

Functional Analyses of Natural Variation in Sp1 Binding Sites of a TATA-Less Promoter

Jeff A. Segal,^{1,*} J. Lynn Barnett,² Douglas L. Crawford²

¹ Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA

² Division of Molecular Biology and Biochemistry, 5007 Rockhill Rd., University of Missouri, Kansas City, MO 64110, USA

Received: 12 February 1999 / Accepted: 2 June 1999

Abstract. Within the lactate dehydrogenase-B (*LdhB*) proximal promoter is a region with multiple in vivo footprinted sites that resembles the binding site for the transcription factor SP1. Like many sequences that regulate transcription rate, these Sp1 binding sites are well conserved among species of the teleost fish *Fundulus*. The only exception is in the northern population of *F. heteroclitus*, where there are many changes in the Sp1 binding sites. These changes affect footprinting patterns, measures of promoter strength, and are associated with the adaptive increase in *Ldh-B* transcription rates. Reported here is data that demonstrates that *Fundulus* hepatocytes have an SP1-like protein; in comparison to human SP1 protein, it has similar specificity and size and a greater affinity for the consensus Sp1 site. This *Fundulus* hepatocyte SP1-like protein as well as the human SP1 protein binds the *Ldh-B* Sp1 sites. Sequence variation in the northern Sp1 region eliminates the “preferred” Sp1 binding site, yet these northern Sp1 sites have significantly greater affinity for the SP1 protein than either the Sp1 sites from southern *F. heteroclitus* (~ 1.6-fold) or the consensus Sp1 site (GGGCGG; ~ 1.8-fold). Furthermore, the *Ldh-B* Sp1 sites also bind non-SP1 proteins, and the extent of binding is affected by the sequence variation in the proximal promoter. These data suggest that natural variation in Sp1 sites affect binding of tran-

scription factors and may effect a modest change in transcription rates.

Key Words: Binding affinities — phylogeny — evolutionary adaptation — transcription

Introduction

Sequence variation is a feature of almost all species (Li 1997) and knowledge of how this variation affects an organism’s biochemical, physiological, and cellular function will be necessary to interpret the significance of the increasing number of genomic sequences being described. For example, natural sequence variation in promoter regions could produce important phenotypic differences. The unexpected diversity in proximal promoter sequences has led to the suggestion that these proximal regions contribute to differential regulation of transcription or subtle changes in transcription rate (O’Shea-Greenfield and Smale 1992; Zenzie-Gregory et al. 1993a, 1993b; Ernst and Smale 1995; Verrijzer et al. 1995; Goodrich et al. 1996). Typically, proximal promoters have a TATA sequence that binds TFIID, an essential factor for all mRNA transcription (Roeder 1991; Smale 1994). Nevertheless, many promoters lack a TATA sequence (e.g., approximately 50% of transcribed genes in *Drosophila* [Arkhipova, 1995]). These TATA-less promoters require Sp1 binding sites (consensus sequence: GGGCGG; Dynan and Tjian 1983) for significant activity (Azizkhan et al. 1993), and the degree of

*Present address: Lilly Research Laboratories, CNS Discover, Bldg. 48, Drop Code 0510, Indianapolis, IN 46285-0510, USA

Correspondence to: D.L. Crawford; e-mail: crawfordDo@umkc.edu

activation from Sp1 tends to be stronger in the context of TATA-less promoters than TATA-containing promoters (Colgan and Manley 1995). Activation of TATA-less promoters by SP1 may involve SP1 recruitment of TFIID to TATA-less promoters (Pugh and Tjian 1991; Wiley et al. 1992). Furthermore, in many of these promoters, SP1 binding is intimately involved in the determination of the transcription start site or sites (Jolliff et al. 1991; Kollmar et al. 1994; Lu et al. 1994; Boam et al. 1995). These facts, together with the observation that a disproportionate number of the TATA-less promoters described are in constitutively regulated genes (Azizkhan et al. 1993), suggest that variation in SP1 binding is one means by which constitutive transcription is modified.

