AD_____

Award Number: DAMD17-01-1-0099

TITLE: Genetic Variation in the HSD3B2 Gene and Prostate Cancer

PRINCIPAL INVESTIGATOR: Juergen K. V. Reichardt, Ph.D.

CONTRACTING ORGANIZATION: University of Southern California
Los Angeles, California 90033

REPORT DATE: July 2004

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2004	Final (1 Jul 2001 - 30 Jun 2004)

4. TITLE AND SUBTITLE
Genetic Variation in the HSD3B2 Gene and Prostate Cancer

5. FUNDING NUMBERS
DAMD17-01-1-0099

6. AUTHOR(S)
Juergen K. V. Reichardt, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Southern California
Los Angeles, California 90033

E-Mail: Reichard@usc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Prostate cancer is the most common cancer in American men. It is also characterized by a substantial racial/ethnic variation in risk: highest in African-American men, lowest in Asian men and intermediate in Caucasian and Latino men. We propose to investigate genetic variants of genes involved in the regulation of prostatic growth and particularly in androgen metabolism, particularly the HSD3B2 gene which encodes the type II b-hydoxysteroid dehydrogenase. Our final progress reported is highlighted by the following four findings. First, our data indicate that the locus under investigation is highly polymorphic in constitutional DNA and mutated in tumor (i.e. somatic) DNA. Second, our population-based investigations are complete. Third, we will complete the biochemical analyses soon. Finally, we note that we have also investigated transcription of the HSD3B2 gene and found that YY1 binding is important.

14. SUBJECT TERMS
No Subject Terms Provided.

15. NUMBER OF PAGES
61

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Abstract

Prostate cancer is the most common cancer in American men. It is also characterized by a substantial racial/ethnic variation in risk: highest in African-American men, lowest in Asian men and intermediate in Caucasian and Latino men. We propose to investigate genetic variants of genes involved in the regulation of prostatic growth and particularly in androgen metabolism, particularly the HSD3B2 gene which encodes the type II β-hydoxysteroid dehydrogenase. Our final progress reported is highlighted by the following four findings. First, our data indicate that the locus under investigation is highly polymorphic in constitutional DNA and mutated in tumor (i.e. somatic) DNA. Second, our population-based investigations are complete. Third, we will complete the biochemical analyses soon. Finally, we note that we have also investigated transcription of the HSD3B2 gene and found that YY1 binding is important.

Overview

This proposal is part of a research program aimed at identifying genes involved in the predisposition to and progression of prostate cancer among various racial/ethnic groups in the US. Prostate cancer will be diagnosed –according to the ACS- in 230,110 men in the US in the year 2004 alone. Some 29,900 individuals will die of this disease this year. Prostate cancer is characterized by substantial racial/ethnic variation in risk: highest in African-American men, lowest in Asian men and intermediate in Caucasian and Latino men. We proposed to investigate as our central hypothesis that genetic variants of genes involved in the regulation of prostatic growth and particularly in androgen metabolism by themselves and in combination significantly contribute to prostate cancer risk and progression. Specifically, we proposed to examine the hypothesis that DNA sequence variations in the type II 3β-hydoxysteroid dehydrogenase (HSD3B2) gene contribute substantially to the risk of prostate cancer particularly across racial/ethnic lines. The "candidate gene", HSD3B2, was chosen because the reaction substrate [i.e. dihydrotestosterone (DHT)] of the enzyme encoded by this gene modulates directly cell division in the prostate. Epidemiologic evidence suggests that variation in DHT levels play an important role in risk of prostate cancer. Thus, 3β-hydoxysteroid dehydrogenase activity encoded by HSD3B2 variant alleles may be important in regulating intraprostatic DHT steady state levels by controlling its degradation. This candidate gene encodes the enzyme that initiates the irreversible inactivation of DHT.

Specifically, in this project we proposed to test, using a case-control study approach within a multi-ethnic cohort study design, the association between prostate cancer risk and its progression and HSD3B2 allelic variants among four major racial-ethnic groups. Our original three interrelated specific aims were:

- To identify all allelic variants in the HSD3B2 locus by sequencing 200 men from four racial/ethnic groups (African-American, Japanese-American, Latino and Americans of European ancestry (Caucasian) men).
- To determine the relationship between the HSD3B2 gene and prostate cancer by genotyping polymorphic DNA markers in the HSD3B2 gene in up to 800 men with prostate cancer and controls from four racial/ethnic groups who are at very different risks of prostate cancer.
- To determine the *in vitro* biochemical properties of HSD3B2 variants identified in specific aims 1 and 2.

Progress Report

Pursuant to specific aim 1, we have sequenced in constitutional ("germline") DNA of 120 men with the following make-up: 30 from each of the four racial/ethnic groups (African-American, Asian-(Japanese-)American, Caucasian and Latino, the entire HSD3B2 gene. Half are prostate cancer cases, half are controls. Our sequencing has identified 17 additional polymorphisms that are highlighted above the gene in Fig. 1.

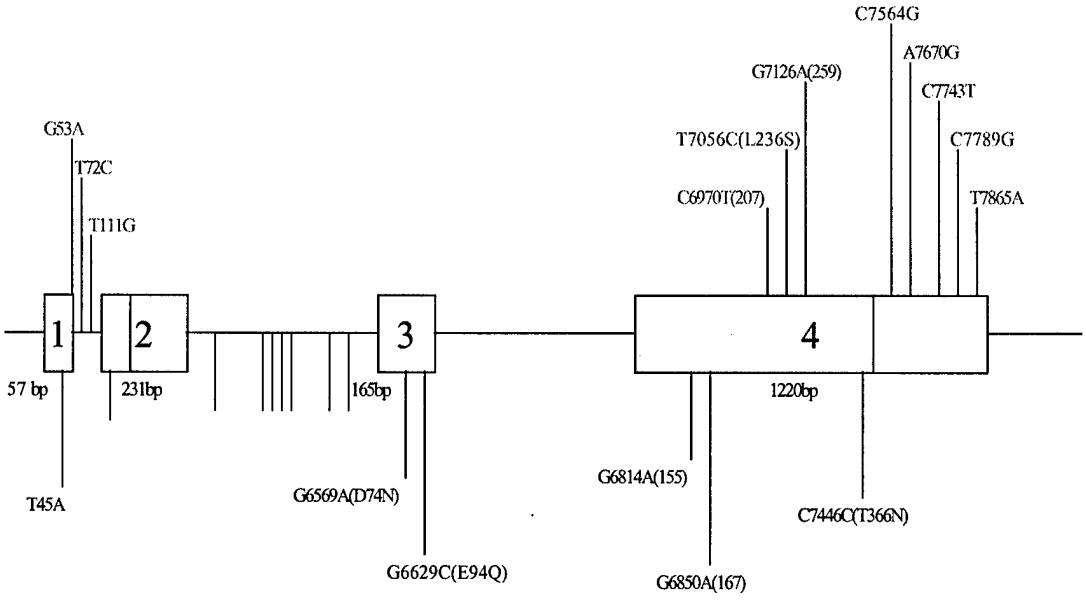


Figure 1: Newly discovered constitutional ("germline") SNPs in the HSD3B2 gene.

In short, the HSD3B2 gene in humans is highly polymorphic and our work has uncovered significant unreported additional genetic variation. This specific aim is complete as of this year.

We also identified a series of somatic mutations in tumor tissue in men with prostate cancer (Fig. 2). These *de novo* mutations are not present in constitutional DNA of the same individual. These mutations may be important in tumor progression.

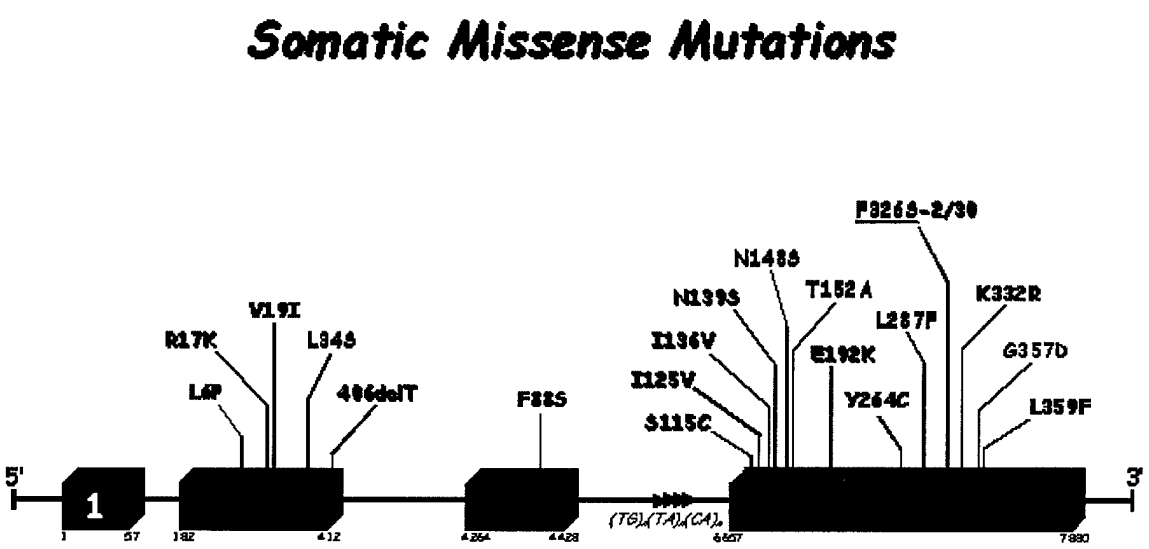


Figure 2: Somatic mutations in the HSDS3B2 gene.

In order to advance specific aim 2 we have begun genotyping men for five SNPs iin the HSD3B2 gene: the D74N, E94Q, L236S and T366N constitutional DNA polymorphism that change amino acids and the T1362G SNP (cf. Fig. 1). Sample genotyping using a multiplex approach (MAPA: Multiplex Automated Primer Extension Analysis; developed in this lab) on an ABI3100 instrument is shown in Fig. 3. The genotyping is complete but the data analysis unfortunately identified no significant associations with prostate cancer (Table 1). However, small effects cannot be excluded at this point(due to the rarity of most SNPs which are generally below the 5 % frequency mark (Table 1).

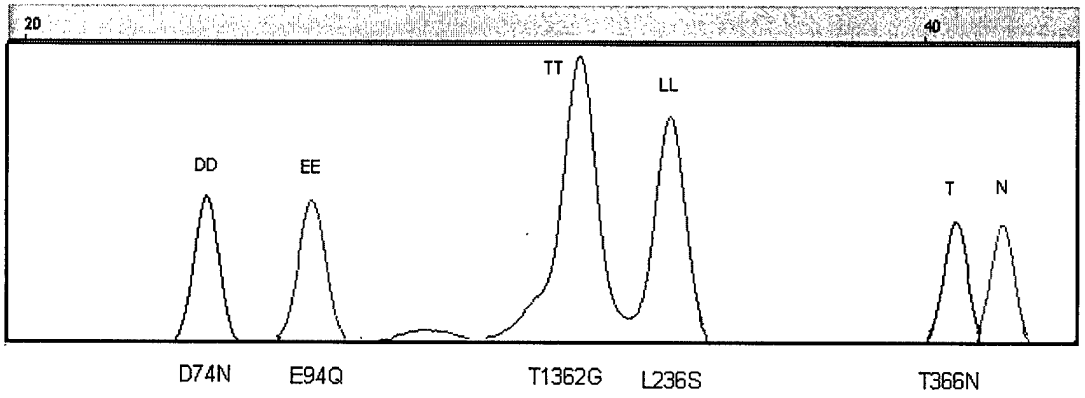


Figure 3: Sample MAPA genotyping of the HSD3B2 gene on an ABI3100 automated DNA sequencer. Only the last, T366N, SNP is polymorphic (heterozygous) in this sample.

Table 1
MAPA Genotyping of Prostate Cancer Cases and Controls

HSD3B2 SNPs by case-control status.

D74N	DD	DN
Controls	342	0
Cases	367	2

E94Q	EE	EQ
Controls	340	2
Cases	366	3

L236S	LL	LS
Controls	332	10
Cases	360	9

T1562G	TT	TG	GG
Controls	325	16	1
Cases	357	12	0

T366N	TT
Controls	342
Cases	369

We have also made progress in specific aim 3 on the *in vitro* biochemistry on several fronts. We have reconstructed all five missense SNPs in constitutional DNA (cf. Fig. 1) and have begun to assay them along the normal cDNA. We expect to make significant progress in the final no cost extension year in this area.

Finally, we have analyzed transcription of the HSD3B2 gene (Foti and Reichardt, 2004). We found that YY1 binding in the first intron is important for proper transcription. The preprint of this paper is attached.

Benchmarks 2004:

Specific aim 1: Complete.
Specific aim 2: Complete.
Specific aim 3: Transcriptional studies are complete. Mutagenesis has also been completed. Biochemical assays have been initiated and are to proceed into the final no cost extension year.

Key Research Accomplishments

- 18 SNPs identified in constitutional DNA in the HSD3B2 gene.
- 19 somatic ("*de novo*") mutations in the HSD3B2 gene identified.

- MAPA genotyping for 5 SNPs in the HSD3B2 gene.
- 5 SNPs are reconstructed in the HSD3B2 cDNA for *in vitro* analyses.
- HSD3B2 transcription requires YY1.

Reportable Outcomes

One paper in press in *J. Molec. Endocrin.* (attached).

Conclusions

Significant progress toward all three specific aims was made in the past three years period. Furthermore, we plan to complete the third specific aim in the coming 12 months of no cost extension.

Our research may result in better presymtomatic diagnosis of prostate cancer and will lead to a better fundamental understanding of the HSD3B2 gene and its enzyme.

