Molecular anatomy of the human glucose 6-phosphate dehydrogenase core promoter

Annamaria Franzè*, Maria Immacolata Ferrante, Francesca Fusco, Antonietta Santoro, Emma Sanzari, Giuseppe Martini, Matilde Valeria Ursini

Istituto Internazionale di Genetica e Biofisica, CNR, Via Marconi 12, 80125 Naples, Italy

Received 29 July 1998; received in revised form 24 September 1998

Abstract The gene encoding glucose 6-phosphate dehydrogenase (G6PD), which plays a pivotal role in cell defense against oxidative stress, is ubiquitously expressed at widely different levels in various tissues; moreover, G6PD expression is regulated by a number of stimuli. In this study we have analyzed the molecular anatomy of the G6PD core promoter. Our results indicate that the G6PD promoter is more complex than previously assumed; G6PD expression is under the control of several elements that are all required for correct promoter functioning and, furthermore, a still unidentified mammalian specific factor is needed.

Key words: Housekeeping gene; Initiator; Glucose 6-phosphate dehydrogenase; Sp1

1. Introduction

In this paper, we analyze the promoter of the human housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD). The enzyme G6PD has a fundamental role in the cell, because it catalyzes the first regulatory step in the pentose phosphate pathway [1]. Moreover, it has been demonstrated that G6PD activity is part of the cell defense against oxidative stress [2,3]. The gene is X-chromosome-linked and its promoter is embedded in a CpG island which is conserved from mice to humans [4]. The expression of this ubiquitous gene is strongly different in various tissues [5,6] and it is regulated by a number of stimuli [1,2,4,7,8]. Identification of the elements, present along the G6PD promoter, involved in the regulation of the gene expression is an important first step in elucidating the mechanisms of physiologic differential regulation of G6PD. In this respect, we decided to analyze the G6PD promoter in detail, starting from the previously reported observation that promoter sequences extending from −126 to +16 are sufficient to drive substantial amounts of G6PD transcription [9,10]. The analysis of this core promoter, in mammalian and *Drosophila* cell lines, identified several regulative sites, which are all strictly required for proper promoter functioning, thus revealing a very complex array of this promoter. We identified two GC motifs, a TATA box and an element that overlaps the transcription start site. This last element, which is essential for basal promoter activity, seems to require a mammalian specific factor in order to drive proper transcription.

*Corresponding author. Fax: (39) (81) 5936123.
E-mail: franze@iigbna.iigb.na.cnr.it

2. Materials and methods

2.1. Plasmids and transfections

The plasmid pGD126 (−126) contains the region of the G6PD gene promoter from −126 to +16 inserted upstream of the CAT gene in the vector pEmblCAT3 which was modified to remove a cryptic AP1 site. In the linker-scanning mutants (LS1–7) 18 bp of wild-type sequence (−126) have been replaced with an *Nde*I-*Xba*I-*Sal*I (NXS) polylinker. All mutants were analyzed by DNA sequencing. The Sp1 and Sp3 cDNAs mammalian expression plasmids (pCMV-Sp1 and pCMV-Sp3, Gal4-Sp1) and the *Drosophila* expression plasmid pPacSp1 have already been described [11–13] and were kindly given by L. Lania. The cells were cultured and transiently transfected as already described [9,14]. CAT activity was quantified using the Molecular Dynamics PhosphorImager system. (The ratio of the acetylated form to the total was normalized by comparison with the luciferase or the β-galactosidase activity.)

2.2. Electrophoretic mobility shift assays and DNase I protection

Gel shift assays were performed using standard methods. The DNA fragment, labeled using the Klenow enzyme, was incubated with HeLa nuclear extracts [15] for 30 min at room temperature. To carry out supershift assays, antibodies were added to the binding assay mixtures and incubated for 2 h on ice, before the addition of the radiolabeled probe. DNase I protection experiments were performed using the Footprinting kit from Pharmacia Biotech. Full length recombinant Sp1 and Sp3 proteins (*Escherichia coli*) were kindly given by L. Lania.

2.3. In vitro transcription

In vitro transcription experiments were performed using the HeLa Cell Transcription kit from Promega. Transcription products were analyzed by primer extension using a CAT primer, as described elsewhere [16].

