Functional and Structural Studies of Wild Type SOX9 and Mutations Causing Campomelic Dysplasia*

(Received for publication, March 5, 1999, and in revised form, May 19, 1999)

Sharon McDowall‡§, Anthony Argentaro‡, Shoba Ranganathan‡, Polly Weller‡, Sabine Mertin‡, Sahar Mansour**‡, John Tolmie‡‡, and Vincent Harley‡

From the ‡Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville 3052, Australia, the §Australian Genome Information Centre, C80 University of Sydney, Sydney, New South Wales 2006, Australia, the †Incyte Genetics, 214 Cambridge Science Park, Milton Road, Cambridge CB4 4WA, United Kingdom, the **Department of Clinical Genetics, St. George’s Hospital Medical School, Tooting, London SW17 0RE, United Kingdom, and the ‡‡Department of Medical Genetics, Yorkhill Hospitals National Health Service Trust, Yorkhill, Glasgow G3 8SJ, United Kingdom

In humans, mutations in SOX9 result in a skeletal malformation syndrome, campomelic dysplasia (CD). The present study investigated two major classes of CD mutations: 1) point mutations in the high mobility group (HMG) domain and 2) truncations or frameshifts that alter the C terminus of the protein. We analyzed the effect of one novel mutation and three other point mutations in the HMG domain of SOX9 on the DNA binding and DNA bending properties of the protein. The F12L mutant HMG domain shows negligible DNA binding, the H65Y mutant shows minimal DNA binding, whereas the A19V mutant shows near wild type DNA binding and bends DNA normally. Interestingly, the P70R mutant has altered DNA binding specificity, but also bends DNA normally. The effects of the point mutations were interpreted using a molecular model of the SOX9 HMG domain. We analyzed the effects upon transcription of mutations resembling the truncation and frameshift mutations in CD patients, and found that progressive deletion of the C terminus causes progressive loss of transactivation. Maximal transactivation by SOX9 requires both the C-terminal domain rich in proline, glutamine, and alanine and the adjacent domain composed entirely of proline, glutamine, and alanine. Thus, CD arises by mutations that interfere with DNA binding by SOX9 or truncate the C-terminal transactivation domain and thereby impede the ability of SOX9 to activate target genes during organ development.

In humans, mutations in SOX9 cause campomelic dysplasia (CD), a skeletal malformation syndrome that is often associated with XY sex reversal (1). Other tissues affected include kidney, heart, and brain, consistent with the expression pattern of Sox9 in developing mouse (2, 3). There are four major classes of mutations causing CD: 1) amino acid substitutions in the HMG domain (Fig. 1A), 2) truncations or frameshifts that alter the C terminus of SOX9 (Fig. 1B), 3) mutations at splice junctions, and 4) chromosomal translocations, of which classes 1 and 2 are investigated here. Most CD patients are heterogeneous for wild type and mutant alleles of SOX9. CD appears to result from haplosufficiency; presumably, a critical dose of SOX9 is required to switch on the appropriate genes during development. The present study reports the identification in a CD patient of a novel amino acid substitution mutation (H65Y) in the HMG domain of SOX9. We report the effects of this and three other point mutations (F12L, A19V, and P70R) on the DNA binding and DNA bending activities of the HMG domain.

SOX proteins represent a large class of transcription factors related to SRY, the testis-determining factor, through their HMG domains that bind and bend DNA in a sequence-specific manner. Expression of these proteins in defined cell types at specific stages of development appears to govern cell fate decisions. SOX9 activates expression of type II and type XI collagen in vitro (4–6), consistent with a role in bone development.

SOX proteins fall within a larger group of HMG domain proteins comprising two classes: 1) those that bind DNA without sequence specificity (such as HMG1, HMGD) and 2) those that bind DNA with sequence specificity (including the TCF1/LEF1 and SOX transcription factors). An amino acid sequence alignment of the SOX9 HMG domain with those of SRY and LEF1 is shown in Fig. 2. Although the three-dimensional structure of the SOX9 HMG domain is not known, the solution structures of the HMG domains of SRY (7) and LEF1 (8), in complex with DNA, have been determined by NMR. The fold of the two HMG domains is similar. The three α-helices of the each HMG domain come together in an L-shape in which the short arm is formed by helices 1 and 2 and the long arm by helix 3 and the N-terminal strand. The concave surface of the “L” contacts the minor groove of the DNA. We have constructed a model of the SOX9 HMG domain based on the solution structure of the SRY HMG domain and have used the model to make interpretations about the effects of point mutations within the HMG domain, on DNA binding. According to the model, three of the SOX9 point mutations studied here (F12L, H65Y, and P70R) occur in residues that lie on or near the DNA binding surface of the HMG domain, and might therefore be expected to affect DNA binding. The fourth mutation (A19V) affects a residue that is not on the DNA binding surface, but might be important in maintaining the structure of the protein.

