Full-length SRY protein is essential for DNA binding

I. Sánchez-Moreno1*, R. Coral-Vázquez2, J.P. Méndez1,3 and P. Canto1,4

1Unidad de Investigación Médica en Biología del Desarrollo, Centro Médico Nacional Siglo XXI, Coordinación de Investigación en Salud, Instituto Mexicano del Seguro Social, C.P. 06720 México, D.F., Mexico; 2Unidad de Investigación Médica en Genética Humana, Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, C.P. 06720 México, D.F., Mexico; 3Present address: Unidad de Investigación en Obesidad: Facultad de Medicina, Universidad Nacional Autónoma de México; Clínica de Obesidad, Instituto Nacional de Ciencias Médicas y Nutrición ‘Salvador Zubirán’, México, D.F., Mexico; 4Correspondence address. Tel/Fax: +52-55-5627-69-00; E-mail: ipcanto@yahoo.com.mx

**SRY** directs testicular development. It has been suggested that the only high-mobility group (HMG) box of the SRY is important for the function of this protein; however, other studies have suggested that the N- and C-terminal regions are also involved in this process. Herein, we analysed and compared *in vitro* the DNA-binding activity of the full-length SRY and three mutants (HMG box alone, N-terminal less and C-terminal less SRY proteins). DNA-binding capability was analysed by mobility shift assays, optical density and dissociation constant by using pure non-fusion SRY proteins. The structure of the full-length SRY was carried out using a protein molecular model. The HMG box SRY alone and C-terminal less SRY proteins had a statistically diminished DNA binding in comparison with the full-length SRY. In contrast, the affinity for DNA of the N-terminal less SRY was relatively similar to the full-length SRY. Likewise, three-dimensional structure of the full-length SRY suggested that some residues of the C-terminal region of the SRY interact with DNA. We demonstrate the importance that full-length SRY has, particularly the C-terminal region of the protein, in DNA binding *in vitro*. Likewise, the affinity of the HMG box alone is clearly reduced when compared with the full-length SRY.

*Keywords*: DNA binding; full-length SRY; HMG box SRY; N-terminal less SRY; C-terminal less SRY

**Introduction**

Sex determination and differentiation are sequential processes regulated by an unknown number of gene loci located on sex and autosomal chromosomes. There is wide evidence that the SRY (sex-determining region of the Y chromosome) ‘GenBank accession: L10102’, directs testicular development (Su and Lau, 1993). This gene is located in Yp11.3 and encodes a transcription factor believed to activate and/or repress target genes. Under the influence of this genetic switch, the indifferent gonad differentiates into testis-specific cell types and organizes into testis-specific morphology (Koopman et al., 1990).

Human SRY is a 204 amino acid nuclear protein containing a central high-mobility group (HMG) box domain, as well as amino-terminal (N-terminal) and carboxy-terminal (C-terminal) regions (Clepet et al., 1993; Su and Lau, 1993). The HMG box domain is highly conserved among mammalian species; whereas variation exists, both in length and sequence, in the N- and C-terminal regions (Tucker and Lundrigan, 1993; Whitfield et al., 1993). The importance of the HMG box has regarding SRY function is highlighted by the fact that most 46,XY patients with sex reversal have been described in the N- and C-terminal regions of this gene (Tajima et al., 1994; Domenice et al., 1998; Mitchell and Harley, 2002; Baldazzi et al., 2003; Shahid et al., 2004; Gimelli et al., 2007).

Since SRY was described, it has been suggested that those regions flanking the HMG domain make little or no contribution to the protein’s interaction with DNA (Ferrari et al., 1992; Mitchell and Harley, 2002). However, several mutations signaled as causative of 46,XY sex reversal have been described in the N- and C-terminal regions of this gene (Oh et al., 2005). The role of the N-terminal region upon SRY function was originally described by Descoleaux et al. (1998) these authors demonstrated that the cyclic AMP-dependent protein kinase phosphorylates the SRY *in vitro*, as well as *in vivo*, on serine residues located in this region. This phosphorylation event was shown to positively regulate SRY DNA-binding activity. In 2002, Assumpcao et al. reported a mutation in the N-terminal region of the SRY, within the phosphorylation site. The authors demonstrated that the mutant SRY was poorly phosphorylated and consequently showed reduced DNA-binding capacity *in vitro*. Regarding the C-terminal region, Poulat et al. (1997) described, for the first time, an interaction between this region and the PDZ domain protein SIP-1/NHERF2, suggesting that this protein could permit the connection of SRY to another protein harboring a PDZ binding motif. Several years later, Thevenet et al. (2005) demonstrated that the interaction between SRY and SIP-1/NHERF2 proteins is conserved in humans and in mice. A recent study (Oh et al., 2005) identified a novel protein, KRAB-O, an SRY-interacting protein located specifically in the C-terminal region of the human and mouse SRY/Sry.