YY1 Binding Within the Human HSD3B2 Gene Intron 1 is required for Maximal Basal Promoter Activity: Identification of YY1 as the 3 1-A Factor

Dolly M. Foti #¶, and Juergen K. Reichardt #§

#Department of Biochemistry and Molecular Biology, and §Department of Preventive Medicine, Institute for Genetic Medicine, USC Keck School of Medicine, Los Angeles, California 90033, USA

Short title: Enhancement of the HSD3B2 gene basal promoter activity by YY1

¶ Corresponding author:
Institute for Genetic Medicine
USC Keck School of Medicine
2250 Alcazar Street, CSA 210
Los Angeles, California, USA
Phone: (323) 442-3914
Fax: (323) 442-2764
E-mail: foti@usc.edu

Reprints should be asked to:
Juergen Reichardt, PhD
USC Keck School of Medicine
2250 Alcazar Street, CSA 210
Los Angeles, California, USA
Phone: (323) 442-1529
Fax: (323) 442-2764
E-mail: reichard@usc.edu

Key Words: HSD3B2, HSD3B1, 3 1-A factor, intronic binding of YY1, transcriptional regulation, basal activity

ABSTRACT

The oxidation and isomerization of 3 -hydroxy-5-ene steroids into keto-4-ene steroids, a pivotal step in the synthesis of all hormonal steroids, is catalyzed by several isoforms of 3 -Hydroxysteroid dehydrogenase isoforms. In humans, two highly homologous isoforms exist, type I expressed by the HSD3B1 gene in peripheral tissues, and type II expressed by the HSD3B2 gene in steroidogenic organs. Previously, it was shown that the HSD3B1 gene 3 1-A element, encompassing 24 nucleotides of intron 1 not perfectly conserved between the two genes and overlapping with a conserved TG box, contributes to maximal basal promoter activity by binding the ubiquitous and unidentified 3 1-A transcription factor. In this study for the first time we report that similarly, the HSD3B2 gene intron 1 is required for maximal basal promoter activity in reporter gene analyses, as lack of intron 1 result in a 5 to 10 fold reduction in promoter activity. Mutational analysis in gel shift assays revealed that the 3 1-A factor binds both the HSD3B2 and HSD3B1 gene intron 1 by requiring only 7 nucleotides of a conserved segment within the 3 1-A element. By competition analysis and use of anti-YY1 antibody in both gel shift and western experiments, we identified the 3 1-A protein as the ubiquitous transcription factor YY1. In addition, we have characterized another similar YY1 binding site differently located in respect of the 3 1-A element in both genes. Deletion analysis in transient transfections experiements have shown that contrarily to as previously shown for the HSD3B1 gene, YY1 binding to the type II 3 1-A element only results in a marginal reduction of basal promoter activity. Instead, the second and inverted YY1 site, placed 35 base pairs downstream the 3 1-A element, is a stronger activator, as its deletion alone results in a 50% decrease of the HSD3B2 gene basal promoter activity. Complete abrogation of YY1 binding within type II intron 1 resulted in a gene reporter activity identical to a reporter construct lacking the whole intron 1. These results designate YY1 as the factor responsible for the intron 1 mediated boost of the HSD3B2 gene basal promoter activity. Similarities and dissimilarities between YY1 binding within the HSD3B1 and HSD3B2 gene intron 1 are discussed involving the conserved intron 1 TG box, that suggests different

mechanisms are implicated in the YY1 mediated stimulation of these two genes basal promoter transcription.

INTRODUCTION

The enzyme 3 -Hydroxysteroid dehydrogenase (3 -HSD) catalyzes a fundamental step in the synthesis of all hormonal steroids in many tissues and organs (Labrie *et al.*), which play a crucial role in the differentiation, development, growth, and physiological function of most human tissues. It is also one of the two enzymes that start the degradation of dihydrotestosterone (DHT), the most powerful androgen (Bartsch *et al.*), thus affecting the availability of this hormone in the regulation of androgen responsive genes. Two isoforms are known in humans, type I and type II 3 -HSD, encoded by the HSD3B1 and HSD3B2 gene, respectively (Luu-The *et al.*, Lachance *et al.* 1990, Rheaume *et al.*, Lachance *et al.* 1991). The two genes are closely linked on the short arm of chromosome 1 (Morrison *et al.*, Morissette *et al.*), and both consist of 4 exons and 3 introns that have similar length and a high degree of homology (Lachance *et al.* 1991) (Fig.1). Type I and type II 3 -HSD enzymes are differentially expressed, being type II activity mostly restricted to steroidogenic organs such as ovary, testis and adrenals (Rheaume *et al.*, Lachance *et al.* 1991), while type I is found in many peripheral tissues, such as placenta, skin, mammary gland, and others (Luu-The *et al.*, Lachance *et al.* 1990, Lachance *et al.* 1991, and references therein).

The importance of type II 3 -HSD activity in steroidogenesis is attested by the many gene mutations resulting in Congenital Adrenal Hyperplasia (CAH), a term that encompasses several recessive autosomal disorders that share the complete or partial deficiency of enzymes involved in corticosteroids synthesis (White 2001, Levine 2000). In both males and females decreased mineralcorticoid secretion results in varying degrees of salt wasting, from lethal to mild phenotypes, and may lead to female masculinization during fetal development. CAH due to 3 HSD type II deficiency may lead to a decrease in all classes of steroid hormones including sex hormones, causing pseudohermaphroditism in males due to gonadal hypospadia (Simard *et al.* 1993, Mendonca *et al.* 1994, Simard *et al.* 2002), and hirsutism, premature puparche and/or infertility in females (Eldar-Geva *et al.* 1990, Mendonca *et al.* 1994, Marui *et al.* 2000,

Pang *et al.* 2002). Type II 3 -HSD enzyme has more recently acquired an active role in prostate cancer research (Ross *et al.* 1998). Androgens are important in the development and maintenance of the prostatic gland (Marker *et al.* 2003), and it is well documented that they play an important role in the etiology of prostate cancer (Ross *et al.* 1996, Bosland 2000). Circulating levels of testosterone and DHT are considered indicative of the risk of getting the disease (Gann *et al.* 1996). Therefore, the ability of the 3 -HSD activity to degrade DHT has gained interest in the potential mutations or functional sequence variants that may affect the overall enzyme performance, either by altering the enzyme catalytic properties or its expression levels. Interestingly, a study from Chang *et al.* (2002) has linked both the HSD3B1 and HSD3B2 gene with the susceptibility of sporadic and hereditary prostate cancer.

The recent discovery of a binding site for steroidogenic factor 1 (SF-1) at nucleotide –64 to –56 of the HSD3B2 gene promoter (Leers-Sucheta *et al.* 1997), which is not conserved in the HSD3B1 gene, might explain the restricted expression of the HSD3B2 gene to steroidogenic organs. SF-1 had indeed earlier been identified as a tissue-specific transcription factor that regulates all the steroidogenic *P-450* genes in the adrenal cortex and gonads (Morohashi *et al.* 1992, Lala *et al.* 1992). Interestingly, a study by Guerin *et al.* (1995) identified the 3 1-A element, a transcriptional activator located within the HSD3B1 gene intron 1 (Fig. 1) that is required for reaching maximal promoter activity in transient transfection assays This region, encompassing 34 nucleotides protected in DNAse footprinting assay, was shown to strongly bind an unidentified ubiquitously expressed nuclear protein named 3 1-A factor, whose molecular weight was estimated as 57 kDa by UV-crosslinking experiments. Mutational analysis revealed that the 4 guanines that lay at the 3' end of the protected region, shared with an overlapping TG box (Fig. 1), are essential for binding of the 3 1-A factor to the 3 1-A element, and their mutation resulted in a 3 to 6 fold lower expression of a reporter gene. As Sp1 was found to bind the TG box very weakly, the 3 1-A protein was attributed a transactivator role. Stressing the importance of the 3 1-A element as a transcriptional activator, Guerin and collaborators also showed that a double-stranded oligomer bearing the DNA sequence of the HSD3B1 gene 3 1-A element cloned in front of the mouse p12 gene promoter

was able to stimulate transcription of the CAT reporter gene from to 2 to 5 fold in human chorioncarcinoma JEG-3 and human adrenal cortex adenocarcinoma SW-13 cells, respectively.

Because of the importance of the type II 3 -HSD activity in steroidogenesis, we were prompted to investigate whether or not the HSD3B2 gene intron 1 has any functional relevance, as in the case of the HSD3B1 gene. We here report that intron 1 is required by the upstream promoter of the HSD3B2 gene to achieve the maximal level of transcription, as deletion analysis in transient transfection assays in several cell lines revealed that removal of intron 1 results in 5 to 10 fold reduction of luciferase reporter activity. Despite mismatches in the region, we have ascertained the presence of a 3 1-A element in the HSD3B2 gene intron 1. Furthermore, we have identified the 3 1-A protein to be the multifactorial transcription factor YY1 (Shi *et al.* 1997), and demonstrate that, in addition to the 3 1-A element, YY1 also recognizes a second binding site within intron 1, whose position is not conserved between the two genes. Through mutational analysis and co-expression of YY1 in transfection experiments, we have demonstrated that differently from the HSD3B1 gene, the type II 3 1-A element alone has little effect as a transcriptional activator. Instead, the deletion of the second YY1 binding site, located in reverse orientation about 35 base pairs downstream the 3 1-A element, reduces the HSD3B2 gene basal promoter activity to half of its full potential. In addition, we provide evidence that Sp3 binds the TG box in the intron 1 of both genes and that the binding efficiency is much higher for the type I TG box. We found the second YY1 binding site in the HSD3B1 gene to be positioned at the 3' end of the TG box. In "*in vitro*" binding conditions, the presence of the weaker YY1 site in such a location appears to favor Sp3 competition for the TG box, thus explaining the stronger Sp3 binding to type I intron 1. Although we have not investigated the functional relevance of those findings for the HSD3B1 gene, our results point to a role for YY1 in the basal expression of 3 HSD activity, and suggest that different mechanisms may be involved in the maintenance the HSD3B1 and HSD3B2 gene basal activity under experimental conditions, as discussed.

MATERIALS AND METHODS

Plasmid constructions

The human HSD3B2 5'-flanking and 5'-UTR (from –1265 to +193) was amplified by PCR from human genomic DNA isolated from blood using the forward primer 5'-CCGACGCGTTAATAAACATTTAAGCCAATAATAAAA-3' and the reverse primer 5'-GGGCTCGAGACCCAGAAGAGGGCTAAAAAC-3'. The amplification product was digested with *Mlu 1* and *Xho 1* and cloned in *pGL3*Basic (Promega) similarly digested, obtaining *pGL3*-HSD2(-1253/+193). To obtain the reporter plasmids used in the study, this construct was digested with *Sac 1* and the resulting plasmid re-ligated to generate *pGL3*-HSD2(-517/+193), or digested with *Mlu 1* and *PstI I*, filled in with Klenow at the 5'- end and re-ligated to generate *pGL3*-HSD2(-217/+193). Plasmid *pGL3*-HSD2(-517/+23) was obtained by digesting *pGL3*-HSD2(-517/+193) with *Drd 1* and *Xho 1*, filling in with Klenow and re-ligating, to generate a similar reporter construct without the intron 1 sequence. The insert nucleotide sequence was verified by automated DNA sequencing using the forward RV3 primer 5'-CTAGCAAAATAGGCTGTCCC –3' and reverse GL2 primer 5'- CTTTATGTTTTTGGCGTCTTCC –3' from Promega, which respectively match upstream and downstream to the *pGL3*Basic vector multiple cloning site.

Site-directed mutagenesis

Mutations were introduced into the reporter plasmids using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) following the manufacturer's instructions and using the same mutant oligomers that are represented in the figures that accompany this paper. Mutations were confirmed by DNA automated sequencing (ABI PRISM 3100 Genetic Analyzer, Applied Biosystems). Each sequence-confirmed mutant insert was released with the appropriate cloning restriction endonucleases, and inserted into similarly digested *pGL3*Basic, to ensure the plasmid used in transfections did not carry any PCR induced nucleotide substitutions within the vector backbone.

Gel shift assays

Crude nuclear extracts were prepared as previously described by Dent and Latchman (1993). Oligonucleotides for radiolabeled probes and competitor DNAs were obtained from Invitrogen Corporation (Carlsbad, CA) and Integrated DNA Technologies (Coralville, IA). Appropriate sense and antisense strands were annealed in Tris 10 mM pH 8.0 by heating at 85 °C for 5 minutes and slowly cooling down to 25 °C. Double-stranded oligomeric probes were 5' –end labeled with [- P] ATP (5,000 Ci/mmol) using T4 polynucleotide kinase (New England Biolabs, Beverly, MA) according to the manufacturer's instructions. Labeled oligomers were separated from unreacted nucleotides by centrifugation through ProbeQuant G-50 Micro Columns (Amersham Biosciences, Piscataway, NJ). One, three or five micrograms of nuclear proteins were pre-incubated with 1 microgram of poly(dI-dC)·poly(dI-dC) (Sigma-Aldrich,) in 20 l of binding buffer (HEPES 20 mM pH 7.9, KCl 50 mM, $MgCl_2$ 5 mM, DTT 0.5 mM, and 4% glycerol) for 5 minutes at room temperature. The labeled probes were then added (100,000 cpm, approximately 0.1-0.2 ng) and complexes were allowed to form at room temperature for 15 minutes. The binding reactions were loaded onto a 4.5% non-denaturing polyacrylamide gels containing 2.5 % glycerol and subjected to electrophoresis in 0.5 x TBE buffer at 17-20 mA for about 3 hours at room temperature. The gel was dried and either autoradiographed overnight with an intensified screen or exposed to a Storage Phosphor Screen (Amersham Biosciences, Piscataway, NJ) and analyzed with a STORM 840 densitometer (Molecular Dynamics, Sunnyvale, CA). The specificity of protein-DNA complexes was analyzed by including 100-fold molar excess of homologous or non-homologous unlabeled competitor DNA into the binding reactions during the pre-incubation step prior to addition of probe. To identify nuclear factors that bind to the probes, 1 microliter of specific antibody (rabbit polyclonal anti-YY1, sc-1730; rabbit polyclonal anti-Sp3, sc-644 X; rabbit monoclonal anti-Sp1, sc-420 X; anti-Sp2 , sc-643 X; Santa Cruz Biotechnology, Santa Cruz, CA) was incubated with the binding mix for 1 hour at 4 °C before addition of the labeled oligomers.