For lactate dehydrogenase-B (*Ldh-B*), the constitutive transcription rate is twofold greater in hepatocytes from northern versus southern populations of the teleost fish *Fundulus heteroclitus* (Crawford and Powers 1992). This difference in transcription rate results in a greater LDH-B₄ enzyme concentration that compensates for the colder temperatures experienced by the northern population (Crawford and Powers 1989, 1992; Segal and Crawford 1994). The variation in *Ldh-B* expression is adaptive (i.e., evolved by natural selection; Pierce and Crawford 1997a) and thus is biologically important because natural selection acts only on traits that affect longevity, reproductive fitness, or survival (Endler 1986).

Much of the adaptive variation in *Ldh-B* expression is associated with sequence variation in the proximal promoter. This sequence variation affects *in vivo* and *in vitro* footprinting patterns and levels of gene expression in transient cell transfection (Segal et al. 1996; Crawford et al. 1999b). Specifically, proximal promoters isolated from northern individuals consistently have greater promoter strength than promoters isolated from southern individuals. The *Ldh-B* gene proximal promoter (Fig. 1) is TATA-less with multiple transcriptional start sites, Inr (initiation of transcription) sequences associated with the start sites and a cluster of binding sites similar to those recognized by the ubiquitous SP1 protein (i.e., SP1 sites that match at least five of the six nucleotides of the consensus SP1 site, Segal et al. 1996). This SP1 region consists of multiple overlapping perfect and imperfect SP1 sites that promote transcription in transient transfection assays (Segal et al. 1996). Similar to many transcriptional regulatory elements (e.g., Gumucio et al. 1994, 1996; Hu et al. 1995) the SP1 binding sites are highly conserved among species of *Fundulus* (Fig. 1). However, a comparison between populations of *F. heteroclitus* reveals significant sequence variation: the SP1 sites from northern populations (northern SP1 sites) have many unique derived nucleotides that are not found in the SP1 sites from southern populations (southern SP1 sites) or other species (Fig. 1; Crawford et al. 1999b). Established phylogenies of *Fundulus* that have been generated from morphological traits, protein isoforms, and DNA se-

quences have all grouped *F. heteroclitus* populations into one species and divided the other taxa into much older species that diverged several million years ago (Wiley 1986; Cashner et al. 1992; Bernardi and Powers 1995; Crawford et al. 1999b). However, the phylogenetic tree established using the sequences from SP1 regions is strikingly different: only the northern *F. heteroclitus* population is a distinct group, and the southern population and the other species all group together. This disparity indicates that the SP1 sites within *F. heteroclitus* have recently accumulated many more substitutions in comparison to homologous SP1 sites among much older species. Importantly, the pattern of nucleotide substitutions in the SP1 sites (and other protein:DNA binding regions in the proximal promoter) is nonrandom and indicates that the *Ldh-B* proximal promoter has evolved by natural selection, specifically directional selection (Crawford et al. 1999b). Thus, the natural sequence variation in the *Ldh-B* proximal promoter appears to be biologically important.

To better understand how the natural sequence variation of the SP1 region affects transcription, we examined the nature of the transcription factors that bind to this region of the *Ldh-B* proximal promoter and analyzed the functional consequences of this variation. The data demonstrate that (1) expression of an SP1 protein is necessary for transcription from the *Ldh-B* proximal promoter, (2) purified human SP1 protein binds to the *Ldh-B* SP1 sites, (3) there is a *Fundulus* SP1-like protein in hepatocyte nuclear extracts from both populations that binds to the SP1 sites, (4) there are non-SP1 proteins in these nuclear extracts that bind these sequences, (5) the *Fundulus* SP1-like proteins have greater affinity for the consensus SP1 site than the human SP1, and (6) the variation in SP1 sites results in different K_d 's for the human SP1 protein and differential binding of *Fundulus* non-SP1 factors. These data demonstrate that the naturally occurring variation in SP1 sites alters the binding of transcription factors and thus may contribute to the difference in constitutive *Ldh-B* transcription rate *in vivo*.

Materials and Methods

Organisms

Fundulus were captured from wild populations (northern and southern *F. heteroclitus*: Wiscasset, ME, and Sapelo Is. Dock, GA, respectively; *F. grandis*. Panacea, FL; *F. diaphanus*, Green Banks, NJ) and acclimated. The acclimation regime was initiated by a 2-week feeding period, a 6-week pseudo-winter (6°C, 16:8 light:dark cycle), followed by minimum of 6 weeks at 20°C, 14:10 light:dark cycle. During this time all species came into reproductive condition and spawned. The reproductive tissues were in regression in all species when assayed. This acclimation regime is sufficient to remove environmentally induced physiological difference (Pierce and Crawford 1997b).