1This work was submitted in partial fulfillment of the requirements for the D.Sc. degree of Sánchez-Moreno I, for obtaining her degree in the Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México.

© The Author 2008. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved.

For Permissions, please email: journals.permissions@oxfordjournals.org
In order to determine precisely the significance that both regions located outside of the HMG box domain of SRY have regarding DNA binding, we analysed and compared this function using the full-length SRY and three SRY mutants (HMG box alone, N- and C-terminal less SRY), with pure SRY proteins. In addition, we analysed the full-length SRY structure by using a protein molecular model.

Material and Methods

Construction of expression plasmids

For glutation-S-transferase (GST) fusion constructs, amplified human DNA encoding the complete open reading frame of SRY (full-length SRY), the HMG box SRY box alone (N- and C-terminal less SRY, amino acids 58–137), the N-terminal-HMG box SRY (C-terminal less SRY, amino acids 1–137) and the HMG box-C-terminal SRY (N-terminal less SRY, amino acids 58–204) were ligated into the BamHI/EcoRI restriction site of pGEX-3X (Amersham Pharmacia Biotech, L. Ch. B, England, UK) (Fig. 1a).

For this purpose, sets of primers carrying BamHI and EcoRI tags were designed according to each sequence to produce GST-fused proteins. Primer sequences are as follows: forward SRY-BamHI 5′-GGCCGGATCCCTATGTGATATGCTGATGCT-3′ and reverse SRY-EcoRI 5′-GGCCGGATCCCAATGTTACCTAGGTGTT-3′ for full-length SRY; forward SRY-BamHI and reverse SRY B-EcoRI 5′-GGCGGGATCCGTAGCAGTTAATCCGATTGTC-3′ for C-terminal less SRY; forward SRY C-BamHI 5′-GGCCGGATCCCTCCCATGATGGACTGCGCGCCTAGAA C-3′ and reverse SRY-EcoRI for N-terminal less SRY; SRY C-BamHI and SRY B-EcoRI for HMG box alone. PCR was carried out using 400 ng of genomic DNA, 0.4 μM of each primer, 200 μM of each dNTP and 1.5 U of Taq polymerase (AmpliTaq, Applied Biosystems, Foster City, CA, USA) in a volume of 50 μl. After 5 min of the denaturing step, amplification was performed as follows: 94°C for 1 min, 68°C for 1 min and 72°C for 1 min for 30 cycles and one step at 72°C for 4 min. PCR products were electrophoresed on 1% agarose gels stained with ethidium bromide, purified by the Qiaex II Gel extraction kit (Qiagen GmbH, Hilden, Germany), and cloned into BamHI/EcoRI sites of pGEX-3X bacterial expression vector. Resulting constructs were confirmed by DNA sequencing using the ABI 377 automated DNA sequencer (Perkin–Elmer, Applied Biosystems Division, Foster City, CA, USA) and BigDye terminator cycle sequencing ready reaction kit (Perkin–Elmer).

Expression and purification of SRY proteins

All expression vectors were transformed in BL21 Escherichia coli strain. The recombinant proteins were expressed and purified as described previously (Schmitt-Ney et al., 1995). Cells harboring the different pGEX-3X-SRY recombinant vectors were grown at 30°C in 250 ml of LB Broth (1% NaCl, 1% Bacto-Tryptone, 0.5% Bacto-yeast extract, pH 7.0) containing 100 μg/ml of ampicillin. When OD600 reached 0.6, IPTG was added to a final concentration of 1 mM, and the culture was incubated for two more hours at 30°C in a shaking chamber. Cells were centrifuged (10 000 g, 4°C, 10 min), resuspended in PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na2PO4, 1.8 mM KH2PO4, pH 7.0) and 1% 100X Triton and incubated for additional 30 min on ice. The suspension was sonicated and the cell debris was removed by centrifugation. The GST–SRY fusion proteins were purified by sepharose columns according to the provider instructions (Amersham Pharmacia Biotech). Full-length SRY and its mutants were released from GST by FaXa digestion (Qiagen GmbH), and then purified with sepharose columns (Qiagen GmbH).