"in vitro" Transcription/Translation Assay

"in vitro" translated YY1 was obtained with the TNT T7/T3 Coupled Reticulocyte Lysate System from Promega (Madison, WI). To obtain the template DNA plasmid, YY1 cDNA was released from *pCMV*-YY1 (gift of Dr. Shi, Harvard University) by digesting with *BamH1* and *Kpn 1* restriction endonucleases and cloned into *pCMX*-L1 vector similarly digested. The "*in vitro*" coupled transcription/translation reaction was carried out following the manufacturer's instructions in presence of T3 RNA Polymerase and 1 g of non-linearized template.

Acrylamide capture of 3 2 oligomer-bound complexes

Oligo A-3 2, the 3 2 upper strand oligonucleotide modified by addition of an Acrydite moiety (Mosaic Technologies, Waltham, MA) at its 5' prime was obtained from Integrated DNA Technologies, Inc. (Coralville, IA). The procedure was carried out as described to Nelson *et al.* (31) with few modifications: 1 M of A-3 2 oligo was annealed with the same amount of the unmodified lower strand and incubated with 150 micrograms of SW-13 cell nuclear extract or 0.5 mg of LNCaP nuclear extract under the same binding conditions herein described for the gel shift experiment (EMSA), scaling up the reaction volume accordingly to the amount of proteins and oligomers. After 15 minutes of incubation, the duplex molecules were immobilized to acrylamide as previously described (29) in the 1 cm large wells of a non-denaturing 4.5 % acrylamide gel (20 cm x 20 cm), and the unbound proteins were separated from the bound complexes by electrophoresis under the same conditions used for the EMSAs. The polyacrylamide fragments containing the covalently linked A-3 2/protein complexes were excised, and the proteins recovered by triturating the band in extraction buffer (Tris 50 mM pH. 7.5, SDS 0.1 %, EDTA 0.25 mM, and 2.5 % glycerol). Samples were kept under rotation overnight at 4 °C. The slurry was filtrated through a 0.22 m Ultrafree-MC Centrifugal Filter Unit and concentrated by a Microcon Centrifugal Filter Device (Millipore Corporation, Bedford, MA). Typically, about 0.8 ml of extraction buffer was reduced to 50 microliters.

Western blot analysis

Protein concentrations were determined by the BioRad protein assay. 10 g of nuclear proteins from SW-13 and LNCaP cells, and same amount of isolated A-3 2-bound proteins, along with 10 l of *in vitro* YY1 were diluted with 2x SDS loading buffer and loaded onto a 12% pre-cast GeneMate Endurance acrylamide gels (ISC BioExpress). Proteins were transferred to Hybond-P PVDF transfer membrane (Amersham Pharmacia Biotech) by a semi-dry procedure. Blots were probed with the same polyclonal antiserum against YY1 (1:500) or against Sp3 (1:5000) used in the gel shift experiments as primary antibodies, and horseradish peroxidase-conjugated goat anti-rabbit were used as secondary antibodies. Bands were visualized using the Western Blotting Chemiluminescence Luminol Reagent (Santa Cruz).

Cell cultures and transfections

All medium and sera were provided by Gibco BRL (Life Technology). SW-13 cells were grown in IMEM medium, and LNCaP cells in RPMI medium, both supplemented with 10% certified fetal bovine serum (FBS), 2 mM glutamine, and 100 IU/ml penicillin and 100 g/ml of streptomycin. COS-7 cells were grown in DMEM medium supplemented with 5% FBS. Cells were maintained at 37 °C with 5% $C0_2$. Unless otherwise specified, 24 hours before transfection mammalian cells were seeded in 12-well plates in medium without antibiotics (SW-13, $3x10^5$ cells/well; COS-7, $8x10^4$ cells/well). Transfections were carried out by the LipofectAmine 2000 method (Invitrogen, Life Science) following the manufacturer's instructions and using 2 g of DNA and 3 l of LipofectAmine (1:1.5) per well. When tested alone, reporter plasmids were transfected at the concentration of 1.8 g/well bringing the DNA amount up to 2 g with 0.2 g of *pCMV* Gal to monitor transfection efficiency. Co-transfection experiments with transcription factor expression vectors were carried out with 1 g/well reporter and 0.5 or 1 g/well of expression plasmid, keeping the DNA amount constant by adding empty vector. Because of the transcription factor interference with beta-galactosidase expression versus the control samples consisting of the reporter plasmid alone, no *pCMV* Gal was added in co-transfection experiments, and luciferase

activity was instead normalized for the protein content. Luciferase activity was measured for 20 seconds in a luminometer using the Luciferase Assay System (Promega Corporation, Madison, WI).

RESULTS

The HSD3B2 gene intron 1 is required for sustaining the basal promoter activity

It was previously shown that deletion of intron1, but not exon 1, reduced the HSD3B1 gene basal promoter activity from 3 to 6 fold (Guerin *et al.* 1995). To test whether intron 1 sequences play a similar role in the HSD3B2 gene basal promoter activity, a luciferase reporter construct bearing the HSD3B2 5' flanking region from nucleotide –517 and the 5' untranslated region up to nucleotide +193, *pGL3*-HSD2(-517/+193), was compared to a similar construct bearing a deletion of intron 1, *pGL3*-HSD2(517/+23), which are represented in Fig. 1 B. As intended also for a study aiming at characterizing polymorphisms that occur within intron 1, which might affect the splicing of the intron, our promoter constructs encompassing the first intron include the conserved splicing portion of the 3'splice junction with exon 2 up to 12 nucleotides (+182/+193). By RT-PCR we confirmed that including the first 12 nucleotides of exon 2 results in intron 1 being properly spliced out in Cos-7 cells (result not shown). When the promoter activity of the full-length construct and the deletion construct were compared upon transfection in several cell lines, it was evident that lack of intron 1 resulted in a 60% to 90% reduction of basal promoter activity, depending on the cell line under examination (Fig. 2). In SW-13 cells, the only steroidogenic cell line we tested, *pGL3*-HSD2(517/+193) had the highest activity, and lack of the intron resulted in a 5 fold reduction of luciferase expression. In the non-steroidogenic cell line Cos-7 the overall promoter activity was 10 fold lower, and the deletion construct residual basal activity accounted to only 10% of activity the full length construct (Fig 1 B). In LNCaP and PC-3 cells, the HSD3B2 gene promoter activity was 20 fold lower than seen in SW-13 cells, however the lower basal activity of *pGL3*-HSD2(517/+23) was also only 30 % the activity of *pGL3*-517/+193 (not shown). The higher promoter activity in SW-13 cells, derived from a small-cell carcinoma of the adrenal cortex (Leibovitz *et al.* 1978) may be attributed to tissue specificity. The lower activity of the intron-less reporter construct in steroidogenic and not steroidogenic cell lines showed that intron 1 is required for maximal levels of HSD3B2 gene basal promoter activity and confirms that, similarly to the HSD3B1 gene, the mechanism involved is not strictly tissue-specific.

The HSD3B2 intron 1 contains a 3 1-A element

To determine if a region with functional similarities to the HSD3B1 gene 3 1-A element might be responsible for the stimulating effect intron 1 demonstrated in transfection experiments, we compared oligo 3 1, matching the 3 1-A element as previously explored (Guerin *et al.* 1995), with oligo 3 2 matching the corresponding HSD3B2 gene intron 1 region in EMSA (electrophoretic mobility shift assay) experiments. The results, shown in Fig. 2 A, indicated that both probes are able to specifically shift a strong complex of identical electrophoretic mobility, and no other major bands were observed. The fact that 100 fold molar excess of unlabeled oligo 3 2, but not an oligo containing the Sp1 consensus, competed specifically the band formed by probe 3 1, and viceversa confirmed that the same protein specifically binds both probes. Guerin *et al.* had previously established the 3 1-A complex formation was dependent on the presence of zinc ions in the binding reactions, and it required the integrity of the stretch of 4 guanines shared by the 3 1-A element and TG box. By using the same criteria, we determined that the 3 1-A factor was the protein responsible for the identical mobility complex formed by the 3 1 and 3 2 probes. Indeed, the complex was (a) prevented by addition of the chelator EDTA at a concentration 10 mM, (b) restored when 5 mM $ZnCl_2$ was added to overcome the EDTA (Fig. 2B), and (c) a mutant oligo bearing a substitution of the 4 Gs to As (3 1m and 3 2m), was unable to shift the same complex when used as a probe (fig. 2A, *left*, lane 4 and 6), as well as unable to compete when added at 100 fold molar excess (Fig. 2B, *left*, last lane). Moreover, the same complex was formed with either probe with the nuclear extract obtained from 5 different cell lines, which complies with the ubiquity observed for the 3 1-A protein. Thus the HSD3B2 gene intron 1 contains an imperfectly conserved 3 1-A element that is still able to bind the 3 1-A protein, or the 3 1-A protein only binds to a strictly conserved nucleotide segment within this *cis*-element.

Sp3 binds the TG boxes of type I and type II intron 1 with different efficiency

The TG box overlapping with the 3 1-A element is conserved between type I and type II introns, however, the previous experiment failed to show significant Sp1 binding to both 3 1 and 3 2 probe, while in the earlier study a weak but more conspicuous Sp1 binding was observed (Guerin *et al.* 1995). While the previous study added at both extremities the unrelated sequence *gatc*, our probe consisted exclusively of nucleotides matching to the region of interest because of a different labeling strategy. Therefore, we reasoned the cause of our undetectable Sp1 binding could be the shorter sequence provided at the 3' end of our probe and that a probe encompassing more of the TG box 3' flanking region would provide a better spatial requirement for binding of Sp1 family members. For both type I and type II intron sequences, we designed an oligo encompassing less 5' flanking nucleotides and more of the 3' flanking region when compared to the 3 1-A element of both genes while maintaining the same length of 36 base pairs (Fig. 2 C). Despite lacking the first five upstream nucleotides of the 3 1-A element, these oligos were still able to form the 3 1-A band in gel shift experiments. In addition, they formed 3 to 4 new complexes of slower electrophoretic mobility, numbered in Fig. 2 B after assigning number 1 to the 3 1-A band. This result matches the 4 bands previously reported (Guerin *et al.* 1995), with the difference that the slower mobility band formed by the type I intron 1 specific probe was remarkably more intense, and comparable if not superior to the 3 1-A band (Fig. 2 B, *left*, first lane). This slower band appeared to be a doublet (named band 4a and 4b) and was efficiently competed by 100 fold molar excess of an Sp1 consensus containing oligonucleotide, which also competed for band 2. As expected, oligos 3 1 and 3 2 only partially competed for band 4 and 2, while efficiently competed out band 1. Oligo 3 1Sm and 3 2Sm, each bearing the same substitution of the 4 Gs to As earlier analyzed in Fig. 1 A, not only did not compete for complex 1 as expected, but also for complex 2 and 4, indicating that both Sp1-related and 3 1-A binding activities rely on the guanine stretch for optimal binding. These mutant oligos did compete the weak complex forming band 3 of intermediate mobility between the Sp1 consensus competed bands. Such competition suggests a protein binding to the 3 1-A element in a region not involving the four

guanines forms the complex. The use of antibodies against Sp1, Sp2 and Sp3 revealed that Sp3 accounts for two of the 4 complexes (Fig. 2 B, *middle*). Indeed, when Sp3 antibody was added to the binding reactions, band 4a, the lower and more conspicuous band of the doublet was abolished, along with band 2 likely formed by the smaller Sp3 (73 kDa) alternative translation product (Kennett *et al.* 1997, Suske 1999). Addition of Sp1 antibody resulted in disappearance of the fainter upper band (4a) of the doublet and only a weak supershift could occasionally be seen. Addition of Sp2 antibody had the same effect, with the difference that a slightly more intense supershifted band was seen with mobility lower than Sp1 supershift. A stronger 4b band was seen when antibody against Sp3 was used, suggesting that, as Sp3 has the strongest affinity for binding to the region, only when Sp3 is prevented from binding other factors can access the site. Surprisingly, the type II intron 1 TG box bound Sp3 less efficiently, as demonstrated by less Sp3 shift when 3 2S oligo was used as a probe (Fig. 2B, *right*), and by its partial ability to compete the doublet formed with labeled 3 1S (Fig. 2B, *left*, lane 4). By densitometry analysis we determined that Sp3 binds to the type II TG box 3 fold less than the type I TG box (results not shown). The binding activity could be brought to the same extent as seen for the type I intron 1 when an oligo bearing a substitution A+138 to G (Fig. 2 C) was used (not shown), which reproduces the 3' flanking region of the type I TG box. Despite the fact that Sp2 antibody did result in disappearance of the faster band of the doublet, we cannot conclude that Sp2 is part of the observed complex, as anti-Sp2 antibody was able to supershift Sp1 protein expressed in Drosophila Schneider's cells (not shown), which do not express Sp1 family members, thus indicating cross-reactivity.

Mapping the boundaries between 3 1-A and Sp3 binding activities

Our EMSA results indicated that 3 1-A and Sp3 binding to the 3 1-A element and TG box respectively are mutually exclusive, as no lower mobility complex indicative of a protein-protein interaction could be seen. Therefore, we were interested in creating mutants able to bind either the 3 1-A protein or Sp3 to be

able to discern the function of each of those two factors in relation to the HSD3B2 gene basal promoter activity upon transfection experiments. To precisely define how and at what extent the G_4 stretch is shared between the apparently overlapping *cis*-elements, we used two mutant oligonucleotides. Oligo M1 changed the first two guanine of the G stretch to As, while oligo M2 changed the last two guanines in the same way (Fig. 3 B). When used in EMSAs (Fig. 3 A, *left*), M1 was unable to bind Sp3, while the 3 1-A band was unaffected. M2 bound Sp3 similarly to as seen with wild type 3 2S, while 3 1-A band was severely diminished. This result indicates that the distal two guanines in the sequence TGGGGT are essential for Sp3 binding, and not necessary for 3 1-A binding. The proximal two guanines are not required for Sp3 binding, but important for 3 1-A factor binding.