The determinants of transactivation by SOX9 have not been fully defined. Many of the mutations that result in CD are...
truncations or frameshifts that alter the C terminus of the protein. We hypothesized that these mutations disrupt the transactivation potential of the protein and we sought to define the limits of the transactivation domain of SOX9 by deletion analysis. At the C terminus of SOX9 lies the PQS-rich domain (Fig. 1B; residues 386–509), a domain rich in proline, serine, and glutamine, which is required for transcriptional activation (9). Preceding this is the PQA domain (residues 339–379) that appears to have resulted fortuitously from exonuclease digestion by

### EXPERIMENTAL PROCEDURES

**Patient Reports**—Patient 10 is the third child of consanguineous Pakistani parents (half first cousins). One older brother died of congenital heart disease; an older sister and younger brother are both well. At birth the proband had macrocephaly, micrognathia, depressed nasal bridge, short limbs, curved femora, small patellae, bilateral talipes equinovarus, normal male genitalia, and mild thoracic kyphosis. Tracheomalacia caused severe respiratory distress and necessitated intubation with ventilation from birth. Radiological features included long iliac bones, bowed femora, straight tibiae, long fibulae, increased acetabular angle (hips not dislocated). Cytogenetic studies showed a normal female karyotype. Parental Haplotyping—Paternity and maternity of patient 10 were confirmed by DNA profiling using 12 fluorolucency labeled PCR primer pairs that amplify microsatellite markers (*beta* globin on Xq28) located on human chromosome 8, as described previously (10). Parental haplotypes were concordant with those of the proband.

### PCR Amplification and HSCPCR—To screen for the H65Y mutation among the family of patient 10, a portion of the SOX9 open reading frame was amplified from genomic DNA by PCR, using primers F and G, and analyzed by SSCP as described previously (10).

**Mutagenesis**—DNA sequences encoding mutant SOX9 HMG domains bearing point mutations were produced by PCR, with the mutation incorporated into one of the primers, or by amplification of patient DNA. Forward and reverse primers also bore Ndel and HindIII sites, respectively, to allow the PCR products to be inserted between the Ndel and HindIII sites in pT7-7. The sequences of all mutants were verified by DNA sequencing.

Deletion mutant SOX9(1–455) was produced by digestion of SOX9 pcDNA3 with PvuII and BstXI, removal of single-stranded termini with mung bean nuclease, and religation with T4 DNA ligase. With the aim of producing a series of nested deletion mutants, a PvuII and BstXI double digest of pcDNA3-SOX9 was treated with mung bean nuclease to remove single-stranded termini, and then with exonuclease III. Only mutant SOX9(1–410) was isolated in this manner, and it appears to have resulted fortuitously from exonuclease digestion by mung bean nuclease past the single-stranded overhang, removing 347 nucleotides upstream of the BstXI site. Other deletion mutants were created by digestion with restriction endonucleases and religation. Mutant SOX9(1–248) was produced by removal of a ApaI restriction fragment from pcDNA3-SOX9. This deletion closely mimics a CD mutation that results from a missense mutation at codon 251. Removal of a SfiI-EcoRV restriction fragment produced deletion mutant SOX9(1–437), which closely mimics a CD mutation resulting from a missense mutation at codon 440. Mutant SOX9(1–454) was produced by removal of sequences between the most 5′ BstUI site of SOX9 and the EcoRV site in the multiple cloning site of pcDNA3. Mutant SOX9(1–465) was produced by removal of sequences between the most 5′ BstUI site in SOX9 and the same EcoRV site. SOX9(3PQA) was produced by removal of a PmI and PvuII restriction fragment from SOX9.

**Production of Mutant and Wild Type HMG Domains**—The plasmids (pT7-7-SOX9 box) were transformed into *Escherichia coli* BL21 and expression of the SOX9 HMG domain was induced by IPTG and soluble protein extracts prepared (11). The HMG domains were expressed in *E. coli* at a level of approximately 15–45 mg/liter. The HMG domain used in this study extends from residues Asn101 to Asn184 of full-length SOX9, with the addition of a Met residue at the N terminus.

**Production of Mutant and Wild Type Full-length SOX9**—Full-length SOX9 was produced in vitro by coupled transcription and translation of SOX9 (wild type and deletion mutants) in pcDNA3, using a TNT kit (Promega), with incorporation of [35S]methionine.

**Electrophoretic Mobility Shift Assays**—Oligonucleotide probes were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer. The sequences of the upper strands are given below. S9WT sequence is GGTTAAGCGGAAATACTCTAATCTGGTAGTA. The high affinity SOX binding site is shown in bold. The high affinity SOX binding site is shown in bold. The four flanking residues (underlined) are those that are least preferred for binding of SOX9 in vitro (12). Col2c1 sequence is GGTCGAGGAAAAAGCCCATCTCCTGAGACC. Col2c2 sequence is GGTCGAAGAAAGCCCATCTCCTGAGACCC. Col2c2 sequence is GGTCGAGGAAAAAGCCCATCTCCTGAGACCC. Col2c1 and Col2c2 are SOX-mutant sequences from the Col2a1 enhancer that are required for chondrocyte-specific expression. In vivo, SOX proteins appear to tolerate considerable sequence variation in their binding sites. The sequences conform loosely to the HMG consensus binding site (A/T)(A/T)CAA(A/T)(A/T)CAA(A/T)CAA(A/T). The residues that correspond to this consensus are shown in bold.