Western blot analysis

The crude extracts, the recombinant purified fusion and the non-fusion proteins were resolved in 10% SDS polyacrylamide gel. Proteins were electrobotted to a PVDF membrane (Amersham Pharmacia Biotech). The blot were probed with an anti-GST primary antibody in a 1:10 000 dilution (Vector Laboratories, Burlingame, CA, USA) in presence of 5% non-fat dry milk (blocking solution). Membranes were revealed with a horseradish peroxidase conjugated secondary antibody diluted 1:1000 (Vector Laboratories, Inc., Burlingame, CA, USA) and a chemiluminiscence detection system, ECL western blotting analysis system (Amersham Biosciences).

Electrophoresis mobility shift assay

Synthetic complementary oligonucleotides containing the consensus binding site for SRY (5′-GGCCGGATCCCTAGGATGCTGATGCTG-3′; bold sequence indicates the SRY–DNA-binding site) (Ferrari et al., 1992), and the mutant probe (5′-GGCGGATCCCTAGGATGCTGATGCTG-3′) (Invitrogen Carlsbad, CA, USA) were prepared as described by Bellon et al. (1991). Briefly, 1.76 pmol of complementary strands were annealed, gel purified and radiolabeled using 10 U of T4 polynucleotide kinase (New England Biolabs, Beverly, MA, USA) in presence of 1 μCi [γ32P]-ATP (Perkin–Elmer Life Sciences Inc., Boston, MA, USA) and purified by sephadex G-25 columns (Amersham Biosciences). Analysis of electrophoresis mobility shift assay (EMSA) for each kind of SRY proteins were performed as described previously (Jäger et al., 1992; Cohen et al., 1994; Schmitt-Ney et al., 1995; Lencnicz-Allen et al., 1996; Desclozeaux et al., 1998; Assumpcao et al., 2002; Mitchell and Harley, 2002; Thevenet et al., 2005). Briefly 10 nM of pure protein from SRY and its protein mutants were mixed with EMSA buffer (40 mM HEPES, 100 mM KCl, 100 mM MgCl2, 1% glycerol and 1 mM DTT) and 1 μg poly(dI–dC) as non-specific competitor. After 30 min at room temperature, 1.76 pmol of double-strand radiolabeled DNA was added to the reaction and the final...
mixture was incubated for 30 additional min at room temperature; samples were electrophoresed in a non-denaturating 10% polyacrylamide gel at 150 V for 6 h, at 4 °C. In competition experiments non-labeled specific DNA competitor was added together with the poly(dI–dC) and the super shift test was performed with 1 µl of anti-GST antibody (Amersham Pharmacia Biotech) Affinity for full-length SRY and for each one of the SRY mutant proteins, without GST in all cases (pure non-fusion protein), was determined by the dissociation constant (kd) as referred by Trimmer et al. (1998). Different concentrations of pure proteins (1–2000 nM) and 218 fmol of radiolabeled probe were handled as the binding reaction previously described. Optical density of bands was obtained by an Alpha Imager Software from Alpha Imager Gel Documentator (Alpha Innotech Corporation, San Leandro, CA, USA). Statistical analyses were carried out using one way ANOVA and a Dunnett test with the SPSS v13.0 for Windows (SPSS, Chicago, IL, USA). The dissociation constant (kd) was defined as the protein concentration required for half-maximal binding. The kd calculated for each protein–DNA interaction is presented as an average of three independent experiments.