Next we asked what is the contribution of the long stretch of adenines to 3 1-A factor binding, which accounts for as many as 8 and 6 residues in the HSD3B1 and HSD3B2 gene, respectively. We disrupted the A_6 sequence by changing the 5 adenines immediately preceding the TG box to CCTTC (oligo 3 2mS) and found that the substitution completely prevents formation of the 3 1-A band (Fig. 3 A, *middle*, first lane) while unable to form any additional complex. To determine what is the minimal number of adenines required for restoring the 3 1-A binding activity, we created a series of 3 oligos that reinstate two, three and four adenines respectively, and found that 4 adenines are required in front of the TG box to restore 3 1-A binding activity to the type II intron 1 (Fig. 3 A, *middle*, lane 2, 3 and 4). Lack of competition with the 3 1-A protein is likely the reason for the stronger Sp3 binding to these mutant oligomers, as well as with the unknown protein that forms band 3, the complex of intermediate mobility between the slower and faster Sp3 band, whose binding to the probe is dependent on the integrity of the A_6 stretch (Fig. 3, *middle*, all lanes lack this complex). We also tested a mutant oligo bearing a deletion of the TGGGGT sequence, $T(G)_4$, which surprisingly abrogated formation of every band seen with the wild type probe except for the 3 1-A band. In addition to the 3 1-A band, $T(G)_4$, M2 and all oligos bearing

mutations in the adenine stretch affected also formation of a faster band that appeared to be as specific as the 3 1-A band, but whose intensity and/or formation was not constantly observed, as later discussed.

Through this mutational analysis and considered that the 3 1-A protein is still able to recognize probe 3 1S and 3 2S despite of their lacking 5 nucleotides of the 3 1-A element, we could confirm that the 3 1-A protein recognizes only a shorter segment of the 24 base pair element. It was indeed evident that the conserved sequence AAAATGG represents the core-binding region for the 3 1-A factor, and that the Sp3 core-binding region is the sequence GGTGG. Therefore, the 3 1-A element and the TG box of the HSD3B2 and HSD3B1 gene appear to be flanking rather than overlapping *cis*-elements. In light of this conclusion, oligo $T(G)_4$ is still able to bind the 3 1-A factor because despite of the deletion of the 4 Gs, the two required G residues are yet provided by the following sequence. Sp3 binding, on the other hand, is no longer supported by this oligo, indicating that the sequence GGAGG is unable to functionally replace the sequence GGTGG.

YY1 is the protein forming the 3 1-A complex.

The sequence AAAATGG, that represent the core-binding region of the 3 1-A factor, is almost identical to the UCR (upstream conserved region) of the Moloney murine leukemia virus, which is known to bind the transcription factor UCRBP (Flanagan *et al.* 1992), later identified as the murine YY1 (Safrany & Perry 1993). UCR, which is only one of many consensus sites known to bind YY1 (Shi *et al.* 1997), all sharing the CCAT core motif, was later found also in many other viral and cellular genes. The 3 1-A factor core-binding region matches the UCR nucleotide sequence 7 out of 9 nucleotides (Fig. 4). To find out if the 3 1-A band is formed by YY1, an oligonucleotide containing the UCR element was tested in EMSA experiments in competition with the 3 2S probe. Addition of 100 fold molar excess of cold UCR oligo resulted in the complete disappearance of the 3 1-A band formed with nuclear extract from three different cell lines (Fig. 4 A), suggesting that the factor binding to the 3 2S probe –and therefore 3 1S as

well as 3 1 and 3 2– is able to recognize the UCR element. The competition was specific, as the Sp3 band was unaffected. Furthermore, when used as a probe, the UCR containing oligo formed a single complex with the same identical electrophoretic mobility as the 3 1-A band (Fig. 4 B). When a rabbit polyclonal antibody directed against full-length YY1 was added to the binding reactions, formation of both the 3 1-A band and the UCR band was lost and a complicated pattern of supershifted bands was seen with both probes. These results suggest that the protein responsible for the 3 1-A band is the transcription factor YY1 or a factor antigenically related to YY1. As YY1 is known as a 68 kDa protein, while the size of the 3 1-A factor was estimated to be 37 kDa by Guerin and co-workers (Guerin *et al.* 1995), we compared the size of the band shifted by labeled UCR oligomer from nuclear extracts with "*in vitro*" translated YY1. The "*in vitro*" transcribed and translated protein from plasmid *pCMX*-YY1 in gel retardation assays did shift the UCR probe forming a single specific complex of the same mobility as the one observed with nuclear extracts (Fig. 5 A). Western Blotting analysis revealed that anti-YY1 antibody –the same antibody used in our EMSAs–, recognized a band of molecular weight slightly above 55 kDa for the "*in vitro*" transcribed/translated YY1 protein (Fig. 5 B, lane 2) and nuclear extracts from both SW-13 (Fig. B A, lane 3) and LNCaP cells (Fig. 5 B, lane 6) as well. This band was not formed in negative control reactions consisting of an "*in vitro*" transcription/translation coupled reaction without template (Fig. 5 B, lane 1). This result indicated that the 3 1-A band formed by nuclear extracts with UCR, as well as 3 2 and 3 2S oligomers, correspond to full-length YY1 protein. In addition to the full-length band, anti-YY1 antibody recognized also two additional bands of lower molecular weight. A band approximately of 35 kDa and another band around 30 kDa could be seen with both nuclear extract and TNT rabbit reticulocyte lysate. To confirm that YY1 binds the 3 2 probe and that it can eventually be co-purified from the bound proteins as differently sized polypeptides, we have used a modification of the Acrydite™ technology method for the isolation of DNA binding proteins recently proposed by Nelson *et al.* (2002). Briefly, proteins from SW-13 or LNCaP cell nuclear extracts were incubated in binding buffer with a 3 2 oligomer whose upper strand carried an Acrydite moiety at its 5' prime. Then the oligomer

was immobilized to acrylamide as described (Nelson *et al.*2002) and the unbound proteins were separated from the bound complexes by electrophoresis. An aliquot of a crude preparation, consisting of eluted and concentrated 3 2-binding proteins from recovered acrylamide slices, reacted with anti-YY1 antibody in Western Blots, forming three bands of the same molecular weight as obtained with unbound nuclear extract proteins (Fig. 5 B, lane 5). Likely because of contaminants in these samples and our loss-prone modification of the original method, the antibody recognized bands were less intense than expected, particularly when compared with the bands from unbound nuclear proteins. Comassie staining of an SDS gel of the same preparation revealed bands whose size were identical to the YY1 specific bands (Fig. 5 C). In addition, an intense band of molecular weight intermediate to 35 and 49 kDa was also observed, that was not recognized by anti-YY1 antibody. Sp3 antibody, also a rabbit polyclonal antiserum, recognized the proper Sp3 specific bands of 105 and 7 kDa, but failed to recognize the 30 and 35 kDa bands (Fig. 5 B, lane 8). These results indicate that at least 3 different YY1 polypeptides were specifically recognized by anti-YY1 antibody, corresponding to the full-length protein, and to two fragments of about 35 and 30 kDa. They also indicate that these YY1 polypeptides bind the 3 2 oligomer, which is consistent with the previously estimated size of the 3 1-A factor and is likely due to proteolytic degradation of YY1.

Existence of a second YY1 site within the HSD3B2 and HSD3B1 gene Intron 1

Established that the sequence AAAATGG within the 3 1-A element binds YY1, as well as that YY1 and the 3 1-A factor are identical, we observed that a similar sequence is located in reverse orientation about 35 nucleotides downstream the TG box in the type II intron 1. Because of a substitution T/G, the site is not perfectly conserved in the type I intron 1 at the same location. However, thanks to another substitution A/G at nucleotide +134, a variant of the sequence is placed immediately following the TG box in the type I intron (AAAATGAGG in type I, AAAATAAGG in type II). To verify whether or not these sequences

are able to bind YY1, we designed the oligos shown in Fig. 6 B and tested them in EMSAs. As shown in Fig. 6 A, the EMSA experiments confirmed the ability of YY1 to bind the HSD3B2 intron 1 35 base pairs downstream the 3 1-A element (oligo 147). YY1 recognized this oligo with an affinity slightly lower than the efficiency exhibited by binding to the UCR and even to the 3 1-A element. This discovery is in agreement with our finding that only the two proximal G residues are absolutely required for YY1 binding within the 3 1-A element (Fig. 3). In accordance to these observations, the corresponding sequence in type I intron 1 is not able to bind YY1 because of its bearing only 3 out of the 4 required adenines. Oligo 125, matching the putative YY1 binding site laying at the end of the type I intron 1 TG box was able to bind YY1, however with much lower affinity, while the same region in the type II intron was unable to bind this transcription factor, as expected accordingly to the results of the mutational analysis shown in Fig. 3. This analysis revealed that YY1 recognizes the same sequence in two different regions within intron 1 of both the HSD3B1 and HSD3B2 genes. One site is conserved between the two genes for nucleotide sequence and location, being it the 7 conserved nucleotides of the 3 1-A element immediately juxtaposed to the 5' prime of the TG box; the second site differs between the two genes for location and orientation. In type I intron 1 the second YY1 site is a slight variant of the AAAATGG sequence located 8 nucleotides downstream the 3 1-A element, immediately following the 3' prime of the TG box. In type II intron 1 the second site lays in reverse orientation 35 base pair apart from the 3 1-A element. These findings are summarized in Fig. 6 C.

YY1 enhances the HSD3B2 gene basal promoter activity by binding to the two identified binding regions within the gene intron 1

To explore whether or not YY1 is functionally involved in the boost of basal promoter activity mediated by the HSD3B2 gene intron 1, we over-expressed YY1 in presence of the same reporter constructs (Fig. 1 B) we earlier used to assess the functional relevance of the intron in transient transfection assays. As

shown in Fig. 7 (A), YY1 stimulated the full-length construct *pGL3*-HSD2(-517/+193) in a dose dependent manner up to about 4 fold, while it failed to stimulate the intron-less construct *pGL3*-HSD2(-517/+23). This result confirmed that YY1 is the effector of the trans-activation mediated by intron 1 on the gene basal promoter activity. To investigate the contribution of each of the two YY1 binding sites we had identified we proceeded comparing the activity of a wild type luciferase reporter construct encompassing the HSD3B2 gene sequence from nucleotide –246 to nucleotide +193, *pGL3*-HSD2(-246/+193), with identical constructs that were abrogated of YY1 binding to either the 3 1-A element or the distal site or both, by changing the 4 adenines required for its binding at both sites. For this purpose, oligo B2mS (shown in Fig. 3) and its antisense were used to abolish YY1 binding in front of the TG box; an oligo identical to oligo 147 (shown in Fig. 6 B), except for having the 4 critical adenines changed to CCTT, was used to eliminate YY1 binding to the distal site by site-directed mutagenesis. In addition, we also tested a construct bearing the $T(G)_4$ deletion that eliminates the sequence TGGGG, thus preventing Sp3 binding but not YY1 binding (Fig. 3). Surprisingly, inability of YY1 to bind its site within the 3 1-A element had little effect on the basal promoter activity (Fig. 7 B). A similar effect was seen with the construct that bears a deletion of the TG box. Therefore, YY1 or Sp3 binding to the juxtaposed 3 1-A element and TG box do not play any major role in the intron 1 mediated increase of the HSD3B2 gene basal promoter activity. Instead, deletion of the distal YY1 binding site resulted in a 50% decrease of basal activity. Complete abrogation of YY1 binding within intron 1 resulted in a basal promoter activity that was only 30% of the wild type activity. Remarkably, this result reproduced the loss of activity seen with the promoter construct missing the whole intronic sequence and also it made evident that Sp3, or any other Sp1 family members, has no ability to direct the boost in basal activity through intron 1 of the HSD3B2 gene. Thus, the proximal YY1 site within the 3 1-A element and the distal YY1 binding site contribute to boost the HSD3B2 gene basal promoter activity. We observed the same effect in several cell lines, including SW-13, Cos-7 and LNCaP cells.

DISCUSSION

Despite the pivotal role of the 3 –Hydroxysteroid dehydrogenase type II enzyme in the biosynthesis of all steroids, the molecular mechanisms that regulate the gene expression are yet to be fully understood. Lately, important studies have been published concerning the significance of steroidogenic factor 1 in the HSD3B2 gene expression and its synergistic effect with phorbol ester (Leers-Sucheta *et al.* 1997), as well as its regulation by cytokines (Cote *et al.* 2000, Gingras *et al.* 2001), gonadotropines (Feltus *et al.* 1999) and glucocorticoids (Feltus *et al.* 2002). However, no study has investigated the role of intron 1 in the HSD3B2 gene basal activity so far, despite a previous finding that the highly homologous HSD3B1 gene requires the activating 3 1-A element within the first intron to keep the basal promoter activity to its full potential by binding the unknown 3 1-A factor. In this study we have shown that the HSD3B2 gene intron 1 is required to achieve maximal basal activity levels of the promoter region. Despite the suggestion that the HSD3B1 and HSD3B2 genes may not share the 3 1-A element (Guerin *et al.* 1995), we have proved the existence of such a *cis* acting element in the intron 1 of the HSD3B2 gene. More specifically, we have proved that the 3 1-A protein binds to a conserved region of 7 nucleotides of the 3 1-A element in both genes, through a mutational analysis that has ultimately helped us pinpointing similarities between the 3 1-A binding core and the UCR core recognized by the multi role transcription factor YY1. We have shown evidence that YY1 is indeed the protein binding "*in vitro*" to the 3 1-A element by competition analysis and use of anti-YY1 antibody.