3. Results

3.1. Analysis of G6PD linker scanning mutants

In order to identify the regulatory elements that are important for proper G6PD core promoter functioning, we have constructed a set of linker scanning mutants, each with 18 bp of the promoter sequence substituted with an unrelated sequence (Fig. 1). Each construct was transfected into three different human cell lines (HeLa, HepG2, WI38) and tested for CAT activity. Fig. 1 shows that the levels of expression of the mutants are comparable in the three cell lines. A strong reduction of promoter activity was observed with the mutants LS6, LS5 and LS3 in which putative Sp1 binding sites are mutated (box B and box C, Fig. 1). Moreover, alteration of sequences including the TATA box (mutant LS2), and the region surrounding the transcription start site (mutant LS1) results in a substantial reduction of CAT expression, suggesting that also these core promoter elements might play a prominent role in G6PD core promoter activity. In vitro transcription assay of wild-type and linker-scanning mutants, using nuclear protein extracts from HeLa cells, reinforced these re-

PII: S0014-5793(98)01259-9

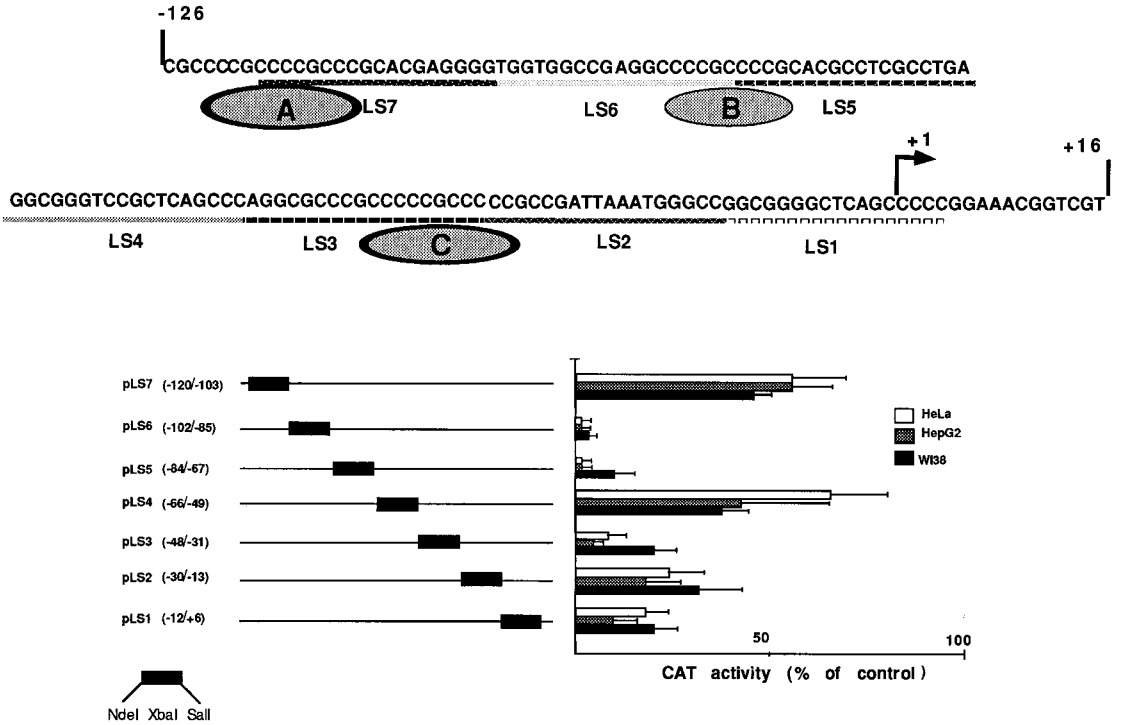

Fig. 1. Promoter activity of linker scanning G6PD mutants. Top, wild-type sequence of G6PD promoter. The underlined regions indicate nucleotides mutated in the LS1–7 constructs; the locations of the three GC boxes determined by sequence computer analysis (A, B, C) and the transcription start site are indicated. Below, diagram of linker scanning mutants and CAT activity in human cell lines (percentage of wild type). Data represent averages ± S.D. of at least three independent experiments.

sults (Fig. 2). Indeed, we found that in vitro transcription from the LS1 mutant mainly starts eight nucleotides downstream of the wild-type start site. The LS2 construct, in which the TATA box was mutated, showed transcription starting at multiple heterogeneous sites. These data suggest that a transcription initiator (Inr) element, located between −12 and +6, is required for efficient promoter activity and also that both TATA and Inr are required for the correct functioning of the promoter.