Full-length SRY homology modeling

The three-dimensional model of the full-length SRY was obtained using the Robetta server (http://robbetta.bakerlab.org/) (Kim et al., 2004). Ten models of the full-length SRY structure were obtained and all of them were validated in the Molprobity server (http://molprobity.biochem.duke.edu/) (Davis et al., 2007), after one or two energy minimization processes, using the Chimera program (Pettersen et al., 2004). The three-dimensional models of the full-length SRY were annealed with the three-dimensional structure of the HMG box domain of the SRY (Protein Data Bank, accession number: 1J46), using the Pymol (http://pymol.sourceforge.net/index.html) and the Superpose v1.0 computational servers (http://wishart.biology.ualberta.ca/SuperPose/) (Maiti et al., 2004). The root mean square distance (RMSD) values were computed and the molecular surface analyses were performed using the Castp server (http://sts-fw.bioengr.uic.edu/castp/calculation.php) (Binkowski et al., 2003).

Results

Before the EMSAs were carried out, the integrity, as well as the concentration of all recombinant proteins, was verified by means of a Coomassie staining gel. Crude extracts and GST fusion proteins were analysed by immunoblotting, using an anti-GST antibody. The expected immunoactive bands for GST alone (26 KDa), GST-full-length SRY (50 KDa), GST-HMG box alone (37.5 KDa), GST-C-terminal less SRY (46.8 KDa) and GST-N-terminal less SRY (44.5 KDa) were observed in both the crude extracts and the purified fusion proteins (data not shown). Full-length SRY and its mutants were released from GST and the pure non-fusion proteins were obtained and electrophoresed in a 12% SDS polyacrylamide gel. The expected bands for full-length SRY (24 KDa), HMG box alone (11.5 KDa), C-terminal less SRY (20.8 KDa), N-terminal less SRY (18.5 KDa) and GST alone (26 KDa) were observed (Fig. 1b, top).

The EMSA analyses (Fig. 1b, EMSA) of the full-length SRY (lane 2) and of the three corresponding SRY mutants (lanes 3–5) obtained from the pure non-fusion proteins revealed the formation of DNA–protein complexes; whereas, the free probe (lane 1) and the GST-protein (lane 6) showed no binding. The C-terminal less SRY (lane 4) and the N-terminal less SRY (lane 5) formed similar DNA–protein complexes to the ones observed with full-length SRY (lane 2); however, the HMG box alone (lane 3) only formed one DNA–protein complex. The slower moving bands corresponding to DNA complexes of the C-terminal less SRY were clearly less intense in comparison with the ones observed with the full-length SRY. In contrast, the slower moving complex band of the N-terminal less SRY was slightly less intense than those present in full-length SRY.

Specificity of each band pattern complex was demonstrated by specific competition with a non-labeled DNA probe, for the full-length SRY and for the three SRY mutant proteins, as well as with a super shift test (data not shown).

In order to evaluate the percentage of DNA binding of all three pure SRY mutant proteins, band intensity was measured in arbitrary optic density units, and comparisons were performed against full strength SRY (100%) (Fig. 1c). The HMG box alone and C-terminal less SRY, showed a diminished binding (35 and 84%, respectively); in both cases, this difference was statistically significant (P < 0.005 and 0.0001, respectively). Additionally, when we used the N-terminal less SRY, we found an increased binding activity; however, this result was not statistically significant.

Furthermore, to compare differences between DNA-binding activities of all three SRY mutant proteins with the full-length SRY, the dissociation constant was determined and plotted (Fig. 2). DNA-binding affinity of the full-length SRY was 57 nM (panel and plot SRY). The HMG box alone reduced protein DNA-binding affinity less than 1-fold, to 73 nM (panel and plot HMG), the C-terminal less SRY reduced it roughly, 2.5-fold, to 123 nM (panel and plot C-TER), and 58 nM for the N-terminal less (plot N-TER).

Figure 2: Determination of the dissociation constant (kd). EMSAs of the full-length SRY (Panel SRY), HMG box alone (Panel HMG), C-terminal less (Panel C-TER), and N-terminal less (Panel N-TER) obtained from pure non-fusion SRY (0–2000 nM). DNA–protein complexes (*) and free probes (**) are displayed. The kd was defined as the protein concentration required for half-maximal binding and was calculated for each protein–DNA interaction and presented as average of three independent experiments. DNA-binding affinity of the full-length SRY was 57 nM (plot SRY), for the HMG box alone, binding affinity was 73 nM (plot HMG), 123 nM for the C-terminal less (plot C-TER), and 58 nM for the N-terminal less (plot N-TER).
Sánchez-Moreno et al.

whilst the N-terminal less SRY had no detectable difference on DNA-binding activity ($k_d = 58 \text{nM}$) (panel and plot N-TER).