YY1 fits perfectly the profile for being the 3 1-A factor, as it is ubiquitously and constitutively expressed (Shi *et al.* 1997, Austen *et al.* 1997). It contains four zinc finger domains with homology to the GLI-Kruppel family of proteins (Shi *et al.* 1991), thus explaining the 3 1-A protein requirement for zinc ions. However, despite a predicted molecular weight of 44 kDa, YY1 is known to migrate as a 65-68 kDa protein in SDS gels (Shi *et al.* 1997), whereas the 3 1-A factor was attributed a molecular weight of 37 kDa by UV-crosslinking analysis (Guerin *et al.* 1995). This apparent discrepancy can be explained by

recent findings that claim YY1 proteolytic mechanisms as part of degradation processes and even tissue-specific signal transduction pathways. Specifically, YY1 was shown to be the target of calpein II in a mechanism aimed at downregulating YY1 protein during muscle development (Walowitz *et al.* 1998), which generates a 40 kDa polypeptide. Moreover, a nuclear cathepsin B-like protease activity appears to degrade YY1 (Pizzorno 2001) generating two fragments of about 30 and 40 kDa, a phenomenon associated with the progression of undifferentiated to differentiated NT2 cells upon treatment with retinoic acid. In this case, the larger cleavage product, representing the carboxy-terminal portion of YY1 containing the zinc-finger domain, was shown to bind the cognate DNA consensus forming a faster complex in gel shift assays. Bovolenta *et al.* (1999) found that administration of Interleukin 2 to individuals affected by HIV induced downregulation of YY1 protein through activation of an unidentified proteolytic activity, which resulted in appearance of a faster YY1 band of unknown molecular weight in gel shift analysis. Interestingly, when YY1 was first isolated for its ability to bind the UCR site, a 40 kDa antigenically related protein was co-purified with it (Becker *et al.* 1994) and found to account for much of the UCR binding activity in T-lymphocytes along with YY1. In light of the latest findings, the 40 kDa protein could have been a truncated form of YY1. As a consequence of these observations, it cannot be excluded that in the previous study on the HSD3B1 gene (Guerin *et al.* 1995) during the procedures required for UV cross-linking analysis, a pre-existing truncated YY1 fragment was also isolated. In addition to the 37 kDa band attributed to the 3 1-A protein, accordingly to Guerin and coworkers, the UV cross-linking analysis produced two additional specific complexes that, once approximately corrected for the contribution of the oligonucleotide, did yield molecular weight compatible with full-length YY1. In support of the above suggestions, we did observe frequently but not constantly appearance of a band of faster mobility in respect of the YY1/3 1-A band in our gel shift experiments, specific to both 3 1S/3 2S and UCR probes (Fig. 3, and seen also in Fig. 6). Formation of this faster band seems to be sensitive to experimental slight changes of pH or ionic strength that facilitate binding of pre-existing factors/fragments, and also it might be influenced by the nucleotides flanking the consensus, as our

deletion construct T(G)$_4$ was able to form this faster complex at all times. In addition, use of anti-YY1 antibody in gel shift experiment showed more than one supershifted band. We exclude that YY1 fragmentation occurred during the manipulation of the nuclear extracts and incubation of the binding reactions, because despite absence of protease inhibitors in our nuclear extracts, incubation of the extract at 37° C for 1 hour before probe addition in EMSA binding reactions did not favor the appearance of faster bands in control experiments that did not exhibit formation of those bands in the untreated samples (not shown). Despite we cannot at this time explain their origin, neither their significance in relation to the HSD3B2 gene basal activity, we have demonstrated that 2 bands of similar molecular weight to the previously identified YY1 fragments could be isolated from nuclear complexes bound to the 3 1-A element. The 35 kDa YY1 fragment could likely account for the 37 kDa band previously attributed to the 3 1-A factor.

In addition to having characterized YY1 as the factor binding the 3 1-A element, we have identified a second YY1 binding site in both type I and type II introns. Indeed, a sequence 100% identical to the region we have found to bind YY1 within the 3 1-A element, and 80% identical to the UCR, was found to bind YY1 at the end of the HSD3B2 intron 1. Interestingly, this sequence is reversed in respect of the orientation of the identical sequence within the 3 1-A element. A weaker YY1 binding site was found at the 3' end of the TG box in the HSD3B1 intron 1. This last finding might explain why Sp3 binds with better efficiency the 3 1S probe than the 3 2S probe. We could bring the band intensity formed by Sp3 bound to 3 2S to that seen for the 3 1S probe when we changed the only nucleotide that differs between the type I and type II sequence. This one nucleotide mismatch in position +134, A in type II, G in type I, does not affect the region strictly required for Sp3 binding (Fig. 2), neither is it part of a region that may represent another Sp1/Sp3 binding site, as demonstrated by the fact that oligos 125 and 126 are unable to shift Sp1 family members (Fig. 6 B). It is possible that a G instead of A at the 3' end of the Intron 1 TG box may facilitate Sp3 binding to a region rich in As and Ts, by reducing or disrupting any intrinsic DNA structure. However, a more likely explanation is that in the fewer cases when YY1 is sitting onto the less

efficient site outside the 3 1-A element, Sp3 displaces it for binding to the TG box easier than it can displace YY1 bound to the stronger site within the 3 1-A element. Our findings suggest the effectors playing a role in the HSD3B1 gene basal activity need to be further investigated, as the conclusions of the previous study were based on the analysis of a mutant that affected not only YY1 binding to the 3 1-A element but also Sp3. While the importance of Sp1 was disregarded in the previous study because of the weaker binding of this factor to the 3 1-A element flanking TG box, we have herein proved the oligomer used in that study could not support proper Sp1 family member binding, and that Sp3 is the Sp1 family member that electively binds the region with intensity comparable to the complex formed by YY1. Nevertheless our results in mammalian cells did not evidence any role for Sp3 in the maintenance of the HSD3B2 gene basal activity, the different organization of the two YY1 binding sites in respect of the TG box in the HSD3B1 gene intron 1, and the consequent improved Sp3 competition for its site in "*in vitro*" experiments, suggest that Sp3 may have a different outcome in the HSD3B1 gene basal promoter activity. The importance of the weaker YY1 binding site immediately flanking the TG box at the 3' end also needs to be assessed.

We have functionally characterized the importance of YY1 in the intron 1 mediated enhancement of the HSD3B2 gene basal promoter activity. Dissimilarly from the HSD3B1 gene (Guerin *et al.*1995), disruption of the YY1 binding site within the 3 1-A element did not result in a 3-6 fold reduction of promoter activity, an outcome that further contributes to make evident differences in the mechanism implicated in the function of type I and type II intron 1. However, disrupting YY1 binding to the distal site with opposite orientation resulted in a 50% decrease in basal promoter activity. When both sites were simultaneously eliminated, the residual promoter activity was only 30% of the wild type promoter activity, a result functionally identical to the lack of the whole intronic sequence. As the two 35 base pairs apart YY1 binding regions have opposite orientation, and YY1 is known to induce DNA bending by binding to its cognate sequence (Natesan & Gilman 1993, Kim & Shapiro 1996), one could speculate that YY1 binding to the two regions within the HSD3B2 intron 1 would induce formation of a loop which

might help binding and/or interaction of other factors involved in the HSD3B2 gene basal transcription. As it is known that TBP and TFIID require bending of the TATA box region for optimal binding (Parvin *et al.* 1995), it is suggestive to hypothesize that YY1 binding to intron 1 would help maintaining a DNA conformation that facilitates the RNA polymerase complex formation onto the imperfect TATA box (Lachance *et al.* 1991) of the HSD3B2 gene. Similarly to our study, Kerr *et al.* (1997) found that extending the human bone sialoprotein gene promoter to include the first 68 nucleotides of intron 1 resulted in a 3 to 7 fold increase in reporter activity in UMR106-01 BPS cells. Interestingly, the human bone sialoprotein gene possesses an inverted TATA box, and one YY1 binding site was found within the first 30 base pairs of intron 1. The explanation we suggest is more likely to describe the mechanism of YY1 action through its binding sites in the HSD3B2 gene intron 1 rather than YY1 acting by changing the DNA structure and/or chromatin organization by its ability to recruit factors with intrinsic histone acetylase activity (Thomas & Seto 1999). Indeed, in preliminary transfection experiments, 66 nM Trichostatin A, a drug known to inhibit hystone deacetylase thus boosting gene activity by favoring an open chromatin structure, did increase the basal activity of the wild type and double YY1 deletion constructs at a similar extent (3 fold) in transfection experiments. This suggests that YY1 is not functioning at the intron 1 level by favoring hystone acetylation.

Taken together our results indicate that the HSD3B2 gene intron 1 is required for maintaining the gene basal promoter activity at the highest levels observed in transfection experiments with a reporter system, and that YY1 is the factor that mediates such effect by binding to two different regions within intron 1. As we have also identified YY1 to be the 3 1-A protein, the *trans*-activator that was attributed a similar role in the HSD3B1 gene, this study points to a major role of YY1 in affecting the basal levels of 3 -Hydroxysteroid dehydrogenase activity. Additional studies will be required to explore the mechanisms implicated in YY1 contribution to the basal activity of both the HSD3B1 and HSD3B2 genes.

An important consideration arising from this study and the previous study on the HSD3B1 gene is that nucleotide sequence variants occurring within the HSD3B1 and HSD3B2 gene intron 1 may result in

more or less profound changes in gene basal activity, and consequently they should not be excluded from being characterized merely because of their occurrence within an intronic sequence.

Lastly, nonetheless the importance of our findings in the contest of the native chromosomal gene needs to be assessed, the information this study provides may be valuable for designing targeted gene therapy strategies, and in particular strategies based on the employment of suicide genes. The HSD3B2 gene offers an example of steroidogenic tissue-specific promoter that could be utilized for targeted gene therapy strategies designed for localized diseases such as ovarian, adrenal and testicular cancer.

Acknowledgements

We thank Dr. Yang Shi for consenting us to use *pCMV*-YY1. We are most grateful to Dr. Baruch Frenkel and Dr. Amy Lee for critical review of the manuscript. We wish to acknowledge Claudia Wehie for skillful technical assistance, and members of the Reichardt lab, Troy Phipps above all, for assistance with the ABI PRISM 3100 DNA Sequencer. This work was supported by DoD grant # 001280 awarded to Dr. Juergen Reichardt

REFERENCES

Austen M, Lüscher B & Lüscher-Firzlaff JM 1997 Characterization of the transcriptional regulator YY1. *The Journal of Biological Chemistry* **272** 1709-1717

Bartsch W, Klein H, Schiemann U, Bauer HW & Voigt KD 1990 Enzymes of androgen formation and degradation in the human prostate. *Annals of the New York Academy of Sciences* **595** 53-66

Becker KG, Jedlicka P, Templeton NS, Liotta L & Ozato K 1994 Characterization of hUCRBP (YY1, NF-E1, delta): a transcription factor that binds the regulatory regions of many viral and cellular genes. *Gene* **150** 259-266

Bosland MC 2000 The role of steroid hormones in prostate carcinogenesis. *Journal of the National Cancer Institute. Monographs* **2000** 39-66

Bovolenta C, Camorali L, Lorini LA, Vallanti G, Ghezzi S, Tambussi G, Lazzarin A & Poli G 1999 Expression and activation of a C-terminal truncated isoform of STAT5 (STAT5) following interleukin 2 administration or AZT monotherapy in HIV-infected individuals. *Journal of Immunology* **163** 6892-6897

Chang BL, Zheng SL, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB & Xu J. 2002 Joint effect of HSD3B1 and HSD3B2 genes is associated with hereditary and sporadic prostate cancer susceptibility. *Cancer Research* **62** 1784-1789

Cote S, Feltus A, Freeman M, Melner MH & Simard J 2000 IL-4 stimulation of ovarian 3 - hydroxysteroid dehydrogenase/ ^{5}- ^{4}-isomerase type II gene expression: mechanisms of activation. In: Proceedings of the 82nd Annual Meeting of Endocrine Society, pp 1295.

Dent CL & Latchman DS 1993 Transcription Factors: A Practical Approach, edn 2, pp 11. Ed Latchman DS. IRL Press, Oxford University Press, Oxford

Eldar-Geva T, Hurwitz A, Vecsei P, Palti Z, Milwidsky A & Rosler A 1990 Secondary biosynthetic defects in women with late-onset congenital adrenal hyperplasia. *The New England Journal of Medicine* **323** 855-863

Flanagan JR, Becker KG, Ennist DL, Gleason SL, Driggers PH, Levi BZ, Appella E & Ozato K 1992 Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. *Molecular and Cellular Biology* **12** 38-44

Feltus FA, Groner B & Melner MH 1999 Stat5-mediated regulation of the human type II 3 -Hydroxysteroid dehydrogenase/ ^{5}- ^{4}-isomerase gene: activation by prolactin. *Molecular Endocrinology* **13** 1084-1093

Feltus FA, Cote S, Simard J, Gingras S, Kovacs WJ, Nicholson WE, Clark B & Melner MH 2002 Glucocorticoids enhance activation of the human type II 3 -hydroxysteroid dehydrogenase/ ^{5}- ^{4} -isomerase gene. *The Journal of Steroid Biochemistry and Molecular Biology* **82** 55-63

Gann PH, Hennekens CH, Ma J, Longcope C & Stampfer MJ 1996 Prospective study of sex hormone levels and risk of prostate cancer. *Journal of the National Cancer Institute* **88** 1118-1126

Guerin S, Leclerc S, Verreault H, Labrie F & Luu-The V 1995 Overlapping cis-acting elements located in the first intron of the gene for type I 3 -hydroxysteroid dehydrogenase modulate its transcriptional activity. *Molecular Endocrinology* **9** 1583-1597

Gingras S, Cote S & Simard J 2001 Multiple signal transduction pathways mediate interleukin-4-induced 3 -hydroxysteroid dehydrogenase/ ^{5}- ^{4}-isomerase in normal and tumoral target tissues. *The Journal of Steroid Biochemistry and Molecular Biology* **76** 213-225

Kennett SB, Ulvadia AJ &Horowitz JM 1997 Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. *Nucleic Acids Research* **25** 3110-3117