3.2. *Sp1 and Sp3 bind to multiple sites in the human G6PD promoter*

It has been suggested [17] that some proteins, which presumably belong to the Sp1 family, bind the two more distal GC boxes (A and B) of the G6PD promoter. However, that research did not distinguish among members of the Sp1 family that possess similar binding specificity. In this respect, we performed gel shift assays with a G6PD probe (−126/−57) containing the GC boxes A and B (Fig. 1), and nuclear extracts from HeLa cells. Three DNA-protein complexes were detected (Fig. 3, lane 2). The formation of all three complexes was specifically inhibited by competition with a 100-fold molar excess of cold oligonucleotide corresponding to the consensus Sp1 binding site (lane 1). Furthermore, pre-incubation of HeLa nuclear extract with Sp1 antibody resulted in the partial abrogation of the complexes α and β (lane 3), whilst

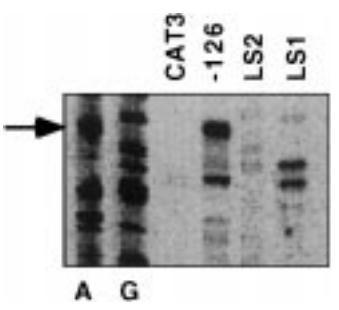

Fig. 2. In vitro analysis of G6PD transcription start site. In vitro transcription products from control vector (CAT3), wild-type G6PD promoter (−126), or mutants LS2 and LS1. The arrow on the left of the gel indicates the major product obtained with the wild type G6PD promoter (−126).

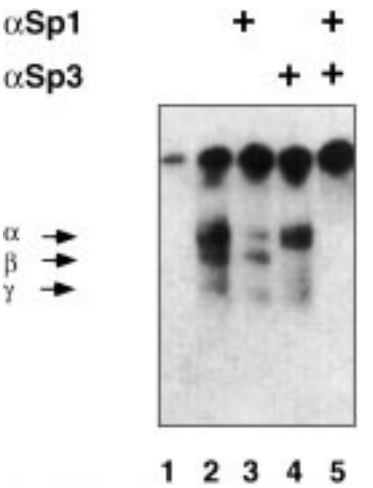

Fig. 3. Gel retardation assay of DNA sequence containing the GC boxes A and B.