Based on the energy minimization and validation process of the three-dimensional structure model of the full-length SRY, one out of 10 models was selected (data not shown). The best model was structurally annealed to the HMG box, and to the local alignment of the Gln57 to Leu138 residues showing a Cα root mean square deviation (RMSD) of 0.79 Å (Fig. 3a). The superposing structures show that DNA threads the SRY model through the Gln193, Arg197 and Trp201 residues of the C-terminal region (Fig. 3b). Furthermore, protein surface analyses suggested the formation of a large DNA interacting cavity, by the residues that contact the DNA located in the HMG box region, as well as by some positive polar residues located in the C-terminal region of the SRY (Fig. 3c).

**Discussion**

SRY pertains to a family of SRY-related HMG-box proteins, designated SOX. It is well known that SOX proteins bind sequence-specifically to DNA by means of a HMG box domain, allowing them to function as transcription factors. This domain is highly conserved among SOX proteins and all of them appear to recognize a similar binding motif. Therefore, these proteins may present some additional features in sequences outside of the HMG box that may confer DNA sequence recognition by different SOX proteins, to their specific target genes either activated or repressed (Wilson and Koopman, 2002).

On this basis, in the present study we compared by EMSA, optical density and dissociation constant, the role accomplished by the full-length, the HMG box alone and the N- and C-terminal regions of SRY upon DNA-binding activity. The assays were undertaken with pure non-fusion SRY proteins.

Most other studies on Sox in the past had been restrained to the use of the isolated HMG box domain in EMSA. Ferrari et al. (1992) studied the mechanics of interaction of HMG boxes with DNA. These authors used full-length-SRY and SRY-HMG box alone, finding that both proteins bind to DNA, therefore suggesting the use of full-length SRY or SRY-HMG box alone interchangeably in DNA-binding studies. In the present study, we found, by EMSA, similar results observing that full-length SRY and its three mutant SRY proteins bind to the DNA. However, we observed that the affinity for DNA of HMG box alone and C-terminal less SRY, displayed a diminished DNA binding in comparison with the full-length SRY, and this difference was statistically significant. These findings were confirmed with the percentage of DNA binding and by dissociation constant. Our results confirm that besides the fact that the HMG box is required for SRY binding, the C-terminal region of SRY plays an important role in DNA-binding activity.

In contrast to the latter observation, affinities of the N-terminal less SRY were relatively similar to the ones observed with full-length SRY, suggesting that the N-terminal region of SRY does little on its own to affect binding to DNA *in vitro*.

Moreover, Ferrari et al. (1992) showed that purified SRY-HMG box alone formed only one retarded band in comparison with the two retarded bands shown in the full-length SRY. Similarly, when we used the HMG box alone obtained from pure non-fusion proteins, we observed only one retarded band and when comparing to full-length SRY obtained from pure non-fusion proteins, the slower moving complexes were not detected. Likewise, this pattern has been observed by Peirano and Wegner (2000) in another Sox gene (Sox 10). These authors observed that when they used only the HMG box of Sox 10, a monomeric DNA binding (one retarded band) was disclosed. In addition, we found that the slower moving complex bands of the C-terminal less SRY obtained from pure non-fusion proteins were clearly less intense in comparison with those observed with the full-length SRY. When examining these bands by super shift assay, we observed that these slower DNA-complexes were super shifted in full-length, HMG box and C-terminal less SRY proteins. Taking these findings together, we suggest that the slower retarded band observed in the EMSA of SRY and SRY mutants, might be two or more SRY molecules and the presence of these molecules might contribute to SRY function, as has been suggested in other Sox proteins (Peirano and Wegner, 2000).

On the other hand, in contrast to the observations with the C-terminal less SRY proteins, the intensities of the slower moving complexes of the N-terminal less SRY proteins were relatively similar to the full-length SRY, being this result not statistically significant.

It has been well established that although the HMG box domain of the SRY is required as a DNA–protein interface, sequences outside of this domain are required to stabilize protein binding and/or generate specificity by helping to discriminate between protein partners.
DNA-binding activity studies of the SRY protein


Submitted on February 14, 2008; resubmitted on April 14, 2008; accepted on April 16, 2008.