Plasmids

The *Ldh-B* proximal promoter is depicted in Fig. 1A. Two proximal promoters with Sp1 sites, TCC repeats and Inrs (STI) were designed, which represent either the typical northern or southern proximal promoter (see Crawford et al. 1999b). These proximal promoters were ligated into a luciferase reporter plasmid (pGL3, Promega). The SP1 expression vector (pPacSp1; Courey and Tjian 1988; Kadonaga et al. 1988) was provided by Dr. R. Tjian (University of California, Berkeley).

Transfections

Drosophila cell lines were used to determine the role of SP1 in *Ldh-B* proximal promoter activity (SL2 ATCC, Rockville, MD; Fig. 2). These cell lines lack SP1 proteins (Courey and Tjian 1988; Hagen et al. 1994). Transfections were performed as in Segal et al. (1996). Promoter activity is expressed as net relative light units (luciferase activity minus the negative control) normalized for transfection activity (net B-galactosidase activity, B-galactosidase activity minus the negative control). Promoter activity was further standardized by dividing by a constant: the mean activity for southern STI proximal promoter. This does not alter the statistical variance and makes it simpler to compare relative activity in different cell lines.

Oligonucleotides

The consensus Sp1 double-stranded oligonucleotide was purchased from Promega (sequence delineated in Fig. 1D) and labeled directly. All other single-stranded oligonucleotides were purchased from Operon (sequences indicated in Fig. 1D: SN, SS1, and SS2; Fig. 6: SNGa and SNGc; Results section: Sp1_N) and purified by polyacrylamide gel electrophoresis. Equimolar quantities of complementary strands were annealed by heating to 90°C for 10 min followed by slow cooling to 25°C overnight.

For mobility shift assays, all oligonucleotides were end-labeled with T4 polynucleotide kinase (PNK; New England Biolabs) in a 10 µl reaction containing 2.5 pmol of double-stranded annealed oligonucleotide, 1.0 µl of PNK 10× buffer (New England Biolabs), 1.0 µl of gamma P³² ATP (3,000 Ci/mmol; Dupont), 5.0 µl of H₂O, and 1 µl PNK. The reaction was incubated at 37°C for 20 min and stopped by the addition of 1.0 µl of 0.5 M EDTA. The volume was brought up to 100 µl with the addition of 1X TE, and the unincorporated radionucleotide was removed by applying it to a G25 column equilibrated with 1X TE.

Mobility Shift Assay

All binding reactions were performed in a final volume of 10 µl and contained 2 µl of 5X binding buffer (5XBB: 20% glycerol, 5 mM MgCl₂, 2.5 mM EDTA, 2.5 mM DTT, 250 mM NaCl, 50 mM Tris-HCl pH 7.5, and 0.25 µg/µl poly dI-dC) and 1 µl 5 µg/µl BSA. Experiments typically used either 0.5 footprinting units of purified human Sp1 protein (rhSP1, expressed in HeLa cells; Promega) or approximately 1–3 µg of northern or southern hepatocyte nuclear extract protein (described below) per reaction. The protein, binding buffer, BSA, and H₂O to 9 µl was incubated on ice for 10 min before the addition of 25 fmol of freshly labeled oligonucleotide. After the addition of labeled oligonucleotide, the reaction was incubated for 20 more min on ice, followed by the addition of 1 µl of 10X loading dye (250 mM Tris-HCl pH 7.5, 0.2% bromophenol blue, 0.2% xylene cyanol, and 40% glycerol). For experiments that included an excess unlabeled competitor oligonucleotide, the competitor was added before the addition of labeled oligo-

nucleotide. For supershift experiments, antibody was incubated with protein, binding buffer, and BSA for 30 min instead of 10 min before the addition of labeled oligonucleotide. All reactions were separated by electrophoresis at 4°C in a 4% (80:1 acrylamide to bisacrylamide) nondenaturing gel with 0.5X TBE and 0.5X TBE running buffer. The gel was subsequently dried and subjected to autoradiography overnight with an intensifying screen.