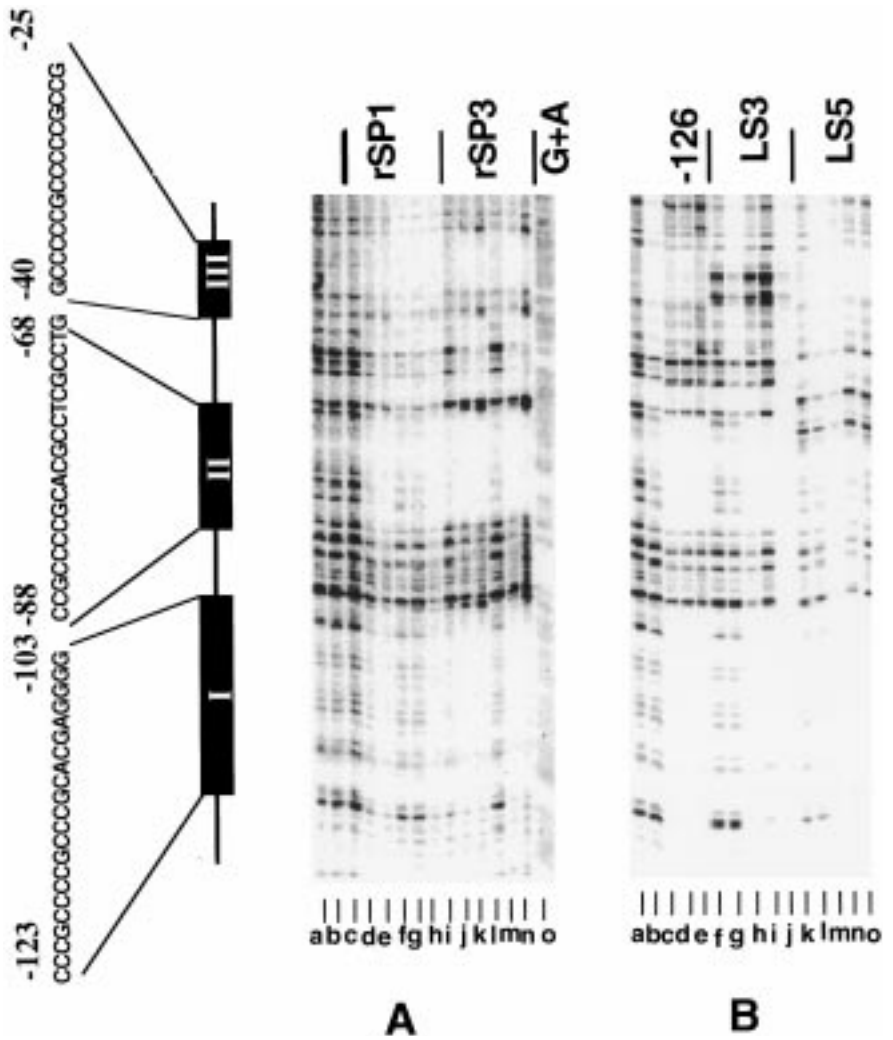

Fig. 4. DNAse I protection assays. A: Gel analysis of wild-type DNA with recombinant protein Sp1 and Sp3. Lanes a, b: naked DNA; lanes c–n: after reacting with increasing amounts (1, 3, 5 μg) of recombinant protein, DNA is incubated with 1 U (lanes c–e and i–k) or 2 U (lanes f–h and l–n) of DNase. Lane o: G+A sequence. B: Gel analysis of DNA from wild-type (−126) or the indicated linker scanning mutants with HeLa nuclear extract. Each group of five lanes is made of two naked DNA controls (lanes a, b; f, g; k, l) and three samples of HeLa-treated DNA digested with increasing amounts of DNase I (1, 2, 3 U) (lanes c–e; h–j; m–o). The regions protected are shown schematically by boxes; protected sequences are reported.

pre-incubation with anti-Sp3 antibody merely resulted in the suppression of complex γ (lane 4). The three complexes were completely eliminated by pre-incubating the nuclear extract with both anti-Sp1 and anti-Sp3 antibodies (Fig. 3, lane 5). Furthermore, we performed DNase protection assays (Fig. 4A,B) to precisely define the regions of DNA-protein interaction. The same pattern of DNase I protection was observed with the HeLa extract or with purified Sp1 and Sp3 proteins. Three protected regions (I, II and III) were detected: region I (−126 to −108) corresponds to the sequences mutated in LS7 (GC box A, Fig. 1), region II (−91 to −73) corresponds to LS6–5 (GC box B), and region III (−45 to −25) corresponds to LS3 (GC box C). As expected, when the mutated sequences were used as probes in the footprinting experiments, box I was not protected in LS7, nor box II in LS5 and LS6 or box III in LS3 (Fig. 4B and data not shown). No protection was detected around the transcription start site.

3.3. Sp1- and Sp3-mediated regulation of the G6PD promoter

The results reported in Figs. 3 and 4 demonstrate that both

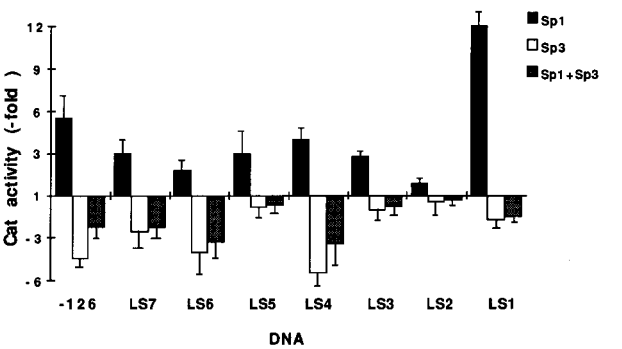

Fig. 5. Sp1 dependence of G6PD promoter mutants in HeLa cells. Linker scanning mutants (LS1–7), or G6PD wild-type (−126) constructs (10 μg) are co-transfected in HeLa cells in the presence of a control or a Sp1 (10 μg) or/and Sp3 expressing plasmid (15 μg). Fold induction was calculated as the activity of each reporter in the presence of the indicated amount of effector (Sp1 or Sp3) divided by the activity of the reporter without effector. Data represent averages ± S.D. of at least three independent experiments.

