODS

2.1. *Distribution of the GTATA sequence in mammalian upstream gene regions*

We used the GenBank Genetic Sequence Data Bank, Release 48.0 (February, 1987) obtained from BBN Laboratories (Cambridge, MA) [9]. The distribution of the 5-mer GTATA was analysed in 160 mammalian promoters at least 400 bp long (total sample size of 64 kbp) using the PROMOTER program [7,8].

2.2. *Synthetic oligonucleotides*

The wild-type synthetic oligonucleotides [GTATA/IFN-γ (44 bp), GTATA-1 (23 bp), GTATA-2 (5 bp), IFN-γ-1 (21 bp) and IFN-γ-2 (9 bp)] and the mutant synthetic oligonucleotides (M1, M2, M3 and M4, corresponding to the GTATA/IFN-γ mer) are shown in Fig. 2B. 5′-3′ strands and the relative complementary strands were synthesized on a Biosearch DNA Synthesizer using the phosphoramidite method. Equimolar amounts of each strand were 5′ labelled with [γ-^{32}P]ATP by the enzyme T4 polynucleotide kinase (Genenco), combined, heated for 5 min at 80°C in 0.5 M NaCl and annealed for at least 30 min at room temperature (RT).

2.3. *Cell lines*

The human cell lines (erythromyeloid K562, B-lymphoid WI-L2, promyelocytic HL-60, melanoma MNT-1 and MNR-1) [12-14] were all grown in α-medium (Gibco) supplemented with 10 % FCS in 5% CO_2 at 37°C. The EPC (cyprine epithelioma) cell line (obtained from Istituto Zooprofilattico di Torino) was cultured in the same conditions, except temperature (21°C) for optimal growth.

2.4. *Electrophoretic mobility shift assay*

The electrophoretic mobility shift assay (gel retardation) was performed as originally described [15] with minor modifications. Nuclear

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

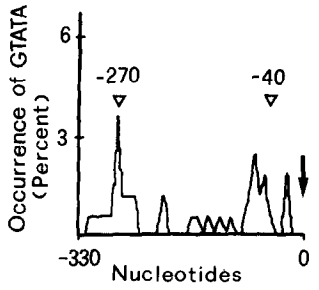


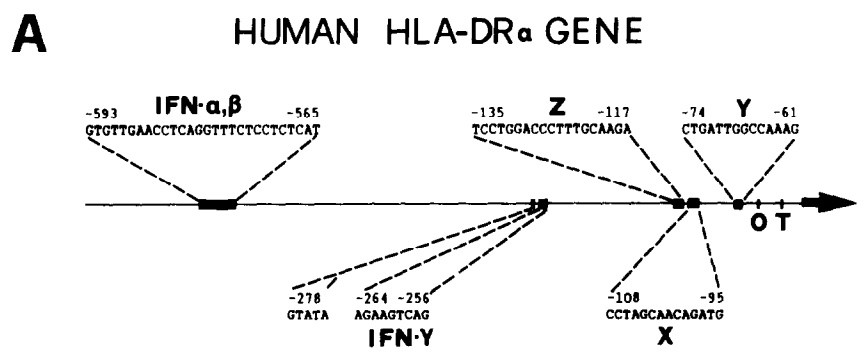
Fig. 1. Distribution of the GTATA rare oligosequence with respect to the start of transcription (solid arrow) of eukaryotic genes. 160 different promoters from the GenBank release 48.0 (February 1987) (BBN Laboratories, Cambridge, USA) were analysed with the program PROMOTER and the location of the GTATA oligonucleotide determined [8]. Results represent the percent of GTATA sequences found within segments of 50 nucleotides of the analysed genes.

extracts were prepared according to Dignam et al. [16] at a protein concentration (BCA assay, Pierce Rockford, IL, USA) of 1–5 mg/ml. Binding reactions, unless otherwise specified, were set up in binding buffer (20 mM Tris-HCl, pH 7.6, 50 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 0.2 mM EDTA, 0.01% Triton X-100, 5% glycerol, 0.5 mM spermidine), in the presence of increasing amounts of poly(dI:dC)·poly(dI:dC) (Pharmacia, Uppsala, Sweden), 1 μg of nuclear extract proteins and 0.25 ng of end-labelled double-stranded oligonucleotides (approximately 50 000 Cerenkow counted cpm), in a total volume of 25 μl. After 30 min at room temperature, samples were electrophoresed at constant voltage (300 V for 2 h) through a low ionic strength (0.35 × TBE buffer) (1 × TBE = 0.089 M Tris-borate, 0.089 M boric acid, 0.008 M EDTA) on a 10% polyacrylamide gel until tracking dye (Bromophenol blue) reached the end of the 16 cm slab. Gels were dried and exposed at −80°C with intensifying screens.