Kerr JM, Hiscock DRR, Grzesik W, Robey PG &Young MF 1997 The human bone sialoprotein gene contains an NF-E1/YY1 cis-acting sequence with putative regulatory activity. *Calcified Tissue International* **60** 276-282

Kim J & Shapiro DJ 1996 In simple synthetic promoters YY1-induced DNA bending is important in transcription activation and repression. *Nucleic Acids Research* **24** 4341-4348

Labrie F, Simard J, Luu-The V, Belanger A & Pelletier G 1992 Structure, function and tissue-specific gene expression of 3 -hydroxysteroid dehydrogenase/5-ene-4-ene isomerase enzymes in classical and peripheral intracrine steroidogenic tissues. *The Journal of Steroid Biochemistry and Molecular Biology* **43** 805-826

Lachance Y, Luu-The V, Labrie C, Simard G, Dumont M, de Launoit Y, Guerin S, Leblanc G & Labrie F 1990 Characterization of human 3 -hydroxysteroid dehydrogenase/ ^{5}- ^{4}-isomerase gene and its expression in mammalian cells. *The Journal of Biological Chemistry* **265** 20469-20475

Lachance Y, Luu-The V, Verreault H, Dumont M, Rheaume E, Leblanc G & Labrie F 1991 Structure of the human type II 3 -hydroxysteroid dehydrogenase/ ^{5}- ^{4}-isomerase (3 -HSD) gene: adrenal and gonadal specificity. *DNA and Cell Biology* **10** 701-711

Lala DS, Rice DA & Parker KL 1992 Steroidogenic factor I, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu-factor I. *Molecular Endocrinology* **6** 1249-1258

Leers-Sucheta S, Morohashi K-I, Mason JI & Melner MH 1997 Synergistic activation of the human type II 3 -hydroxysteroid dehydrogenase/ ^{5}- ^{4}-isomerase promoter by the transcription factor Steroidogenic Factor-1/Adrenal 4-binding protein and phorbol ester. *The Journal of Biological Chemistry* **272** 7960-7967

Leibovitz A, McCombs WM, Johnston D, McCoy CE & Stinson JC 1973 New human cancer cell culture lines. I. SW-13, small-cell carcinoma of the adrenal cortex. *Journal of the National Cancer Institute* **51** 691-697

Levine LS 2000 Congenital adrenal hyperplasia. *Pediatrics in Review* **21** 159-170

Luu-The V, Lachance Y, Labrie C, Leblanc G, Thomas JL, Strickler RC & Labrie F 1989 Full length cDNA structure and deduced amino acid sequence of human 3 -hydroxy-5-ene steroid dehydrogenase. *Molecular Endocrinology* **3** 1310-1312

Marker PC, Donjacour AA, Dahiya R & Cunha G 2003 Hormonal, cellular, and molecular control of prostatic development. *Developmental Biology* **253** 165-174

Marui S, Castro M, Latronico AC, Elias LL, Arnhold IJ, Moreira AC & Mendonca BB 2000 Mutations in the type II 3 -hydroxysteroid dehydrogenase (HSD3B2) gene can cause premature pubarche in girls. *Clinical Endocrinology* **52** 67-75

Mendonca BB, Russell AJ, Vasconcelos-Leite M, Arnhold IJP, Bloise W, Wajchenberg BL, Nicolau W, Sutcliffe RG & Wallace AM 1994 Mutation in 3 -hydroxysteroid dehydrogenase type II associated with pseudohermaphroditism in males and premature pubarche or cryptic expression in females. *Journal of Molecular Endocrinology* **12** 119-122

Morissette J, Rheaume E, Leblanc JF, Luu-The V, Labrie F & Simard J 1995 Genetic linkage mapping of HSD3B1 and HSD3B2 encoding human types I and II 3 -hydroxysteroid dehydrogenase/ ^{5}- ^{4}-isomerase close to D1S514 and the centromeric D1Z5 locus. *Cytogenetics and Cell Genetics* **69** 59-62

Morohashi K, Honda S, Inomata Y & Omura T 1992 A common trans-acting factor, Ad4-binding protein, to the promoters of steroidogenic P-450s. *The Journal of Biological Chemistry* **267** 17913-17919

Morrison N, Nickerson DA, McBride MW, Mueller UV, Boyd E & Sutcliffe RG 1991 Regional chromosomal assignment of human 3-beta-hydroxy-5-ene steroid dehydrogenase to 1p13.1 by non-isotopic *in situ* hybridization. *Human Genetics* **87** 223-235

Natesan S & Gilman MZ 1993 DNA bending and orientation-dependent function of YY1 in the c-fos promoter. *Genes & Development* **7** 2497-2509

Nelson C, Hendy S, Reid K & Cavanagh J 2002 Acrylamide capture of DNA-bound complexes: electrophoretic purification of transcription factors. *BioTechniques* **32** 808-815

Pang S, Wang W, Rich B, David R, Chang YT, Carbunaru G, Myers SE, Howie AF, Smillie KJ & Mason JI 2002 A novel nonstop mutation in the stop codon and a novel missense mutation in the type II 3 -hydroxysteroid dehydrogenase (3 -HSD) gene causing, respectively, nonclassic and classic 3 -HSD deficiency Congenital Adrenal Hyperplasia. *Journal of Clinical Endocrinology and Metabolism* **87** 2556-2563

Parvin JD, McCormick RJ, Sharp PA & Fisher DE 1995 Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor. *Nature* **23** 724 727

Pizzorno MC 2001 Nuclear cathepsin B-like protease cleaves transcription factor YY1 in differentiated cells. *Biochimica et Biophysica Acta* **1536** 31-42

Rheaume E, Lachance Y, Zhao HF, Breton N, Dumont M, de Launoit Y, Trudel C, Luu-The V, Simard J & Labrie F 1991 Structure and expression of a new complementary DNA encoding the almost exclusive 3 -hydroxysteroid dehydrogenase -4 -isomerase in human adrenals and gonads. *Molecular Endocrinology* **5** 1147-1157

Ross RK & Schottenfeld D 1996 Prostate cancer. In: *Cancer epidemiology and prevention*, edn 2 pp 1180-1206. Eds D Schottenfeld & J Fraumeni. Oxford: Oxford University Press

Ross RK, Pike MC, Coetzee GA, Reichardt JK, Yu MC, Feigelson H, Stanczyk FZ, Kolonel LN & Henderson BE 1998 Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. *Cancer Research* **58** 4497-4504

Safrany G & Perry RP 1993 Characterization of the mouse gene that encodes the /YY1/NF-E1/ UCRBP transcription factor. *Proceedings of the National Academy of Sciences of the United States of America* **90** 5559-5563

Simard J, Rheaume E, Sanchez R, Laflamme N, de Launoit Y, Luu-The V, van Seters AP, Gordon RD, Bettendorf M & Heinrich U 1993 Molecular basis of Congenital Adrenal Hyperplasia due to 3 - Hydroxysteroid dehydrogenase deficiency. *Molecular Endocrinology* **7** 716-728

Simard J, Moisan AM & Morel Y 2002 Congenital adrenal hyperplasia due to 3 -hydroxysteroid dehydrogenase/ ^{5}- ^{4} -isomerase deficiency. *Seminars in Reproductive Medicine* **20** 255-276

Shi Y, Seto E, Chang LS & Shenk T 1991 Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67** 377-388

Shi Y, Lee, J-S & Galvin KM 1997 Everything you have ever wanted to know about Yin Yang 1. *Biochemica et Biophysica Acta* **1332** F49-F66

Suske G 1999 The Sp-family of transcription factors. *Gene* **238** 291-300

Thomas MJ & Seto E 1999 Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* **236** 197-209

Wa[illegible]owitz JL, Bradley ME, Chen S-J & Lee T-C 1998 Proteolytic regulation of the zinc finger transcription factor YY1, a repressor of muscle-restricted gene expression. *The Journal of Biological Chemistry* **273** 6656-6661

White PC 2001 Congenital adrenal hyperplasias. Best Practice & Research Clinical Endocrinology & Metabolism **15** 17-41

FIGURE LEGENDS

Fig. 1: The HSD3B2 gene intron 1 is required for maximal promoter basal activity. A A schematic representation of the HSD3B2 gene organization and partial sequence alignment of the HSD3B1 and HSD3B2 intron 1 is shown. Open boxes indicate untranslated exon sequences as well as the 5' and 3' untranslated regions; solid boxes indicate the coding exons. Roman numbers identify each exon. The transcription initiation start in exon 1 is shown. In the sequence alignment, the nucleotides part of the control elements previously identified in the HSD3B1 gene intron 1 (26) are wrapped by a colored box and the name of each element is shown above each box. The nucleotides previously identified as protected in DNAse footprinting experiments are underlined. Small letters evidence nucleotide differences between the two aligned sequences. Dotted lines replace missing nucleotides. **B** SW-13 cells were seeded in 6-well plates at $6x10^5$ cells/well. After 16 hours cells were transfected by Lipofectamine 2000 with 3.5 g of *pGL3*-HSDB2(–517/+193), a luciferase reporter gene driven by the HSD3B2 promoter region linked to the 5-UTR, spanning the first exon and intron 1, or *pGL3*-HSDB2(-517/+23), which lacks the intron 1 sequence, separately added in triplicate wells. Total DNA was kept at 5 g by addition of backbone vector and a ratio DNA to lipofectamine of (1:2.5) was used. Cells were harvested 48 hours later. The promoter activity is reported as relative luciferase units (rlu) mean of 3 independent experiments. The standard deviations are shown. Cos-7 cells were seeded at $1.6x10^5$/well and similarly treated. The luciferase reporter constructs are schematized next to their own activity bar. Negative numbers indicate the first nucleotide of the HSD3B2 gene 5' flanking region included in the construct, positive numbers identify the nucleotidic delimitation of the untranslated coding regions included in each construct. Open boxes schematize untranslated coding regions. The transcription initiation start (+1) is shown. A straight bold line represent the intron 1 sequence, two interconnected lines indicate its deletion.

Fig. 2: Identification of a 3 1-A Element and TG box in the HSD3B2 gene Intron 1. A *Left*: EMSA result obtained when 3 g of nuclear extract from the indicated cell lines were incubated with labeled 3 1 oligo encompassing the HSD3B1 gene 3 1-A element, labeled oligo 3 2 encompassing the HSD3B2 gene homologous region, and their mutants 3 1m and 3 2m bearing the substitutions indicated in C. The complexes were resolved as indicated in "Materials and Methods". The names of each probe specific to either type I or type II intron, and the competitor oligos used in the experiment are indicated. -, indicates no competition. A major band of identical electrophoretic mobility was formed by each wild type probe, which was specifically competed by 100 fold molar excess of the same unlabelled oligo (self), but not competed by 100 fold molar excess of an oligo containing the Sp1 consensus. The two mutant probes were not shifted, which indicates the protein binding to the wild type probes requires the Gs of the overlapping TG box, as established for the 3 1-A protein (26). *Right*: to further verify the authenticity of the major complex as the 3 1-A protein, as well as its ubiquity, 5 g of nuclear extracts from LNCaP cells were incubated with the 3 1 probe in presence of EDTA 10 mM, with and without 5 mM zinc chloride. The complex was not seen in presence of EDTA, an ion chelator, while it was restored when Zn^{++} was added to the EDTA containing binding reaction. The competition with 100 fold molar excess cold 3 2 oligomer is also shown as further evidence that the same protein binds both 3 1 and 3 2 oligomers. **B** *Left*: EMSA result obtained with oligo 3 1S that, in respect of 3 1 oligo, encompasses ten nucleotides at the 3' end of the TG box, while it is depleted of the twelve 5'-flanking nucleotides, of which 5 are part of the 3 1-A element. Competitor oligos were used at 100 fold molar excess. Four major bands were observed, numbered 1 to 4 after assigning nr 1 to the 3 1-A band. Band 4 appeared to be a doublet, indicated as 4a and 4b. Self-competition indicates the complexes forming each band are specific. Oligo 3 2S competed all the bands but band 4a less efficiently. Oligo Sp1 efficiently competed band 2 and the double band 4, while it did not affect the 3 1-A band. Oligos 3 1 and 3 2 competed the 3 1-A band, but only weakly the Sp1 specific bands, as expected for their shorter 3' end sequence when compared to the probe. The mutant 3 1m and 3 2m oligo were unable to compete any of the bands

except for the weaker complex forming band 3 (only 3 1m competition is shown). *Middle*: the same amount of nuclear extracts as above were pre-incubated with anti- Sp1, anti-Sp2 and anti-Sp3 antibodies for 1 hour at 0° C, than added of probe and further incubated for 15 minutes at room temperature. Sp1 and Sp2 antibodies diminished the weak 4b band, and a faint supersfhit was seen. Sp3 antibody completely prevented formation of band 2 and band 4a. *Right*: comparison of the 3 1S and 3 2S oligonucleotide binding ability. While the two oligos have the same affinity for the 3 1-A protein, they bind Sp3 with different efficiency. **C** The upper nucleotide sequence of the oligomers employed in the study is shown. Oligos are aligned to center the shared regions encompassing the 3 1-A element, which is evidenced by a line above the 3 1 oligo. The outlined nucleotide in 3 2S and 3 1S refers to the single nucleotide mismatch existing in their 3' end that might affect Sp3 affinity for the TG box.