Nuclear Extracts

Hepatocyte nuclear extracts were prepared from livers of northern and southern *F. heteroclitus* according to Gorski et al. (1986) with modifications. All the procedures were carried out in a cold room; all buffers, solutions, and rotors were cooled to 4°C before use. Livers from three to five northern or two to four southern *F. heteroclitus* individuals (typically 500 mg of tissue) were excised and homogenized with four strokes of a motor-driven Teflon-glass homogenizer in 5.5 ml of homogenization buffer (2.0 M sucrose, 10 mM Hepes, pH 7.6, 0.5 mM spermidine, 0.15 mM spermine, 1.0 mM EDTA, and 15 mM KCl) supplemented just before use with 1% nonfat milk, 1 mM DTT, and 1X protease inhibitor mix (0.5 mM PMSF, 2 µg/ml aprotinin, 2 µg/ml leupeptin, and 2 µg/ml pepstatin). This homogenate was filtered through glass wool to remove solid material and mixed with an equal volume of homogenization buffer. The sample was then layered over 4 ml of sucrose cushion (same as homogenization buffer, but with 2.2 M sucrose and without the lowfat milk) and centrifuged for 45 min at 24,600 rpm in a SW41 rotor (75,000 g; Beckman). Each nuclear pellet was resuspended by gentle pipetting in 2.5 ml of nuclear lysis buffer (10 mM Hepes, pH 7.6, 100 mM KCl, 0.1 mM EDTA, 10% glycerol, and 3 mM MgCl₂), supplemented just prior to use with 1 mM DTT and 1X protease inhibitor mix. 375 µl of 3 M (NH₄)₂SO₄ was added to each sample, followed by gentle rocking for 30 min at 4°C. The samples were then centrifuged for 60 min at 35,000 rpm in a 70.1 Ti rotor (90,000 g; Beckman) to pellet the chromatin. Two ml of the supernatant were transferred to a new tube, and 0.66 g of solid (NH₄)₂SO₄ (0.33 g/ml) was slowly dissolved into the sample. After 45 min at 4°C, the sample was centrifuged for 60 min at 24,600 rpm in the SW41 rotor (75,000 g) to pellet the protein precipitate. Each protein pellet was resuspended at 4°C in 60 µl of nuclear dialysis buffer (25 mM Hepes, pH 7.6, 0.1 mM EDTA, 40 mM KCl, and 10% glycerol), supplemented just prior to use with 1 mM DTT and 1X protease inhibitor mix. The samples were then dialyzed twice for 90 min at 4°C against 150 ml of nuclear dialysis buffer supplemented with 1 mM DTT. The extracts were then centrifuged in a microcentrifuge (16,000 g) at 4°C for 5 min to clarify. Ten microliters of the supernatant were removed for protein assay determination, and then an equal volume of 5 mg/ml BSA (in nuclear dialysis buffer supplemented with 1 mM DTT and 1X protease inhibitor mix) was added to the remaining 50 µl. Aliquots were quick-frozen in liquid nitrogen.

Binding Affinities

Mobility shift assays were used to separate bound from free oligo. Mobility shifts, used for the nonlinear analyses of quantitative binding rates, were done as described above. For each experiment, nine different amounts of oligonucleotides were used: 500 fm, 400 fm, 335 fm, 270 fm, 205 fm, 140 fm, 75 fm, 50 fm, and 30 fm; in a 10-µl reaction, yielding 50 to 3 nM of oligonucleotides. Binding reactions contained either 0.013 footprinting units of rhSP1 or 0.5 µg of northern or southern nuclear extract. In all cases, these conditions gave a range of protein to oligonucleotide ratio that approached protein binding saturation at the highest oligonucleotide concentration. After electrophoresis as described above, the gel was dried and the radioactivity corresponding to bound and free oligonucleotide was quantified with a phosphorimager (Molecular Dynamics). Nonlinear regression analyses of binding af-

finities were generated from the data for each experimental gel and the K_d determined with the use of GraphPad Prism 2.0 software (GraphPad). Each oligonucleotide:protein interaction was tested in at least three separate experiments, and the mean K_d was calculated from these separate measures. Statistical significance of differences between different oligonucleotide K_d s was determined by analysis of variance (ANOVA) using Minitab.