3. RESULTS

3.1. *Distribution of the GTATA 5-mer in mammalian promoters and its location within the upstream sequence of the human HLA-DRα* gene

Fig. 1 shows the distribution of the GTATA mer in the upstream region of 160 mammalian promoters. The distribution is not random, peaking at about −270 bp and at −40 bp from the start of transcription. While the −40 bp peak identifies the TATA box [17], the −270 bp peak identifies a distinct region. Interestingly, the GTATA oligonucleotide of the HLA-DRα gene is



B

Synthetic Oligonucleotides employed

A. Wild-type oligonucleotides

GTATA/IFN-γ mer	5'-CTAATGTGCTTCAGGTATATCCCTGTCTAGAAGTCAGATTGGGG-3' 3'-GATTACACGAAGACCATATAGGGACAGATCTTCAGTCTAACCCC-5'
GTATA-1	5'-CTAATGTGCTTCAGGTATATCCC-3' 5'-GATTACACGAAGACCATATAGGG-5'
GTATA-2	5'-GTATA-3' 3'-CATAT-5'
IFN-γ-1	5'-TGTCTAGAAGTCAGATTGGGG-3' 3'-ACAGATCTTCAGTCTAACCCC-5'
IFN-γ-2	5'-AGAAGTCAG-3' 3'-TCTTCAGTC-5'

B. Mutant oligonucleotides

M-1	5'-CTAATGTGCTTCAGGTATATCCCTGTCTTCTCTTCAGATTGGGG-3' 3'-GATTACACGAAGTCCATATAGGGACAGAAGAGAAGTCTAACCCC-5'
M-2	5'-CTAATGTGCTTCAGAGCCCTCCCTGTCTAGAAGTCAGATTGGGG-3' 3'-GATTACACGAAGTCTCGGGAGGGACAGATCTTCAGTCTAACCCC-5'
M-3	5'-CTAATGTGCTTCAGGTATAGAGAGAGAGAAGTCAGATTGGGG-3' 3'-GATTACACGAAGTCCATATCTCTCTCTCTTCAGTCTAACCCC-5'
M-4	5'-CTAATGTGCTTCAGAGCCCTCCCTGTCTTCTCTTCAGATTGGGG-3' 3'-GATTACACGAAGTCTCGGGAGGGACAGAAGAGAAGTCTAACCCC-5'

Fig. 2. (A) Location and sequences of putative regulatory boxes (*x*, *y*, *z*, IFN-γ, IFN-α,β) of the human HLA-DRα gene [10]. The location of the rare GTATA oligosequence (−278) is indicated. (B) Sequences of the synthetic wild-type (GTATA/IFN-γ, IFN-γ-1, IFN-γ-2, GTATA-1, GTATA-2) and mutant (M1, M2, M3, M4) oligonucleotides employed. The GTATA and the consensus IFN-γ sequences are underlined. The mutated sequences are boxed.

positioned at −278 nucleotides from the start of transcription, and in close proximity (only 9 nucleotides) is located the consensus sequence AGAAGTCAG (known as IFN-γ box), shared by other interferon-inducible genes [18,19] (Fig. 2A).