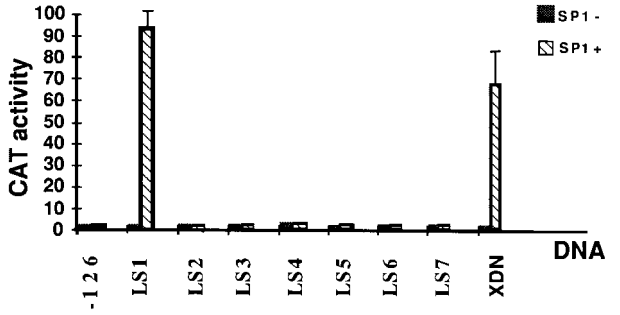

Fig. 6. Sp1 dependence of G6PD promoter mutants in *Drosophila* cells. Linker scanning mutants, LS1–7 (2 μg), or G6PD wild-type, −126 (2 μg) or myc, XDN (2 μg) constructs are co-transfected in *Drosophila* SL2 cells in the presence of a control or a Sp1 expressing plasmid (1.5 μg). Data represent averages ± S.D. of at least three independent experiments.

Sp1 and Sp3 proteins bind the G6PD promoter at regions whose integrity is crucial for proper promoter activity. Since Sp1 and Sp3 have been shown to activate or repress GC box-containing promoters in a promoter-specific manner [11,12,18–22], we wanted to determine the transacting ability of these factors on the G6PD promoter. The analysis of the constructs described in Fig. 1 indicates that all the constructs are activated by Sp1 and repressed by Sp3 (Fig. 5), albeit to different degrees, suggesting that the boxes A, B and C can complement each other and that no exclusive region for the

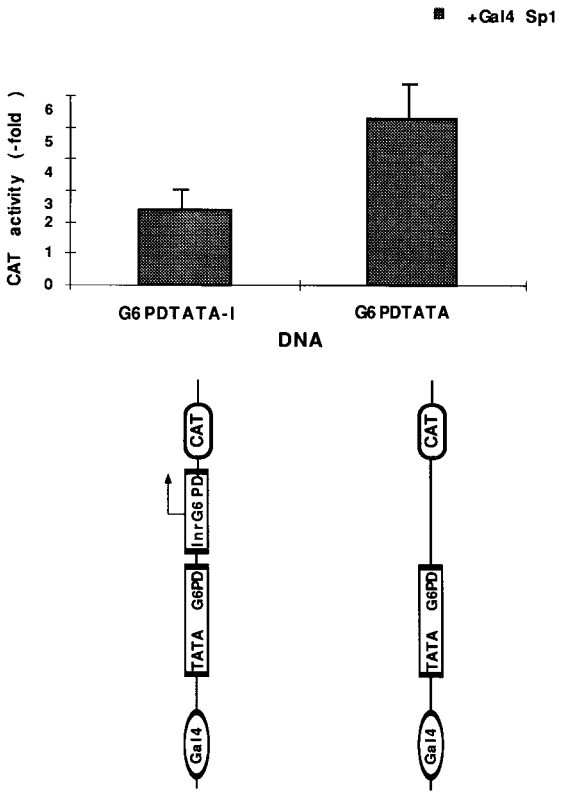

Fig. 7. Activation by Sp1 targeted to G6PD promoter. Left, diagram of G6PDTATA-I and G6PDTATA constructs. Right, the two constructs (12 μg) are co-transfected in HeLa cells with the Gal4-Sp1 expressing plasmid (12 μg). Fold induction was calculated as indicated in the legend of Fig. 5. Data represent averages ± S.D. of at least three independent experiments.

binding of the Sp1 family proteins exists in the G6PD promoter. As regards the relative transcription efficiency of the various mutants, we noted surprisingly that LS1 is significantly stimulated by Sp1 (fold induction = 11.8), much more than the wild-type construct (fold induction = 5.7).