Results

Sequence variation in Sp1 region

The structure of the *F. heteroclitus* *Ldh-B* proximal promoter and the variation in Sp1 sites are shown in Fig. 1. Previous work has demonstrated that the *Ldh-B* Sp1 sites contribute significantly to promoter activity (Segal et al. 1996). These Sp1 sites consist of a number of heterogeneous sequences that match at least five of the six nucleotides for the canonical Sp1 binding site (GGGCGG; Dynan and Tjian 1983). All of these sequences are the reverse complement relative to the coding strand. However, in either orientation Sp1 sites will bind SP1 protein and activate transcription (Kadonaga et al. 1987). The Sp1 sequence region can be subdivided into two general subregions: the 5' GA region (GGGAGG/CCTCCC) and the 3' GC region. In the GC subregion from the southern population of *F. heteroclitus* and in other *Fundulus* species there are perfect (six of six nucleotide match) Sp1 sites. In contrast, from northern populations of *F. heteroclitus* there are no "perfect" Sp1 binding sites because of a substitution of a G for a C (#5, Fig. 1B) that changes GGGCGG to GGCCGG. Within the GA subregion, all northern sequences have four potential Sp1 binding sites (i.e., five of six match; see Fig. 1D) while southern *F. heteroclitus* and other *Fundulus* species have only two or three potential Sp1 binding sites. This difference is due to a unique insertion in the northern population of an extra C (#1, Fig. 1) or deletions in this region (#4 CCT, Fig. 1). The other major difference specific to the northern population is the deletion of GGCCA (Fig. 1) that eliminates a Sp1 site (five of six match) and decreases the distance to the TCC repeat (the putative TFIID binding site; Segal et al. 1996).

SP1 Protein Activates Transcription

The *Drosophila* SL2 cell line (lacking SP1 protein) was used to determine if SP1 transactivation is required for *Ldh-B* proximal promoter function. *Ldh-B* STI proximal promoters (Sp1, TCC, and Inr, Fig. 1A) were transfected with or without a human SP1 expression vector (pPacSp1). Promoter strength of a STI proximal promoter is highly correlated with the amount of human SP1 expression vector that is cotransfected ($r^2 = 0.969$; Fig. 2A). Moreover, expression of SP1 (by cotransfection of pPacSp1) with northern or southern STI promoters sig-

nificantly increases promoter strength (Fig. 2B). For the northern promoter, this increase in expression was largely dependent on the presence of the Sp1 sites (compare STI north with TI north in the presence of SP1 protein). Surprisingly, this was not the case for the southern proximal promoter: SP1 expression increases both the southern TI and STI promoter activities equally. These results suggest that the southern Sp1 sites are not important for mediating Sp1 transactivation. However, transfection into fish cell lines has demonstrated that both northern and southern promoter activities are critically dependent on the Sp1 sites (Segal et al. 1996). The only difference between the northern and southern TI promoters is that the northern promoter has a single T to C transition among the TCC repeats (Crawford et al. 1999b). A possible explanation for these disparate results is that significant overexpression of exogenous SP1 protein in the SL2 cell lines allows for SP1 binding and transactivation from southern TI promoters, whereas the T to C transition in the northern promoters precludes this.

Northern and Southern Sp1 Regions Bind a Human Recombinant SP1 Protein

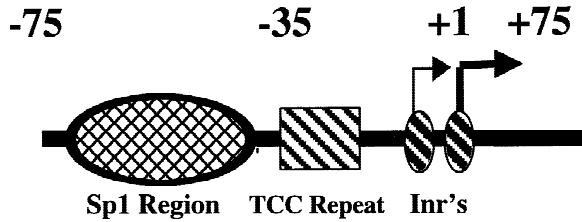
To investigate whether SP1 proteins interacts with the *Ldh-B* Sp1 sites and whether sequence variation between northern and southern promoter types influences such interactions, oligonucleotides were designed for use in mobility shift assays (Fig. 1D). Three oligonucleotides were designed based on the *Ldh-B* Sp1 sites found in northern or southern individuals: a northern sequence (SN) and two southern sequences (SS1 and SS2; Fig. 1D). These three *Ldh-B* oligonucleotides reflect the type and extent of variability found in natural populations (Fig. 1B; Crawford et al. 1999b). Additionally, an oligonucleotide that contains a single, perfect Sp1 site was utilized (Sp1; Promega).