3.2. *The GTATA 5-mer identifies a sequence required for binding nuclear proteins*

When 5′ end ^{32}P-labelled GTATA/IFN-γ mer (the sequences of the synthetic oligonucleotides used are shown in Fig. 2B) is incubated with nuclear extracts from cell lines of different histotype, a reproducible gel retardation pattern is obtained. At the highest (400 ng) poly(dI:dC) concentration, the binding of non-specific protein(s) (B1) and low-affinity factors (B2,B4) to ^{32}P-labelled GTATA/IFN-γ mer is either diminished or absent, while binding of the B3 factor appears to be only slightly affected (Fig. 3A). This result suggests that B3 exhibits the highest affinity for the GTATA/IFN-γ mer, and was confirmed by five independent experiments (see also Fig. 4 and Fig. 5). Nuclear extract proteins from the epithelioma papulosum carp fish cell line in the presence of the GTATA/IFN-γ mer, did not generate any band shift comparable to the B3 retarded band (Fig. 3B). Competition experiments demonstrate that the B3 band disappears only when cold GTATA/IFN-γ mer is added to the binding reaction (data not shown). In contrast, no competition resulted with the z-box oligonucleotide except when added at high concentrations (200 ng/25 μl of reactions) (data not shown).

Fig. 4 shows that while the GTATA/IFN-γ mer is able to bind nuclear factors, both IFN-γ-1 and GTATA-1 mers do not give rise to any specific complex with the B3 factor, suggesting that simultaneous presence of the GTATA and AGAAGTCAG boxes is mandatory for optimal binding of the nuclear factors.

3.3. *Effects of mutations in the GTATA and AGAAGTCAG sequences on binding of the B3 factor(s)*

In order to investigate the role of GTATA and AGAAGTCAG sequences on the binding of the B3 nuclear factor two sets of experiments were performed using the mutant oligonucleotides shown in Fig. 2B. In the first set of experiments (Fig. 5A) the binding of the B3 factor to the mutant oligonucleotides was determined. In the second set of experiments cold mutant oligonucleotides were used to compete for the binding of the B3 factor to the ^{32}P-labelled GTATA/IFN-γ mer (Fig. 5B).

The results shown in Fig. 5A demonstrate that, since the M3 oligonucleotide binds the B3 factor, mutating the TCCCTGTCT sequence of the oligonucleotide GTATA/IFN-γ does not affect the binding. This suggests that the sequence of the 9 bp spacer between GTATA and AGAAGTCAG in the upstream region of the human HLA-DRα gene does not affect recognition of the B3 factor. Conversely, mutations in the GTATA, AGAAGTCAG or both do affect binding of the B3 nuclear factor, suggesting that the simultaneous presence of GTATA and AGAAGTCAG sequences is required for binding of the B3 factor(s).

Additionally, Fig. 5B shows that M1, M2 and M4 mers have low capacity for competing with the GTATA/IFN-γ oligomer, while the extent of competition by the M3 oligonucleotide was similar to that of the GTATA/IFN-γ mer.

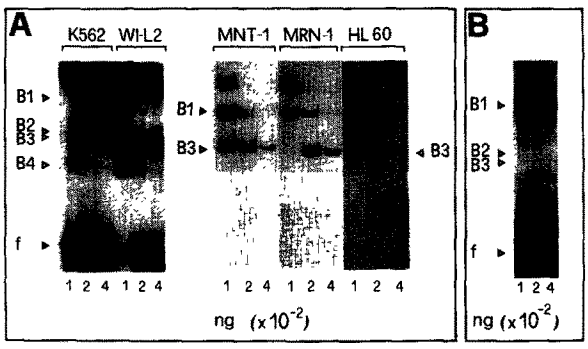


Fig. 3. Gel retardation assays of the binding of nuclear proteins from different cell lines to the GTATA/IFN-γ oligonucleotide. (A) 1 μg of nuclear extracts from the erythromyeloid K562, the B-lymphoid WI-L2, the promyelocytic HL60 and the melanoma, MNT-1 and MNR-1 cell lines were allowed to bind to 0.25 ng of 5′ end-labelled GTATA/IFN-γ mer in a final volume of 25 μl in presence of the indicated amounts of poly(dI:dC)(dI:dC). After 30 min binding at room temperature, the reaction mixtures were electrophoresed on 8% polyacrylamide gels. (B) Binding of the GTATA/IFN-γ mer to nuclear extracts from the EPC carp fish cell line. B1–B4 indicate retarded bands; f = free GTATA/IFN-γ 5′ end labelled mer.