**Electrophoretic Mobility Shift Assays**—Oligonucleotide probes were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer. The sequences of the upper strands are given below. S9WT sequence is GGTTAAGCGGAAATACTCTAATCTGGTAGTA. The high affinity SOX binding site is shown in bold. The high affinity SOX binding site is shown in bold. The four flanking residues (underlined) are those that are least preferred for binding of SOX9 in vitro (12). Col2c1 sequence is GGTCGAGGAAAAAGCCCATCTCCTGAGACC. Col2c2 sequence is GGTCGAAGAAAGCCCATCTCCTGAGACCC. Col2c2 sequence is GGTCGAGGAAAAAGCCCATCTCCTGAGACCC. Col2c1 and Col2c2 are SOX-mutant sequences from the Col2a1 enhancer that are required for chondrocyte-specific expression. In vivo, SOX proteins appear to tolerate considerable sequence variation in their binding sites. The sequences conform loosely to the HMG consensus binding site (A/T)(A/T)CAA(A/T)CAA(A/T)CAA(A/T). The residues that correspond to this consensus are shown in bold. The residues that correspond to this consensus are shown in bold. The residues that correspond to this consensus are shown in bold.

**Electrophoretic Mobility Shift Assays**—Oligonucleotide probes were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer. The sequences of the upper strands are given below. S9WT sequence is GGTTAAGCGGAAATACTCTAATCTGGTAGTA. The high affinity SOX binding site is shown in bold. The high affinity SOX binding site is shown in bold. The four flanking residues (underlined) are those that are least preferred for binding of SOX9 in vitro (12). Col2c1 sequence is GGTCGAGGAAAAAGCCCATCTCCTGAGACC. Col2c2 sequence is GGTCGAAGAAAGCCCATCTCCTGAGACCC. Col2c2 sequence is GGTCGAGGAAAAAGCCCATCTCCTGAGACCC. Col2c1 and Col2c2 are SOX-mutant sequences from the Col2a1 enhancer that are required for chondrocyte-specific expression. In vivo, SOX proteins appear to tolerate considerable sequence variation in their binding sites. The sequences conform loosely to the HMG consensus binding site (A/T)(A/T)CAA(A/T)CAA(A/T). The residues that correspond to this consensus are shown in bold. The residues that correspond to this consensus are shown in bold. The residues that correspond to this consensus are shown in bold.
To correct for varying transfection efficiencies, CAT enzyme-linked immunosorbent assay kit (Roche Molecular Biosciences) was determined by enzyme-linked immunosorbent assay, using a reporter plasmid, which was then treated with shrimp alkaline phosphatase and labeled with [γ-32P]ATP using T4 polynucleotide kinase. Probes (0.2–0.8 ng) were mixed with extract containing 180 ng of wild type or 600 ng of A19V or P70R mutant SOX9 HMG domain in binding buffer (13), in a total volume of 16 μl, and kept for 15 min on ice. Products were resolved by SDS-PAGE (Fig. 4). The affinities of wild type and mutant HMG domains were expressed in E. coli upon induction with IPTG. The proteins were soluble and stably expressed as judged by SDS-PAGE (Fig. 4). The affinities of wild type and mutant HMG domains for DNA probes S9WT, SOXCORE, and two sequences from the Col2a1 enhancer, Col2c1 and Col2c2, were compared by EMSA (Fig. 5). The probe, S9WT, bears the high affinity SOX9-binding site selected in vitro (AGAACATG). This sequence includes the high affinity binding site defined for other SOX proteins (A/AT/AT/CA/A/AT), shown in bold and termed SOXCON here; Ref. 20–22), flanked on either end by sequences from the 5′ and 3′ ends of SOX9 (Table I).