Fig. 3: Dissection of the 3 1-A element and TG box in the HSD3B2 gene intron 1. A *Left*: EMSA result obtained with 5 g of nuclear extract from SW-13 cells aimed at comparing the effect of nucleotide changes within the G_4 stretch shared by the 3 1-A element and TG box. The binding ability of mutants M1 and M2, shown in (B), is compared with the band pattern formed by oligo 3 2S. M1, that changes the last two Gs to As, was unable to bind Sp3 while it could still bind the 3 1-A protein. M2, that replaces the first two Gs with two As, could bind Sp3 but its ability to bind the 3 1-A was severely reduced. *Middle*: EMSA result comparing the binding ability of the wild type 3 2S probe with oligos bearing mutations that alter the stretch of adenines in the type II 3 1-A element as shown in (B). The substitution A_5 to CCTTC (3 2mS) completely abolished formation of the 3 1-A band, while it did not form any extra complex. Mutant oligos A_2 and A_3, that restore 2 and 3 of the adenines in front of the TG box, were unable to support 3 1-A protein binding, while addition of four adenines (A_4) resulted in a fully restored 3 1-A complex. *Right*: Gel shift assay of oligo $T(G)_4$, bearing a deletion of nucleotides from +125 to +129, was tested with nuclear extract from SW-13 cells. The deletion completely eliminated Sp3 binding

and did not affect 3 1-A protein binding. **B** The nucleotide sequence of the 3 2S oligonucleotide and its mutants employed in the study is shown. Small letters reveal the mutated nucleotides. The name of each oligo is indicated on the left side of each sequence. The sequence shown at the bottom summarizes the analysis results, with the minimal nucleotides required for 3 1-A factor binding showed in bold capital letters, and the minimal nucleotides required for Sp3 binding showed in cursive capital letters. Dotted lines represent the flanking nucleotides not involved in direct binding. The ability of each oligo to bind YY1 and Sp3 is summarized. +, binding; -, no binding; +/-, partial binding.

Fig. 4: The 3 1-A protein is antigenically related to transcription factor YY1. **A** EMSA: 3 g of nuclear extracts from the indicated cell lines were pre-incubated on ice with 100 fold molar excess of self and UCR unlabelled oligomers for 10 minutes and then added of 100,000 cpm of labeled 3 2S oligonucleotide. Complexes were allowed to form for 15 minutes at room temperature, and the binding reactions were loaded on a 4.5% non-denaturing polyacrylamide gel and subjected to electrophoresis to allow separation of the complexes from the unbound probe. The UCR-containing oligomer prevented formation of the 3 1-A band with the same efficiency observed for the 3 2S self-competition, while it did not affect Sp3 binding. **B** g of nuclear extracts from SW-13 cells were pre-incubated with and without 1 l anti-YY1 polyclonal antibody in binding buffer for 1 hour on ice. Then the reactions were added with either UCR or 3 2S labeled oligomers and further incubated at room temperature for 15 minutes. UCR self-competition is shown in the far right lane to confirm band specificity. Both 3 2S and UCR probes shifted a complex of identical electrophoretic mobility. In samples added of anti-YY1 antibody (YY1) the band corresponding to the 3 1-A complex was supershifted, suggesting that YY1 and the 3 1-A factor are the same protein. The two arrows between the two gel shift results point to the 3 1-A band here shown as YY1. **C** The sequences of the oligomers encompassing the 3 1-A element and UCR utilized in the experiment are shown. The region known to bind YY1 in the UCR probe, and the

homologous region in the 3 2S probe are shown in bold letters. The arrow indicates orientation of the YY1 binding region as previously proposed.

Fig. 5: The 3 1-A band corresponds to full-length YY1. **A**: 1 g of *pCMX*-YY1 expression plasmid was added to 50 l of TNT rabbit reticulocytes cell-free lysates in presence of T3 polymerase as described in Materials and Methods. Five l of the "*in vitro*" transcribed and translated product and same volume of a no-template reaction as negative control were incubated in gel shift binding buffer with labeled UCR oligomer, with and without self-competition, and the electrophoretic mobility of the bound probe was compared with the complexes obtained from 3 g of SW-13 cell nuclear extracts in a gel retardation assay. The "*in vitro*" translated YY1 shifted the UCR probe generating one specific complex, whose mobility was identical to the band observed with the UCR probe when reacted with nuclear extracts from SW-13 cells. **B** "*in vitro*" synthesized YY1 was compared with YY1 protein from cell nuclear extracts by western blot analysis. Lane 1, 10 l TNT control reaction in presence of empty *pCMX-L1* vector; lane 2, 10 l of TNT reaction on *pCMX*-YY1; lane 3, 20 g of SW-13 nuclear extract; lane 4 and 7, molecular weight standards (Santa Cruz, sc-2031) recognized by the secondary antibody; lane 5, 20 l of crude isolate on A-3 2-bound proteins from SW-13 cell nuclear extract; lane 6 and 8, 20 g of nuclear extract from LNCaP cells. Samples from lane 1 to 7 were treated with rabbit polyclonal anti-YY1 antibody direct against the full-length protein. The sample in lane 8 was treated with rabbit polyclonal anti-Sp3 antibody directed against the carboxy terminus. **C** Comassie staining of 12% SDS acrylamide gels. *Left*, 20 l of A-3 2-bound proteins isolated from 300 g of nuclear extracts from SW-13 cells; *Right*, part of the low range Biorad pre-stained SDS-PAGE standards are shown. Asterisks mark the position of the two bands recognized by anti-YY1 antibody in the western blot experiments.

Fig. 6: Existence of a second YY1 binding site within the HSD3B1 and HSD3B2 gene Intron 1. A EMSA experiment carried out with 5 g of SW-13 nuclear extracts and 100,000 cpm of each of the

oligomers indicated on top. Competition with 100 fold molar excess of UCR and Sp1 oligomers is also shown for the 3 2S probe. -, no competition. An arrow points at the YY1 formed band, previously known as 3 1-A band. **B** The double stranded sequence of the oligomers used in the study is shown. The name of the mutant oligomers specifies the first and last nucleotide of the intron 1 sequence they encompass. (I), (II) identifies oligos matching to type I or type II introns, respectively. The YY1 binding core in the 3 1-A element shared by the 3 1 and 3 2 oligomers is shown first as a reference, with dotted lines replacing the 5' and 3' flanking nucleotides. The black arrow indicates orientation of the YY1 binding site. Dotted lines encase the putative YY1-like binding region in each oligomer for easier sequence comparison. The ability of each oligomer to bind YY1 is summarized. ++, strong binding; + binding; +/-, weak binding; -, no binding. **C** *Upper*: A cartoon summarizes the different organization of the two YY1 binding sites in respect of the TG box in intron 1 of the HSD2B1 and HSD3B2 gene. Solid and dashed boxes represent stronger and weaker YY1 binding respectively. An open or filled square box represents the weaker or stronger binding of Sp3 to the TG box, respectively. A long and narrow rectangular box represents the 24 nucleotides previously identified as the 3 1-A element in the HSD3B1 gene intron 1. A similar schematization is used to represent the not perfectly conserved region in the HSD3B2 gene. *Lower*: The sequence alignment of the type I and type II intron 1 from nt +111 to nt +174 (type I) or to nt +173 (type II) shows the nucleotide changes involved in the formation of the second and not conserved YY1 binding site. Small letters identify the nucleotide substitutions, dotted lines replace missing nucleotides. Bold letters outline the nucleotide essential for Sp3 binding. Arrowheads identify the nucleotides involved in YY1 binding and the orientation of such binding. Bold solid arrowheads indicate strong binding, the bold open line arrowhead indicate a weaker YY1 binding.

Fig. 7. Functional analysis of YY1 in the HSD3B2 gene intron 1 mediated enhancement of basal promoter activity (A) YY1 is responsible for the boost in promoter activity mediated by the HSD3B2 intron 1. SW-13 cells were plated at 3.2 x 10^5 cells per well in 12-well plates. Sixteen hours later cells

were transfected as detailed in Materials and Methods with 0.5 g of *pGL3*-HSD2(-517/+193) and *pGL3*-HSD2(-517/+23) reporter constructs, previously shown in Fig 1 B, in presence or absence of 0.6 and 1.5 g of *pCMX*-YY1, and the total amount of DNA was kept at 2 g per well by addition of *pCMX* backbone vector where necessary. Cells were harvested 24 hours after transfection and the promoter activity was measured as relative luciferase units (rlu). The result of a representative experiment is shown, with the promoter activity expressed as rlu mean of duplicate wells. The standard deviations are shown. Each experiment was repeated at least 3 times. The YY1 dose-dependent boost in basal promoter activity seen with *pGL3*-HSD2 (-517/+193) was lost when *pGL3*-HSD2 (-517/+23) was used, which lacks intron 1. **(B)** Functional analysis of YY1 binding sites and TG box in the context of the HSD3B2 gene intron 1. The activity of luciferase reporter plasmid *pGL3*-HSD2(-246/+193) is compared with the activity of its deletion constructs that are represented in C. SW-13 cells were transiently transfected as described above except no YY1 expression plasmid was added. **(C)**. Schematization of the nucleotide deletion and substitutions involved in the study: 1. *pGL3*-HSD2(-246/+193) parental vector; 2. The B2mS substitution changes A5 to CCTTC eliminating YY1 binding site to the 3 1-A element; 3. Deletion $T(G)_4$ eliminates Sp3 binding to the TG box without affecting YY1 binding to the 3 1-A element; 4. Substitution of the 4 adenines to CCTT prevents YY1 binding to the distal binding site; 5. YY1 binding to both sites was prevented by constructing a double mutant using the same nucleotide substitutions described in 2. and 4.

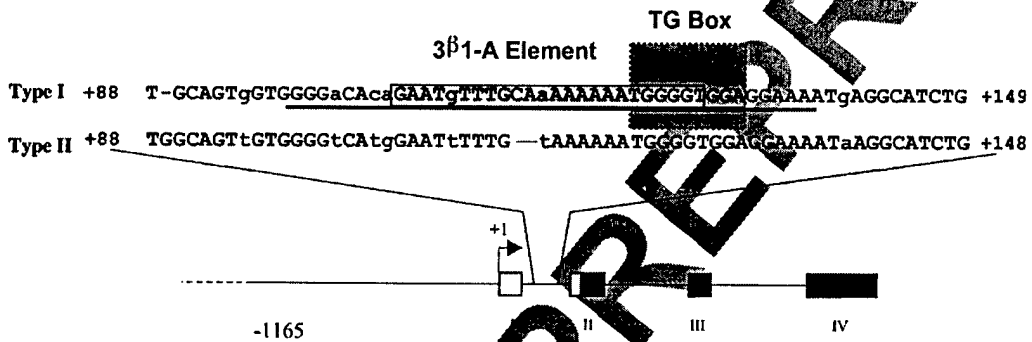


Fig. 1A

B

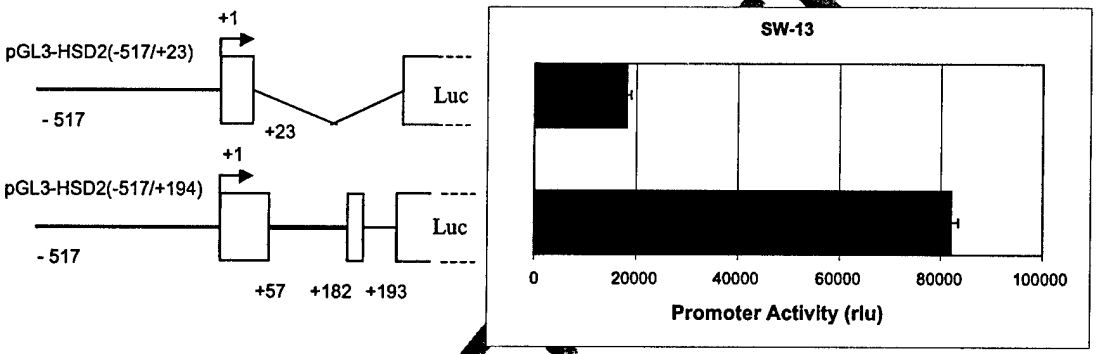


Fig. 1 B

A

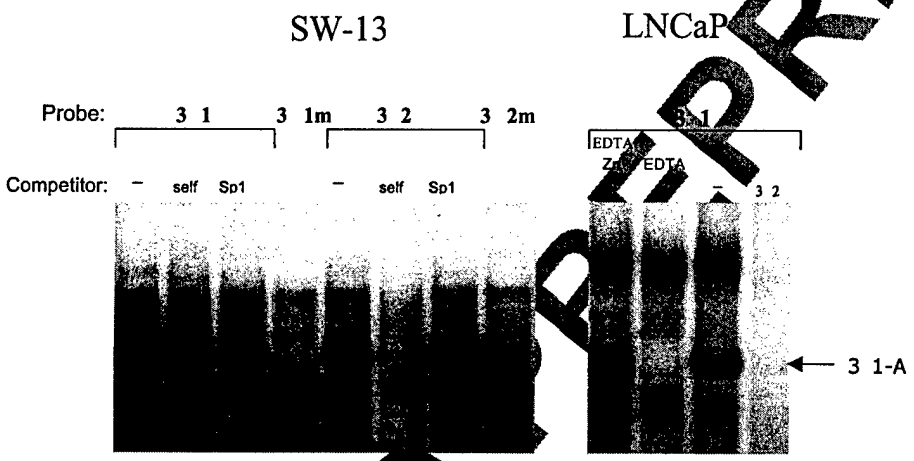
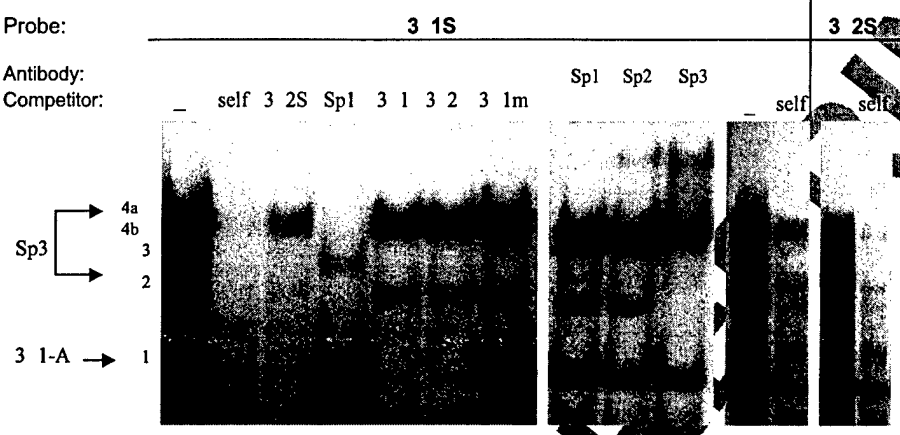