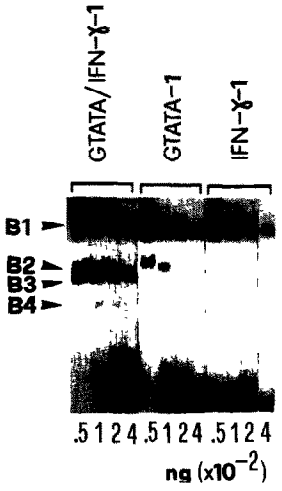


Fig. 4. Analysis by gel retardation of the binding of nuclear proteins to the GTATA/IFN-γ, GTATA-1 and IFN-γ mers. The assay was performed using 1 μg of proteins of a nuclear extract prepared from a B-lymphoid WI-L2 cell line. B1–B4 indicate retarded bands. The indicated concentrations of competitor poly(dI:dC)(dI:dC) were included in the binding reactions (25 μl). The sequences of the oligonucleotides employed are shown in Fig. 2B.

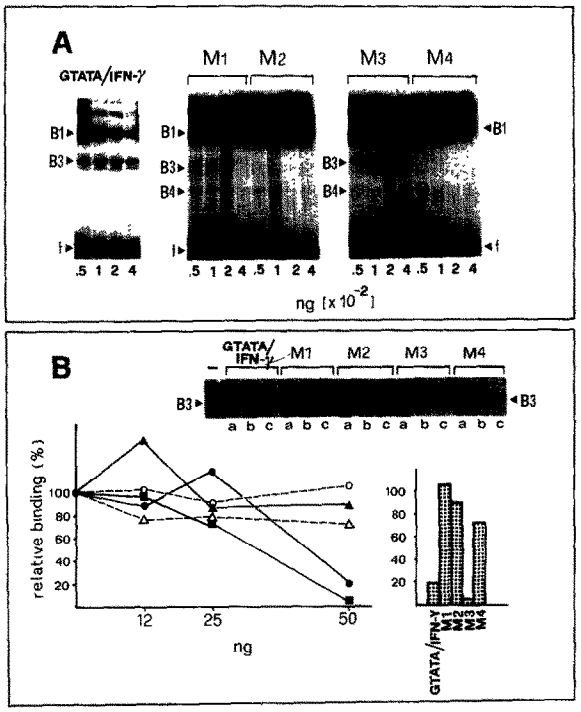


Fig. 5. Effects of mutations on the binding of the GTATA/IFN-γ mer to the B3 factor(s). In this experiment nuclear extracts from the MRN-1 cell line were used. (A) Binding of GTATA/IFN-γ, M1, M2, M3 and M4 oligonucleotides to nuclear factors. B1–B4: retarded bands; f = free oligonucleotide. The amounts of poly(dI:dC)(dI:dC) competitor are indicated. (B) Competition by GTATA/IFN-γ (●), M1 (○), M2 (▲), M3 (■) and M4 (△) mers in the binding of nuclear factors to the 5′-end labelled GTATA/IFN-γ oligonucleotide. Up = autoradiograms; down left = densitometric analysis; down right = comparison of the effects of 50 ng of GTATA/IFN-γ and mutant oligonucleotides in the binding of ^{32}P-labelled GTATA/IFN-γ mer to the B3 factor(s). Binding reactions (25 μl) were carried on for 30 min at room temperature in the presence of the indicated ng of cold competitors. The sequences of the oligonucleotides employed are shown in Fig. 2B.

4. DISCUSSION

In this paper we describe the identification of a sequence located within the upstream region of the human HLA-DRα gene, that (i) contains a rare 5-mer, GTATA, and (ii) specifically binds nuclear proteins. Rare 5-mers were identified by means of a computer-assisted analysis using a non-homology search approach [7,8] and were found in some instances to occur in close proximity to regulatory elements of mammalian promoters [8]. The DNA region centered around the sequence GTATA is located just 5′ of the so-called IFN-γ consensus sequences [19–21] as shown in Fig. 2A.

By means of gel-retardation assay, we demonstrate that the synthetic double-stranded oligonucleotide containing both GTATA and IFN-γ boxes is able to form specific complexes with nuclear factors (retarded band B3 of Fig. 3, Fig. 4 and Fig. 5) from cell lines of different histotype.