DNA Binding and Bending Activities of Mutant HMG Domains from Patients with CD—The wild type and mutant HMG domains were expressed in E. coli upon induction with IPTG. The proteins were soluble and stably expressed as judged by SDS-PAGE (Fig. 4). The affinities of wild type and mutant HMG domains for DNA probes S9WT, SOXCORE, and two sequences from the Col2a1 enhancer, Col2c1 and Col2c2, were compared by EMSA (Fig. 5). The probe, S9WT, bears the high affinity SOX9-binding site selected in vitro (AGAACATG). This sequence includes the high affinity binding site defined for other SOX proteins (A/AT/AT/CA/A/AT), shown in bold and termed SOXCON here; Ref. 20–22), flanked on either end by two probes preferred by SOX9 (12). SOXCORE bears the sequence GCAACAATCT, in which the four flanking residues of S9WT are mutated to those selected by SOX9 at lowest frequency in these positions (underlined). The wild type SOX9 HMG domain bound S9WT (relative binding 100%) more strongly than the other probes. Binding of SOXCORE, was about 8-fold lower. These results are consistent with our previous finding that the 5′-AG and 3′-GG in S9WT enhance binding of SOX9 (12). Interestingly, binding of the wild type HMG domain to Col2c1 and Col2c2 was about 5- and 3-fold lower than to S9WT. Note that Col2c1 has a single HMG binding site, which includes the 3′-flanking G in S9WT, whereas Col2c2 has two sites, one of which includes the 3′-flanking G and the other of which includes both 3′-flanking G nucleotides in S9WT. Presumably only one of the two sites on Col2c2 can be occupied at a time, as only a single shifted band is seen, even with high concentrations of SOX9 HMG domain.

Binding of the F12L mutant to any of the four probes was not detectable (relative binding <0.1%), suggesting that Phe12 is essential for DNA binding. The H65Y mutant showed barely detectable binding to S9WT, Col2c1, and Col2c2 (relative binding of 0.07%, 0.01%, and 0.01%, respectively), and undetectable binding to SOXCORE. Thus the H65Y mutation has a drastic effect on sequence-specific DNA binding. Binding of the A19V mutant to each of the four probes was only 3-5-fold lower than wild type, suggesting that Ala19 is not essential for DNA binding and that the A19V mutation does not drastically disrupt the structure of the HMG domain.

Interestingly, the P70R mutant showed altered DNA binding specificity compared with the wild type HMG domain. As stated above, binding of the wild type HMG domain to S9WT was about 8-fold higher than to SOXCORE. In comparison, whereas binding of the P70R mutant HMG domain to S9WT was only 7-fold lower than the wild type HMG domain, its
binding to SOXCORE was undetectable (<0.01% relative binding). Thus, the four residues in S9WT that flank the core SOX consensus site appear to be essential for binding of the P70R mutant to DNA, whereas they enhance binding of the wild type SOX9 HMG domain only moderately. This suggests that the P70R mutant is missing some of the key contacts that contribute to binding to the core SOX consensus site. Finally, we found binding of the P70R mutant to the Col2c1 and Col2c2 probes to be barely detectable. The presence of at least one of the flanking residues of S9WT in the binding sites on these probes is presumably responsible for the small amount of binding observed.

Some point mutations in the HMG domain of SRY in patients with XY gonadal dysgenesis alter the DNA bending properties of the protein (23). Therefore, we determined the bend angles induced upon binding of the wild type and mutant HMG domains to S9WT and SOXCON, using a circular permutation assay. The bend angle induced upon binding of the wild type HMG domain to S9WT was $\theta = 71^\circ \pm 0.4^\circ$. The A19V and P70R mutants bent this probe similarly (Fig. 6A). The bend angle induced upon binding of the wild type HMG domain to SOXCON was $\theta = 78^\circ \pm 0.6^\circ$ (Fig. 6B). This is similar to the angle induced upon binding of the SRY HMG domain to SOXCON (results not shown). The A19V mutant bent the SOXCON probe similarly (Fig. 6B). Thus, the A19V and P70R mutations do not appear to alter the DNA bending properties of SOX9.

Molecular Modeling—To investigate further the function of specific amino acid residues in the SOX9 HMG domain, we built a model of the structure of the HMG domain of SOX9, based on the known solution structure of the HMG domain of SRY in complex with DNA. The SRY and SOX9 HMG domains differ at 39 of the 77 amino acids in the SRY structure. The homology model of the SOX9 HMG domain fits closely to the structure of the SRY HMG domain; 72 of the 77 C-$\alpha$ carbons have been aligned (root mean square deviation = 0.72 Å; Fig. 7A).

The DNA binding surfaces of the SRY and SOX9 HMG domains are depicted in Fig. 8 (A and B). Of the four mutated
residues of SOX9 studied here, Phe\textsubscript{12} (cyan), His\textsubscript{65} (magenta),
and Pro\textsubscript{70} (yellow) are located on or near the DNA binding
surface in similar positions to the homologous residues in SRY.
Ala\textsubscript{19} of SOX9 is not part of the DNA binding surface; it faces
away from the DNA, into the solvent. In the SOX9 model, as in
the SRY structure, the side chain of Phe\textsubscript{12} interacts with the
base of T12. Pro\textsubscript{70} lies at the end of helix 3 of both SRY and
SOX9 HMG domains and is likely to be important in determin-
ing the orientation of the C-terminal tail that includes residues
Lys\textsubscript{73} (blue) and Tyr\textsubscript{74} (green). These residues are thought to be
instrumental in DNA binding and bending by SRY, and the
present model of the SOX9 HMG domain suggests that their
positions on the DNA binding surface are conserved in the
SOX9 HMG domain.