Figure 2 A

B



C

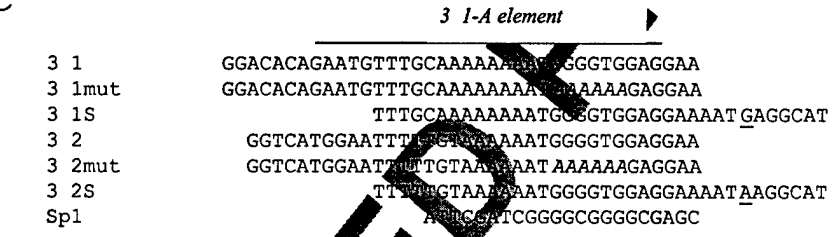
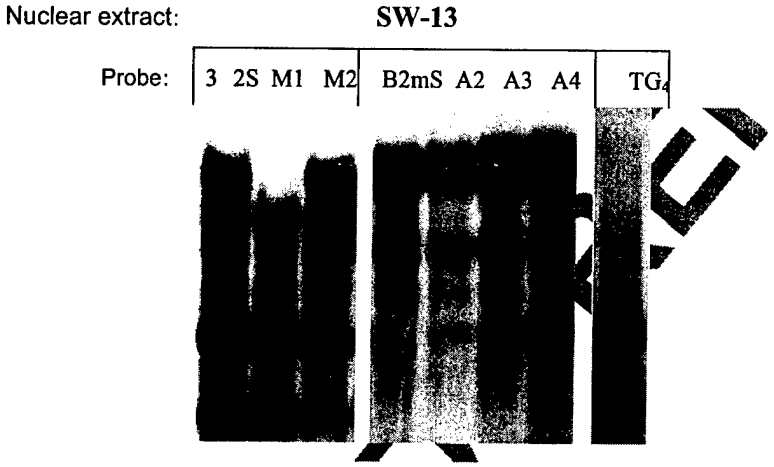


Fig. 2 B, C

A



B

		3 1-A	Sp3
3 2S	TTTTTGTAAAAAATGGGGTGGAGGAAAATAAGGCAT	+	+
3 2mS	TTTTTGTAccttcTGGGGTGGAGGAAAATAAGGCAT	-	+
A_2	TTTTTGTAcctAATGGGGTGGAGGAAAATAAGGCAT	-	+
A_3	TTTTTGTActAAATGGGGTGGAGGAAAATAAGGCAT	-	+
A_4	TTTTTGTAcAAAATGGGGTGGAGGAAAATAAGGCAT	+	+
M1	TTTTTGTAAAAAATGGaaTGGAGGAAAATAAGGCAT	+	-
M2	TTTTTGTAAAAAATaaGGTGGAGGAAAATAAGGCAT	-/+	+
TG_4	GGAATTTTTGTAAAAAA-----TGGAGGAAAATAAGGCATCTGC	+	-

------**AAAATGG** *GGTGG* ------

Fig. 3

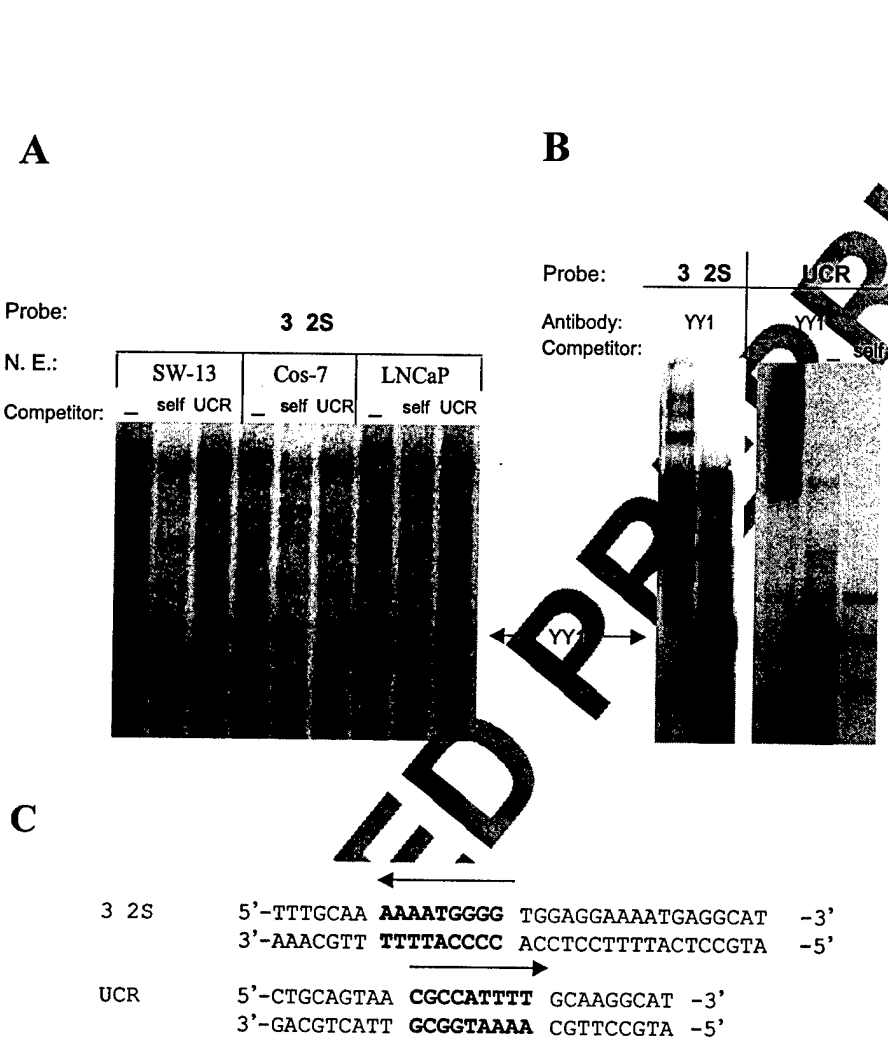


Fig. 4

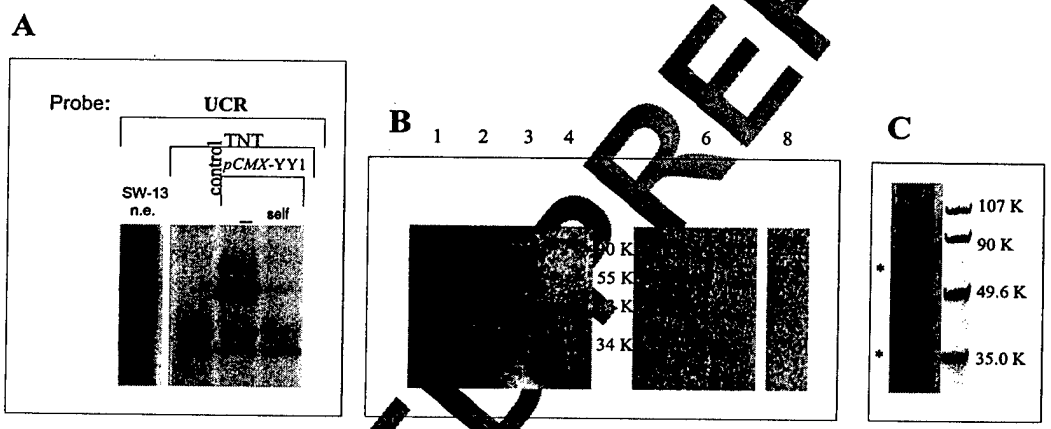
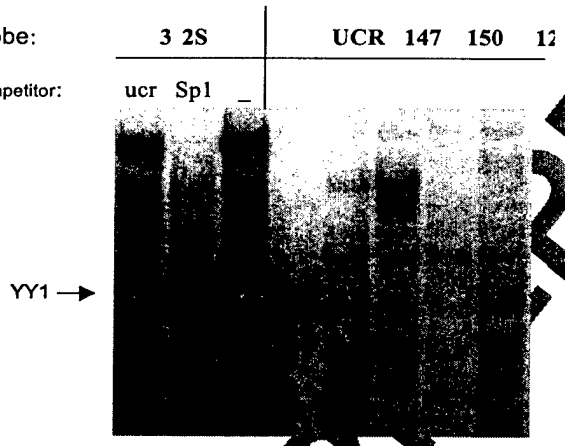


Fig. 5

A



B

		YY1 binding
3 1-A/YY1	- - - AAAATGGGG - - - - -	++
UCR	CTG[illegible]GTAA[illegible]GCCATTTTGCAAGGCAT [illegible]ACGT[illegible]TTCGGTAAAACGTTCCGTA	++
147/175(II)	[illegible]TGAGTG[illegible]TAAACCATTTTACCTCTTGTT GACTCA[illegible]ATATTGGTAAAATGGAGAACAA	+
150/178(I)	CTGAGTATATAACCATTTGACATCTCTTT GACTCATATATTGGTAAACTGTAGAGAAA	-
125/150(II)	GGTGGAGGAAAATAAGGCATCTGCTG CCACCTCCTTTTATTCCGTAGACGAC	-
126/151(I)	GGTGGAGGAAAATGAGGCATCTGCTG CCACCTCCTTTTACTCCGTAGACGAC	+/-

Fig. 6 A and B

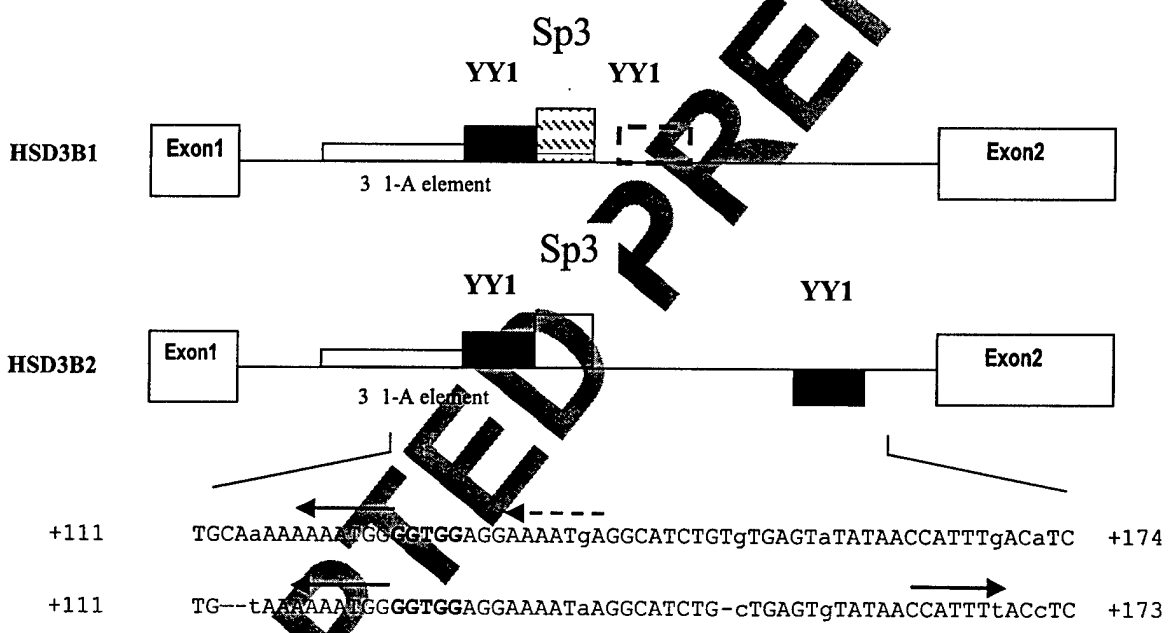


Fig. 6 C

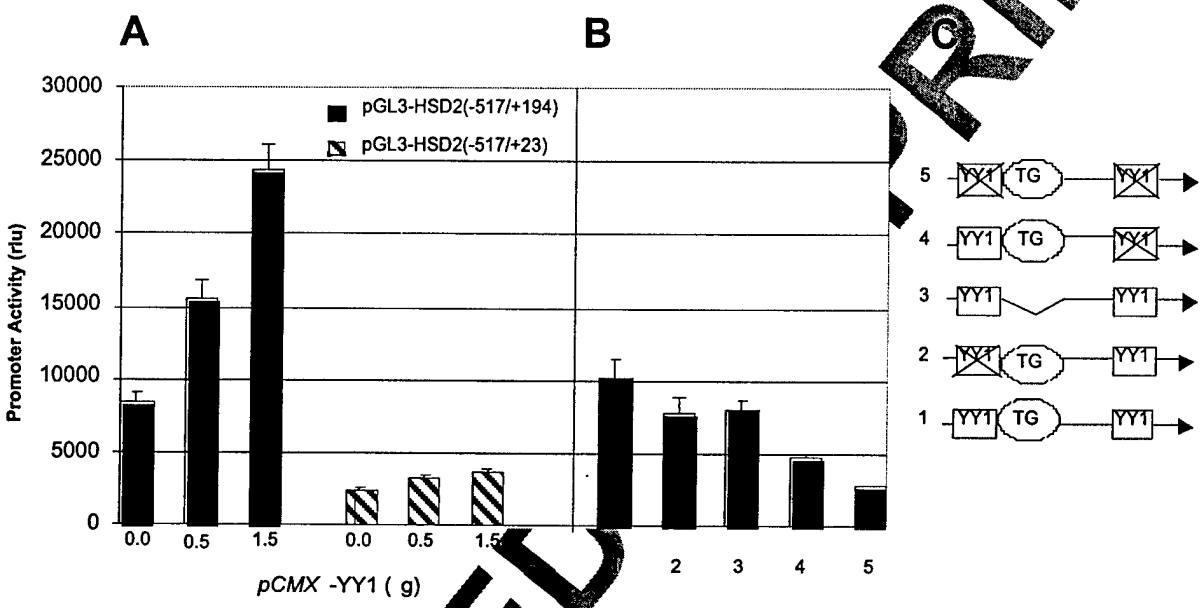


Fig. 7