3.4. *Sp1-mediated regulation of the G6PD promoter in Drosophila SL2 cells*

In order to define the regulation of the G6PD promoter by the Sp1 family members in a more controlled manner, we expressed our constructs in *Drosophila* cells, which are devoid of any Sp1 and Sp1-like DNA binding activity [23,24]. Surprisingly, we did not find any detectable transcriptional activity of G6PD core promoter either in the presence or in the absence of the Sp1 co-transfected expression vector. Under the same conditions, the transcription of a bona fide Sp1-dependent promoter (human c-*myc*) was stimulated ∼40 times by Sp1 (XDN, Fig. 6). The analysis of our promoter mutants in *Drosophila* cells showed that only the LS1 reporter was activated by Sp1 (Fig. 6). This construct was also very efficiently activated by Sp1 in HeLa cells (Fig. 5). Furthermore, in order to exclude the possibility that the heterologous 18 bp sequences present in LS1 may contain a fortuitous positive transcription element, we constructed a G6PD promoter mutant that contains different sequences around the transcriptional start site. The LS8 construct is similar to LS1, 18 bp of wild-type being substituted by a different sequence. This promoter construct, when co-transfected into HeLa and *Drosophila* SL2 cells along with the Sp1 expression plasmid, shows the same extent of Sp1-mediated activation as was obtained with the mutant LS1 (data not shown). These data suggest that the sequence around the start site plays a critical role in the control of G6PD efficient transcription. Furthermore, as Sp1 stimulates the transcription of this promoter only in the absence of the natural region around the start site, we suggest that an essential factor missing in *Dro-*

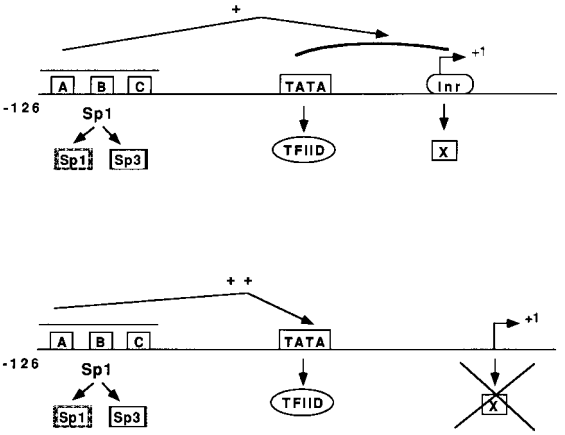

Fig. 8. Proposed model of transcriptional regulation of the G6PD promoter. The binding proteins are shown below, the binding sites are shown on the line, and the interactive effects of the binding proteins are indicated above. The Sp1 and Sp3 proteins can bind to sites A, B, C. Top, wild-type promoter: the Inr site has a basal functional activity and together with the TATA box is essential for promoter activity. Sp1 factors can act on the wild-type G6PD promoter only in the presence of a mammalian-specific factor(s) (X). Below, LS1 mutant: in the absence of the Inr element the promoter is strongly dependent on Sp1, and the Sp1 factors interact directly with the TATA box.

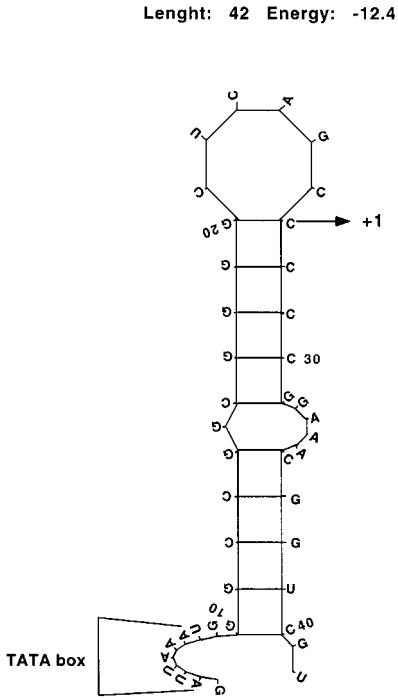

Fig. 9. Prediction of secondary structure. The sequence of the G6PD promoter containing the TATA box and the initiator element (−26/+16) has been analyzed with the software 'Fold' of the Genetics Computer Group (GCG), Madison, WI.

sophila cells is required to obtain full promoter activity in the wild-type configuration.