As a first approach to determine whether SP1 binds to the *Ldh-B* Sp1 sites, mobility shift assays were done with a recombinant human SP1 protein (rhSP1 expressed in HeLa cells; Promega) in the presence of 0.05 mg/ml poly dI-dC. The results from this experiment demonstrate that rhSP1 binds specifically to the Sp1, SN, SS1, and SS2 oligonucleotides (Fig. 3). Specificity of binding is demonstrated by the reduction of mobility shift with 30- to 40-fold excess of Sp1 consensus oligo. The Sp1 oligo binds a single rhSP1 whereas the three *Ldh-B* promoter regions appear to be capable of binding more than one rhSP1 protein (compare lane 2 to lanes 8, 14, and 20, Fig. 3). This pattern reflects the greater number of Sp1 binding sites that occur in these oligonucleotides.

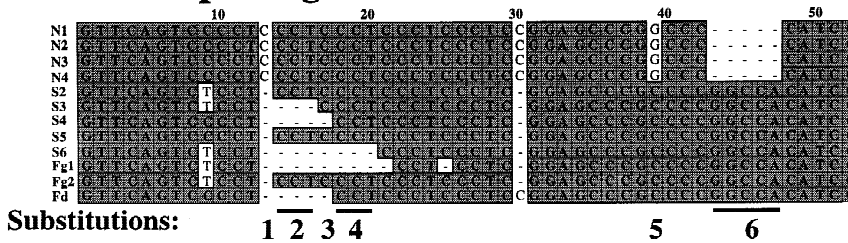
Northern and Southern Hepatocyte Nuclear Extracts Contain an SP1-Like Protein

Whether the Sp1 sites bind SP1 protein in vivo is dependent initially on whether *Fundulus* hepatocytes contain

A

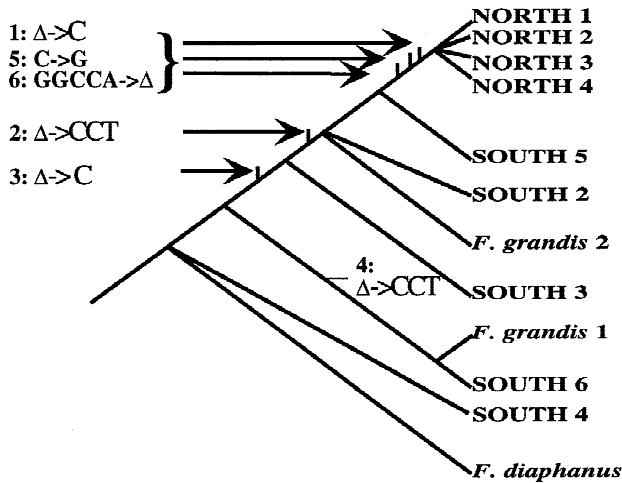


B *Ldh-B* Sp1 Region



C

Substitutions



D Northern and Southern Sp1 Oligos



Fig. 1. The *Fundulus Ldh-B* proximal promoter and natural sequence variation for Sp1 region. **A** The *Ldh-B* proximal promoter has Sp1 sites, is TATA-less, and has TCC repeats instead of a TATA box, and has two initiators (Inr). The two clusters of transcriptional start sites are indicated by large (major start sites) and small (minor start sites) arrows. **B** Sp1 region from three species of *Fundulus* (*F. heteroclitus*, *F. grandis*, and *F. diaphanus*). N and S refer to the northern and southern populations of *F. heteroclitus*, respectively. Fg and Fd refer to the othertwo species (*F. grandis* and *F. diaphanus*, respectively). Numbers and lines under sequences are traits that are consistently different for