Inspection of the model of the SOX9 HMG domain allows us
to speculate on how the SOX9 mutations studied here affect
DNA binding. The F12L mutation affects a key aromatic con-
tact between the HMG domain and the DNA; in the F12L
mutant, the Leu\textsubscript{12} side chain is unlikely to interact with the
bases in the same way that Phe\textsubscript{12} does in the wild type HMG
domain. In the A19V mutant, the larger hydrophobic side chain
of Val\textsubscript{19} is likely to be stabilized by interaction with Phe\textsubscript{12},
Tyr\textsubscript{15}, and Tyr\textsubscript{43}. These interactions could alter the interaction
of Phe\textsubscript{12} with the DNA.

His\textsubscript{65} of SOX9 lies in a hollow on the DNA binding surface,
with most of its side chain accessible to solvent, away from the
DNA binding surface. Replacement of this residue with tyro-
sine would replace a positively charged or hydrophilic neutral
residue with a larger and more hydrophobic side chain, which
would prefer to be buried, with a consequent change of confor-
mation. In this changed conformation, Tyr\textsuperscript{65} might protrude

---

**Fig. 7.** MOLSCRIPT C-α trace overlay of HMG domains of wild
type SOX9 model (cyan) and SRY (1hry, magenta) (A) and HMG
domain models of the P70R mutant of SOX9 (green) and wild
type SOX9 (cyan) (B). Selected residues implicated in DNA binding
are shown in ball-and-stick representation.

**Fig. 8.** Molecular surfaces from models of HMG domains in
complex with a DNA octamer d(GCACAAAC), showing the loca-
tions of specific residues on the DNA binding surface. A, HMG
domain of SRY; B, HMG domain of wild type SOX9; C, HMG domain
of P70R mutant SOX9. The contributions of residues pertinent to the
molecular surface have been color-coded; Phe\textsubscript{12} (cyan), His\textsubscript{65} (magenta), Pro\textsubscript{70} (yellow), Lys\textsubscript{73} (blue), Tyr\textsubscript{74} (green), and Arg\textsubscript{77} (red). The SRY structure
does not have coordinates for residue 77. The backbone of the upper
strand of the DNA is shown in red and that of the lower strand in purple.
The bases of the DNA are color-coded (G, green; C, yellow; A, red; T, purple), and the deoxyribose units are shown as white pentagons.
To investigate the effect of the P70R mutation on the structure of the SOX9 HMG domain, we built a model of the P70R mutant HMG domain, in which 71 of the 77 Cα carbons have been aligned with those of the wild type SOX9 HMG domain (root mean square deviation = 0.69 Å; Fig. 7B). The DNA binding surface of the P70R mutant HMG domain is depicted in Fig. 8C. Comparison of the models of wild type and P70R mutant SOX9 HMG domains allows us to speculate on how the P70R mutation affects the interaction of the HMG domain with DNA. The major difference between the model of the P70R mutant and that of the wild type HMG domain is that the C-terminal tail is oriented differently in the P70R mutant (Fig. 7B). This tail includes residues Lys73 and Tyr74, which appear to be important DNA-contact residues. In the NMR structure of SRY, Tyr74 interacts hydrophobically with base A3 (closest distance = 3.2 Å). In the model of the SOX9 HMG domain, Tyr74 is not close enough to this base to make hydrophobic contact (closest distance = 6.9 Å), but its side chain is able to flip over so that the phenolic hydroxyl group can interact with base C16 by hydrogen bonding. Two contacts with C16 are possible: one with N1 of C16 (closest distance = 3.3 Å) and the other with the sugar oxygen of C16 (closest distance = 2.7 Å). In the SOX9 model, Lys73 is able to interact with the phosphate group of C16 (closest distance = 2.7 Å). Thus, the present model of the wild type SOX9 HMG domain suggests that the side chains of Lys73 and Tyr74 are likely to help to stabilize the interaction of the SOX9 HMG domain with one of the residues flanking the 5′-end of the core SOX binding site in S9WT. This might explain the preference of SOX9 for the 5′-flanking G in S9WT. According to the models, a consequence of the altered orientation of the C-terminal tail of the P70R mutant is that Lys73 is no longer oriented toward the DNA binding surface, and Tyr74, while still on the DNA binding surface, is also unable to make contact with the DNA (closest distance = 5.5 Å). Thus, some of the key contacts that the wild type SOX9 HMG domain makes with the DNA appear to be lost in the P70R mutant. However, a compensating interaction is possible; in the model of P70R mutant, the altered orientation of the C-terminal tail may allow Arg57 (Fig. 7C; red) to interact with one of the residues flanking the core SOX binding site in S9WT. In conclusion, the models are consistent with the hypothesis that the reduced DNA binding affinity and altered DNA binding specificity of the P70R mutant are, at least in part, due to a different orientation of the C-terminal tail of the HMG domain from that of the wild type protein.