3.5. Transcriptional activation by Sp1 targeted to G6PD promoter

A bias in the interpretation of the data from co-transfection experiments in HeLa cells could be due to the constitutive expression of both endogenous Sp1 and Sp3 proteins [11,12]; therefore, we extended our study to HeLa cells using gene fusion experiments on reporter constructs, in which Sp1 boxes were substituted by the GAL4 binding site (Fig. 7). The expression vector Gal4-Sp1 [25] was transfected in HeLa cells and its activity assayed on the reporter constructs. The expression vector Gal4-Sp1 contains the Sp1 transcriptional activation domain fused to the yeast Gal4-DNA binding domain. Two different reporter plasmids were used; the first (G6PDTATA-I) contains five Gal4 binding sites in front of the G6PD core promoter sequence ranging from −34 to +14; the second (G6PDTATA) is isogenic to the first, except for the fact that the sequence from −12 to +6 is substituted by the corresponding sequence in the mutant LS1. The results (Fig. 7) clearly show that the G6PDTATA construct is more responsive to Gal4-Sp1 than the wild-type construct, confirming the data already obtained both in mammalian (Fig. 5) and *Drosophila* cells (Fig. 6), which show that the G6PD core promoter is more responsive to Sp1 when the sequence around the start site is mutated.

4. Discussion

In this paper, we identify several DNA elements in the core promoter of the G6PD gene which are essential for its expression. Transfection experiments in mammalian cells (Fig. 1) indicate a prominent role for the sequences corresponding to the B box (mutants LS6, LS5), the C box (mutant LS3), the TATA box (mutant LS2) and an element surrounding the transcription start site (mutant LS1). This last element acts as an initiator, which defines the site and rate of transcription together with a non-canonical TATA box (Fig. 2). Moreover, using electrophoretic mobility shift assays and DNA footprinting analysis, we demonstrated that Sp1 and Sp3 proteins are both capable of binding to promoter DNA sequences spanning the GC motifs, and we have precisely defined the region of DNA-protein interactions. We demonstrated that, in mammalian cells, Sp1 acts as a transcriptional activator while Sp3 acts as a repressor (Fig. 5). These results suggest that changes in the cellular ratio of Sp1 and Sp3 could affect G6PD expression in specific physiological contexts; the binding of Sp1 and Sp3 to the sites depends on the transcription factor milieu that may determine the variability in the expression of the G6PD gene in various tissues. We also observed that the mutant LS1 is activated much better in mammalian cells by Sp1 than the wild-type construct (Figs. 5 and 7) and that the G6PD promoter could not be activated in *Drosophila* cells by Sp1 (Fig. 6); activation of the G6PD promoter by Sp1 in *Drosophila* cells was only possible in the presence of mutations at the origin of transcription (mutant LS1). These results suggest that Sp1 is not sufficient to activate the wild-type promoter; in order to function the G6PD promoter may require another essential factor, which is absent in *Drosophila* cells but is present in HeLa. This factor is bound, either directly or indirectly, to the element around the transcription start site and is fundamental for the interaction with the transcriptional machinery (Fig. 8). An alternative transcription complex, which is less efficient (Fig. 1) but strongly dependent on Sp1 (Figs. 5 and 6), may be active in the case of the LS1 mutant; in this case the promoter is essentially under the control of the TATA box because the element around the transcription start site is mutated. The amount of endogenous Sp1, present in mammalian cells, fails to stimulate efficient transcription and/or to overcome Sp3 repression in the particular LS1 promoter conformation (Fig. 1), but a high level of Sp1 can compensate this effect (Fig. 5).

In conclusion, the results reported here clearly indicate that the G6PD promoter is more complex than previously assumed and reinforce the model suggesting that transcriptional regulatory proteins may prefer to modulate transcription through core promoters with specific structures. As shown in Fig. 4, in vitro protection experiments did not reveal any protein bound to the sequences around the start site. The lack of a footprint on this region may be due to a specific secondary structure assumed by this promoter in vivo. In fact, around the transcription start site we can observe an inverted repeat (−10/−6, +1/+5), which can loop as shown in Fig. 9. Such a looping interaction may alter the configuration of the promoter and thereby facilitate its interactions with some specific factor(s). In order to explore the possibility that one or more factors are bound to the element overlapping the transcriptional start site only in the specific secondary structure assumed by the G6PD promoter in vivo, we are currently setting up in vivo footprinting experiments.