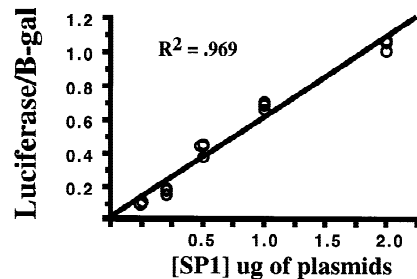
the maximum parsimony tree below (Fig. 1C, i.e., consistency index, CI = 1). Numbers 2 and 4 refer to deletions of the nucleotides CCT. Number 6 refers to the deletions of nucleotides GGCCA. Sequences contain four extra nucleotides at both 5' and 3' end that do not bind protein in vivo. Sequences were aligned with Clustal W in MacVector 6.0 (Oxford Molecular) and subsequently altered to minimize the number of gaps. **C** Evolutionary relationships for Sp1 sites (above, 1B). Maximum parsimony tree for Sp1 region (PAUP 3.1; Swofford 1989) and rooted with *F. diaphanus*. Tree has significant structure ($g = 0.824057$; $p > 0.01$; Huelsenbeck and Hillis 1992), and branch leading to

SP1 or a SP1-like protein (SP1-like: a SP1 protein with similar specificity and transcriptional activation). Humans, rats, and mice all express SP1 protein ubiquitously, however, *Drosophila* does not have such a protein. The extent to which other vertebrates, such as fish, have SP1 proteins is unknown. To address this question, hepatocyte nuclear extracts were made from the livers of northern or southern *F. heteroclitus* and used in mobility shift assays with the consensus Sp1 oligonucleotide (Fig. 4). There is a protein present in both northern and southern hepatocyte nuclear extracts that binds specifically to the consensus Sp1 site (lanes 2 and 12, Fig. 4). The specificity of binding is demonstrated by the ability of excess unlabeled Sp1 oligonucleotide to outcompete the shifted band (lanes 3–5 and 13–15, Fig. 4) and the inability of (1) a nonspecific unlabeled oligonucleotide (containing the CREB transcription factor binding site; lanes 9 and 19, Fig. 4); (2) a single-stranded oligonucleotide (lanes 10 and 20, Fig. 4); and (3) over 500-fold mass excess of herring sperm DNA or poly dI-dC to outcompete the shifted band. Furthermore, an excess of unlabeled SN, SS1, or SS2 is capable of outcompeting this shifted band (lanes 6–8 and 16–18, Fig. 4). The shifted oligo is composed of two bands of nearly identical mobility and is comparable to the pattern seen in mobility shifts with mammalian extract preparations. These two bands are typical of mobility shifts using cellularly derived SP1 due to post-translational modifications of the protein (Jackson and Tjian 1988; Jackson et al. 1993). Neither the northern nor the southern SP1-like protein interacts with the human Sp1-antibody (targeted to amino acids 436 to 454 of the human SP1 protein, Santa Cruz Biotechnology; data not shown). However, this pattern of two closely migrating bands, the absolute mobility of the doublet (migrating with a mobility identical to rhSP1) and the specificity of binding to the Sp1 consensus sequence together strongly indicate that this factor is a *Fundulus* SP1 or SP1-like protein.

The *Fundulus* SP1-like protein binds to SN, SS1, and SS2, however, non-SP1 factors bind as well. The effective competition of SN, SS1, and SS2 against the Sp1 oligo (lanes 6–8 and 16–18, Fig. 4) indicates that these oligos can bind the *Fundulus* SP1-like protein. To directly determine the interactions that occur between these oligos and the SP1-like protein in the northern and southern nuclear extracts, mobility shifts with the *Ldh-B* Sp1 oligos and the *Fundulus* extracts were performed with various unlabeled competitors (Fig. 5). It is appar-

Fig. 1. Continued. *F. heteroclitus* north has 100% bootstrap support ($n = 1,000$). All deletions or insertions, regardless of size, are scored as a single character in the phylogenetic analysis. Listed between nodes, "Substitutions" are changes that have a CI of 1.0. Numbers refer to changes underlined in the sequence alignment (Fig. 1B). Changes with CI < 1.0 that affect tree topology are not listed. Δ denotes a deletion or gap. Notice, except for *F. heteroclitus* north, none of the other taxa can be separated from each other. **D** Oligos used in gel shift assays. The sequences of the northern (SN oligonucleotide) and two southern

A Promoter Activity with Varying [SP1]