Our data show that, if the GTATA, the AGAAGTCAG sequences or both are mutated (Fig. 5) or deleted (Fig. 4), the binding to the B3 factor is much reduced. Taken together, these results consistently indicate that the presence of both GTATA and IFN-γ box is required for the binding of the B3 factor(s). Although the role of the GTATA/IFN-γ sequence and the binding factors remains to be elucidated, our data suggest that the combination of different computer-aided analyses is a useful tool in identifying upstream sequences which bind nuclear proteins.

Acknowledgements: Work supported by CNR P.F. Ingegneria Genetica (R.G.), by 60% MPI (R.G.), by Regione Veneto (R.G.), by CNR P.F. Biotecnologia e Biostrumentazione (P.G.) and by A.I.R.C. (P.G.). We thank Mr Maurizio Matteuzzi, Giancarlo Cortese, Rocco Fraioli and Cynthia Full for the expert technical assistance. We thank Dr Elizabeth Ormondroyd for the revision of the manuscript. R.B. is a recipient of a A.I.R.C. fellowship.

REFERENCES

[1] Wingender, E. (1988) Nucleic Acids Res. 5, 1879–1902.
[2] Boss, J.M. and Strominger, J. (1986) Proc. Natl. Acad. Sci. USA 83, 19139–9144.
[3] Sullivan, K.E., Kalman, A.F., Nakanishi, M., Tsang, S.Y., Wang, Y. and Peterlin, B.M. (1987) Immunol. Today 8, 289–292.
[4] Dorn, A., Durand, B., Marfing, C., Le Meur, M., Benoist, C. and Mathis, D. (1987) Proc. Natl. Acad. Sci. USA 84, 6249–6254.
[5] McKnight, S.L. and Kingsbury, R., Science 217, 316–324.
[6] Gorman, C.M., Moffat, L.F. and Howard, M. (1982) Mol. Cell. Biol. 2, 1044–1056.
[7] Volinia, S., Gambari, R., Bernardi, F. and Barrai, I. (1989) Computer Appl. Biosci. (CABIOS) 5, 33–40.
[8] Volinia, S., Bernardi, F., Gambari, R. and Barrai, I. (1988) J. Mol. Biol., 203, 385–390.
[9] Bilofsky, H.S., Burks, C., Fickett, J.W., Goad, W.B., Lewitter, F.I., Rindone, W.P., Swindell, C.D. and Tung, C.S. (1986) Nucleic Acids Res. 14, 1–11.
[10] Sullivan, K.E., Calman, A.F., Nakanishi, M., Tsang, S.Y., Wang, Y. and Peterlin, B.M. (1987) Immunol. Today 8, 289–293.
[11] Das, H.K., Lawrence, S.K. and Weissman, S.M. (1983) Proc. Natl. Acad. Sci. USA 80, 7024–7029.
[12] Gambari, R., Barbieri, R., Piva, R., Tecce, R., Fisher, P.B., Giacomini, P. and Natili, P.G. (1987) Ann. NY Acad. Sci. USA 511, 292–304.
[13] Lozzio, C.B. and Lozzio, B.B. (1975) Blood 45, 321–329.
[14] Collins, S.J., Gallo, R.C. and Gallagher, R.W. (1977) Nature 270, 347–348.
[15] Fried, M. and Crothers, P.M. (1981) Nucleic Acids Res. 9, 6505–6525.
[16] Dignam, J.D., Lebowitz, R.M. and Roeder, R.G. (1983) Nucleic Acids Res. 11, 1475–1489.
[17] Goldberg, M. (1979) Ph.D Thesis, Stanford University.
[18] Shamboeck, A., Korman, A.J., Kamb, A. and Strominger, J. (1983) Nucleic Acids Res. 11, 8663–8671.
[19] Basta, P.V., Sherman, P.A. and Ting, J.P.Y. (1987) J. Immunol. 138, 1275–1280.
[20] Basta, P.V., Sherman, P.A. and Ting, J P.Y. (1988) Proc. Natl. Acad. Sci. USA 85, 8616–8622.
[21] Tsang, S.Y., Nakanishi, M. and Peterlin, N. (1988) Proc. Natl. Acad. Sci USA 85, 8598–8602.