Acknowledgements: We thank L. Lania for donating the Sp1 and Sp3 expressing plasmids, and the recombinants proteins. We also thank L. Lania and E. Avvedimento for the critical review of the manuscript

and helpful advice. Furthermore, we thank M. Terracciano for her skilful technical assistance.

References

[1] Luzzatto, L. and Mehta, A. (1988) The Metabolic Basis of Inherited Disease, pp. 2237–2265.
[2] Ursini, M.V., Parrella, A., Rosa, G. and Martini, G. (1997) Biochem. J. 323, 801–806.
[3] Pandolfi, P.P., Sonati, F., Rivi, R., Mason, P., Grosveld, F. and Luzzatto, L. (1995) EMBO J. 14, 5209–5215.
[4] Toniolo, D., Filippi, M., Dono, R., Lettieri, T. and Martini, G. (1991) Gene 102, 197–203.
[5] Battistuzzi, G., D'Urso, M., Toniolo, D., Persico, M.G. and Luzzatto, L. (1985) Proc. Natl. Acad. Sci. USA 82, 1465–1469.
[6] Corcoran, C.M., Fraser, P., Martini, G., Luzzatto, L. and Mason, P.J. (1996) Gene 173, 241–246.
[7] Kletzien, R.F., Harris, P.K. and Foellmi, L.A. (1994) FASEB J. 8, 174–181.
[8] Martini, G. and Ursini, M.V. (1996) BioEssays 18, 1–8.
[9] Ursini, M.V., Scalera, L. and Martini, G. (1990) Biochem. Biophys. Res. Commun. 170, 1203–1209.
[10] Ursini, M.V., Lettieri, T., Braddock, M. and Martini, G. (1993) Virology 196, 338–343.
[11] Hagen, G., Müller, S., Beato, M. and Suske, G. (1994) EMBO J. 13, 3843–3851.
[12] Majello, B., De Luca, P., Hagen, G., Suske, G. and Lania, L. (1994) Nucleic Acids Res. 22, 4914–4921.
[13] Emami, K.H., Navarre, W.W. and Smale, S.T. (1995) Mol. Cell. Biol. 15, 5906–5916.
[14] Di Nocera, P.P. and Dawid, I.B. (1983) Proc. Natl. Acad. Sci. USA 92, 1955–1959.
[15] Lee, K.A.W., Bindereif, A. and Green, M.R. (1988) Gene Anal. Tech. 5, 22–31.
[16] Strazzullo, M., Majello, B., Lania, L. and La Mantia, G. (1994) Virology 200, 686–695.
[17] Philippe, M., Larondelle, Y., Lemaigre, F., Mariamé, B., Delhez, H., Mason, P., Luzzatto, L. and Rousseau, G.G. (1994) Eur. J. Biochem. 226, 377–384.
[18] Hagen, G., Dennig, J., Preib, A., Beato, M. and Suske, G. (1995) J. Biol. Chem. 270, 24989–24994.
[19] Majello, B., De Luca, P., Suske, G. and Lania, L. (1995) Oncogene 10, 1841–1848.
[20] De Luca, P., Majello, B. and Lania, L. (1996) J. Biol. Chem. 271, 8533–8536.
[21] Li, R., Hodny, Z., Lucianova, K., Barath, P. and Nelson, B.D. (1996) J. Biol. Chem. 271, 18925–18930.
[22] Majello, B., De Luca, P. and Lania, L. (1997) J. Biol. Chem. 272, 4021–4026.
[23] Courey, A.J. and Tijan, R. (1988) Cell 55, 887–898.
[24] Courey, A.J., Holtzaman, D.A., Jackson, S.P. and Tijan, R. (1989) Cell 59, 827–836.
[25] Gill, G., Pascal, E., Tseng, Z.H. and Tijan, R. (1994) Proc. Natl. Acad. Sci. USA 91, 192–196.