The PQS-rich and PQA Domains of SOX9 Both Contribute to Transactivation—Many CD mutations result in truncation of the C terminus of SOX9. We constructed a series of mutants with successively larger deletions of their C termini. These mutants mimic mutant SOX9 proteins seen in CD, and by assaying their transactivation activity we have defined the transactivation domain of SOX9.

Transfection of COS-7 cells with full-length SOX9 gave a 31-fold induction of CAT transcription compared with transfection of cells with the vector, pcDNA3. SOX9 mutants with successively larger deletions of their C termini gave successively lower levels of CAT activation (Fig. 9A). Transactivation by the deletion mutant SOX9 1–454 (which lacks the C-terminal 55 amino acids) was 4.4-fold lower than wild-type SOX9 (Sceffe, p < 0.05). Transactivation by mutant SOX9 1–248 was reduced another 9.5-fold (Sceffe, p < 0.05), to levels that are near background. A mutant lacking only the PQA domain gave 31-fold induction of CAT activation compared with transfection of cells with the vector, pcDNA3. SOX9 mutants with successively larger deletions of their C termini gave successively lower levels of CAT activation (Fig. 9A). Transactivation by the deletion mutant SOX9 1–454 (which lacks the C-terminal 55 amino acids) was 4.4-fold lower than wild-type SOX9 (Sceffe, p < 0.05). Transactivation by mutant SOX9 1–248 was reduced another 9.5-fold (Sceffe, p < 0.05), to levels that are near background. A mutant lacking only the PQA domain gave 31-fold induction of CAT activation compared with transfection of cells with the vector, pcDNA3.
DISCUSSION

In this study, we have investigated the structure and function of wild type SOX9 and two classes of mutation that occur in patients with campomelic dysplasia. We report the identification of a novel amino acid substitution mutation (H65Y) in the SOX9 HMG domain of a CD patient. The mutation appears to have been inherited from the unaffected father. It is unlikely that the mutation is a rare polymorphism that is unrelated to the CD phenotype, as a screen of the DNA from 62 Pakistanis failed to detect another instance of the mutation. Furthermore, the DNA binding activity of the H65Y mutant protein was barely detectable, suggesting that the patient's CD phenotype results from failure of the mutant SOX9 protein to bind DNA and its consequent failure to activate expression of target genes. The father’s normal phenotype might be explained by mosaicism, whereby a certain proportion of his cells (including germ cells) are heterozygous for the mutant SOX9 allele and the remainder have two copies of the wild type allele.

A comparison of the amino acid sequences of the HMG domains that bind DNA with sequence specificity with those that bind without sequence specificity (such as HMG-1 and HMG-D) gives some clues as to which residues are important in DNA binding and sequence specificity (24). His65 of SOX9 is conserved in sequence-specific HMG domain proteins, but is replaced by a Tyr residue in most of the nonspecific HMG domains, suggesting that a His residue at this position is required for sequence-specific DNA binding. The greatly reduced DNA binding activity of the H65Y mutant compared with the wild type HMG domain supports this interpretation. In the model of the wild type HMG domain of SOX9 presented here, His65 lies in a hollow on the DNA binding surface. Replacement of this residue with a larger and more hydrophobic tyrosine residue might interfere with interaction with the DNA by protruding into the DNA binding pocket. An alternative explanation for the lack of DNA binding activity observed in the H65Y mutant is suggested by inspection of the solution structure of SRY. Packing of His65 (homologous to His65 of SOX9) with Pro9, Leu61, and Gln62 helps stabilize the middle of the long arm of the “L” (7). The model of the SOX9 HMG domain presented here suggests that the larger side chain of Tyr65 in the H65Y mutant would cause Pro9 to relocate and modulate an interaction between Arg2 and the DNA.

The present study also investigated the effect of three other point mutations within the HMG domain of SOX9: F12L, A19V, and P70R. Phe12 is conserved in the sequence-specific HMG domains and is conserved or replaced with another aromatic residue (Tyr) in the nonspecific HMG domains and thus is expected to be important for sequence-specific DNA binding. Phe12 of SRY appears to interact with an AT base pair in the minor groove of the DNA and aid the deformation of the DNA that is brought about by intercalation of the adjacent residue, Ile18, between two AT base pairs (25). In our model of the SOX9 HMG domain, Phe12 lies in a similar position on the DNA binding surface and can therefore be expected to function similarly. Thus, the F12L mutation is likely to interfere with a key contact between the SOX9 HMG domain and the DNA. The present study supports this prediction, as the F12L mutant protein shows undetectable binding to the DNA sequences tested. It is also possible that the mutation disrupts the structure of the SOX9 HMG domain, as Phe12 of SRY is one of 12 residues that lie at the junction of the three helices and form a large hydrophobic core, which stabilizes the “L” shape and has an exposed surface that contacts the DNA (7).

The A19V mutant HMG domain bound the DNA probes 3–5-fold less strongly than the wild type HMG domain, but both mutant and wild type HMG domains bent DNA probes bearing S9WT or SOXCON to the same extent. The lack of conservation of Ala19 of SOX9, among other SOX proteins, and the substantial DNA binding activity observed in the A19V mutant, suggest that Ala19 is not critical for sequence-specific DNA binding and that the A19V mutation does not drastically disrupt the structure of the HMG domain. Ala19 in SOX9 corresponds to Gln19 in SRY, which does not contact the DNA directly; instead, it helps to maintain the orientation of the long and short arms of the L-shape of the molecule (7). Thus, the reduced DNA binding activity observed in the A19V mutant HMG domain might be explained if this mutation modulates some of the interactions responsible for maintaining the orientation of the long and short arms of the wild type HMG domain. Inspection of the model of the SOX9 HMG domain suggests that interaction between Val19 in the A19V mutant with Phe12, Tyr15, and Tyr43 is likely to modulate the interaction of Phe12 with the DNA. The CD phenotype that results from the A19V mutation indicates that the residual DNA binding activity observed in this mutant protein is insufficient to allow the protein to bind and activate target genes to levels required for normal development. Alternatively, the A19V mutant may be deficient in some activity not measured here.

Pro70 of SOX9 is conserved among sequence-specific HMG domain proteins but not among the nonspecific HMG domains, suggesting that it is important for sequence-specific DNA binding. Meyer et al. (26) found that the P70R mutant HMG domain showed reduced binding to a DNA probe that includes S9WT. We extend these results to show that the P70R mutant has altered DNA binding specificity. Although the wild type HMG domain binds S9WT with only 8-fold higher affinity than it binds SOXCORE, the P70R mutant showed moderately strong binding to S9WT, but no detectable binding to SOXCORE. These results suggest that the mutant protein lacks some of the contacts that the wild type protein makes with the core SOX consensus site, but retains contacts with the flanking residues in S9WT.

An inspection of the structures of SRY and LEF1 suggests that Pro70 in SOX9 is likely to be important for determining the orientation of the C-terminal tail of the HMG domain (residues 71–84). In these structures, the corresponding proline residues break helix 3 and produce a kink, such that helix 3 is shorter than that found in the HMG-1 and HMG-D domains, which lack DNA sequence specificity. The C-terminal strand then bends back toward the N terminus, forming a small hydrophobic cluster, that brings the N- and C-terminal strands into proximity and is thought to be important for sequence-specific DNA binding (8). Tyr74 of SRY forms part of the hydrophobic core, and its aromatic ring is packed against the bases of A3 and T14 and appears to push the A3 base toward the major groove, thereby disrupting base stacking and base pairing. Thus, Tyr74 appears to be important for sequence-specific DNA binding and bending. Lys72 of SRY appears to form a salt bridge with the phosphate group of C16 (7). Lys72 and Tyr74 are highly conserved in SOX proteins. In our model of the wild type SOX9 HMG domain, Lys73 and Tyr74 also appear to make important contacts with the DNA. We predicted that replacement of Pro70 in the SOX9 HMG domain with Arg would alter the orientations of these residues. Accordingly, the molecular models of the respective HMG domains, presented here, show Lys72 to be on the DNA binding surface of the wild type HMG domain, but not on that of the P70R mutant. Although Tyr74 is located on the DNA binding surface in the P70R mutant, it is not close enough to make contact with the DNA. Thus, we propose that the P70R mutation interferes with the ability of the protein to bind and activate target genes by modulating the interaction of Lys73 and Tyr74 with the DNA.
The NMR structure of SRY suggests that Lys\textsuperscript{73} and Tyr\textsuperscript{74} of SRY play an important role in DNA bending (7). As the homologous residues have altered orientations with respect to the DNA binding surface in our model of the P70R mutant of SOX9, we predicted that this mutant would have altered DNA bending properties. Using a circular permutation assay, we estimated the bend angle induced upon binding of the P70R mutant HMG domain to probes bearing S9W7 to be close to the angle induced upon binding of the wild type HMG domain. Thus, the P70R mutation does not appear to affect the DNA bending properties of the SOX9 HMG domain. The model of the P70R mutant suggests that a compensating interaction between Arg\textsuperscript{77} and the extended DNA sequence of S9W7 might stabilize DNA binding and bending by the P70R mutant.

Many of the mutations that give rise to CD truncate the C terminus of the protein. We constructed a number of SOX9 deletion mutants, which mimic the truncation and frameshift mutations seen in CD patients, and used these mutants to define the transactivation domain of SOX9. Analysis of the transactivation activity of the deletion mutants shows that progressive truncation of the C terminus of SOX9 results in progressive loss of transactivation activity and therefore demonstrates that most of the transactivation activity is conferred by the PQS-rich domain. However, the PQA domain is also required for maximal transactivation. The present results contrast with those of Sudbeck et al. (9), who also found that the transactivation activity of SOX9 was conferred by the PQS-rich domain, but the PQA domain was not required for maximal transactivation. The apparent discrepancy between these results and the present results might be explained by the fact that Sudbeck et al. studied the effect of removal of the PQA domain in the context of a fusion protein in which SOX9 (with or without the DNA binding domain) was fused to the Gal4 DNA binding domain. In contrast, the present study examined the effect of deletions, on the transactivation activity of SOX9, in the context of the native protein. The results of the present study are therefore more likely to reflect the activity of the native protein and mutant SOX9 proteins found in patients with CD.

Although the PQA domain of SOX9 varies greatly in length among species, the PQS-rich domain is highly conserved. Analysis of the mutations present in CD patients suggests that CD often arises from truncation of the C terminus of SOX9, and it is likely that in these cases the CD phenotype results from failure of the mutant SOX9 protein to activate target genes to levels sufficient for normal development.

In conclusion, the question of how the PQS and PQA domains mediate transcriptional activation remains to be answered. It is likely that they do so via interactions with other transcriptional activators or components of the basal transcription apparatus.

Acknowledgments—This work was initiated in the Laboratory of Human Molecular Genetics in the Department of Genetics at Cambridge University under the supervision of Professor Peter Goodfellow, to whom we are grateful for guidance and interest in this work. We also thank Alan Schafer for helpful discussion, Steven Matthias for pBEND2SOXCON, Harry Ostrer for the plasmid p85EIbCAT, and John Todt for microsatellite markers used in parental haplotyping.

REFERENCES

1. Foster, J. W., Dominguez Steglich, M. A., Guioli, S., Kwock, G., Weller, P. A., Stevanovic, M., Weissenbach, J., Mansour, S., Young, I. D., Goodfellow, P. N., Broek, J. D., and Schafer, A. J. (1984) Nature 312, 525–530
2. Ng, L. J., Wheatley, S., Macsotai, G. E., Conway Campbell, J., Bowles, J., Wright, E., Bell, D. M., Tam, P. P., Cheah, K. S., and Koopman, P. (1997) Biol. 183, 108–127
3. Zhao, Q., Eberspaecher, H., Lefebvre, V., and de Crombrugghe, B. (1997) Dev. Dyn. 209, 377–386
4. Bell, D. M., Leung, K. K., Wheatley, S. C., Ng, L. J., Zhou, S., Ding, K. W., Shen, M. H., Koopman, P., Tam, P. P., and Cheah, K. S. (1997) Nat. Genet. 16, 174–178
5. Zhou, G., Lefebvre, V., Zhang, Z. F., Eberspaecher, H., and de Crombrugghe, B. (1998) J. Biol. Chem. 273, 14989–14997
6. Bridgewater, L. C., Lefebvre, V., and de Crombrugghe, B. (1998) J. Biol. Chem. 223, 14998–15006
7. Werner, M. H., Huth, J. R., Gronenborn, A. M., and Clore, G. M. (1995) Cell 81, 705–714
8. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) Nature 379, 791–795
9. Sudbeck, P., Schulz, M. L., Baeuerle, P. A., and Scherer, G. (1996) Nat. Genet. 13, 230–232
10. Kwock, C., Weller, P. A., Guioli, S., Foster, J. W., Mansour, S., Zuffardi, O., Punnett, H. H., Dominguez Steglich, M. A., Broek, J. D., and Young, I. D. (1995) Am. J. Hum. Genet. 57, 1028–1036
11. Smith, D. B. (1993) Methods Mol. Cell. Biol. 4, 220–229
12. Martin, S., McDowell, S. G., and Harley, V. R. (1999) Nucleic Acids Res. 27, 1359–1364
13. van de Wetering, M., Oosterwijk, M., Donijies, D., and Clevers, H. (1991) EMBO J. 10, 123–132
14. Kwock, C., Zwieb, C., Wu, C., and Adhya, S. (1989) Gene (Amst.) 85, 15–23
15. Thompson, J. P., and Landy, A. (1988) Nucleic Acids Res. 16, 9687–9705
16. Salis, A., and Blundell, T. L. (1993) J. Mol. Biol. 234, 779–815
17. Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
18. Nicholls, A., and Honig, B. (1991) J. Comput. Chem. 12, 434–445
19. Gauss, G. H., and Lieber, M. R. (1992) Nucleic Acids Res. 20, 6739–6740
20. Harvey, R., Lovell Badge, R., and Goodfellow, P. N. (1994) Nucleic Acids Res. 22, 1500–1501
21. Kanai, Y., Kanai Azuma, M., Noe, T., Saide, T. C., Shiroishi, T., Hayashi, Y., and Yazaki, K. (1996) J. Cell Biol. 133, 667–681
22. Denny, P., Swift, S., Connor, F., and Ashworth, A. (1992) EMBO J. 11, 3705–3712
23. Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell Badge, R., and Bianchi, M. E. (1994) EMBO J. 13, 6115–6124
24. Grosschedl, R., Giese, K., and Pagel, J. (1994) Trends Genet. 10, 94–100
25. Tang, Y., and Nilsson, L. (1998) Trends Genet. 14, 611–618
26. Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991) Genes Dev. 5, 880–894