B

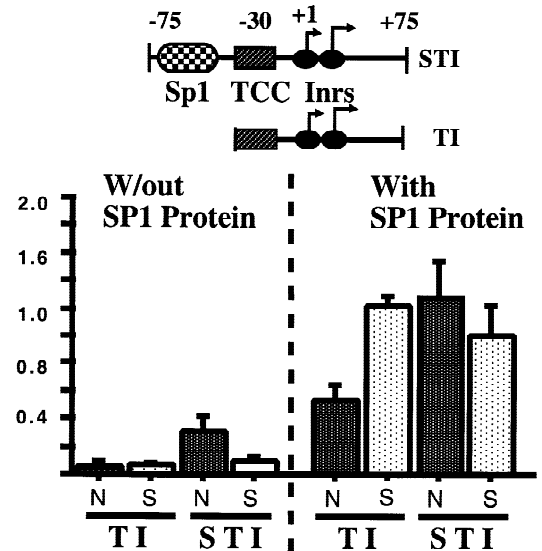


Fig. 2. Promoter strength is dependent on SP1 expression. Promoter strength (measured as the amount of luciferase expressed relative to the amount of cotransfected CMV β -Gal) was determined in *Drosophila* cells that lack SP1 expression (Courey and Tjian 1988; Hagen et al. 1994). **A** Relative promoter strength with different amounts (0, 0.25, 5, 1, and 2 μ g per flask) of human SP1 expression vector (pPacSp1) was cotransfected with STI-luciferase (Sp1, TCC repeat and Inr, Fig. 1A) and CMV β -gal. For each concentration of pPacSp1 cotransfected, there are three transfections measured in triplicate. Means for all three transfections are displayed. **B** STI and TI (lacking Sp1 like sequences) proximal promoters transfected into *Drosophila* cell lines with and without human SP1 expression vector. N and S refer to northern or southern proximal promoters, respectively.

ent that several complexes form between SN and the northern extract (lane 2, Fig. 5A). The slowest migrating band represents the bound SP1-like protein as increasing concentrations of the unlabeled Sp1 oligonucleotide specifically outcompetes just this band (lanes 4–6, Fig. 5A). Additionally, the band is a doublet and migrates at the

(SS1 and SS2 oligonucleotides)Sp1 regions are indicated. Sequences matching six of the six nucleotides of the consensus Sp1 site are indicated by horizontal lines with a shaded oval, while sequences matching five of the six nucleotides of the core consensus Sp1 site are indicated by horizontal lines. Vertical lines are sites with deleted sequences outlined in boxes (insertion #1 or deletion #6, relative to northern sequence). Shaded boxes, #5, are substitutions of G for C (S \rightarrow N) that removes the consensus Sp1 site. Deletions #2 and #3 within the southern Sp1 sites are not shown.

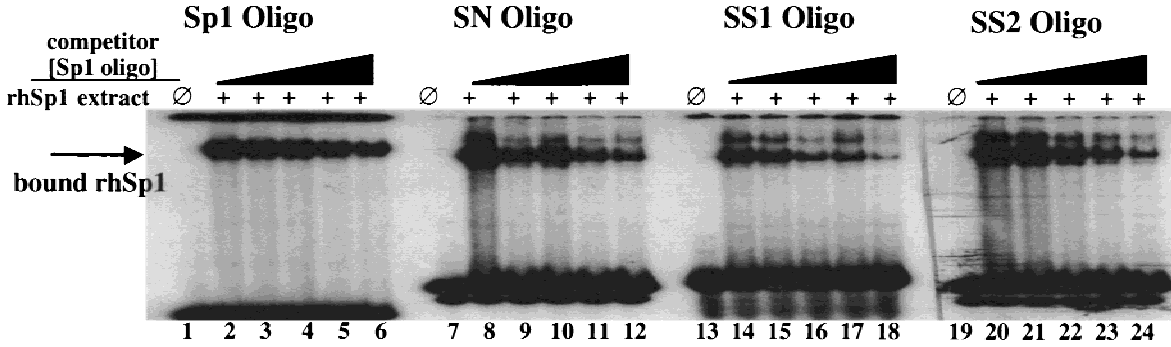


Fig. 3. Mobility shift analysis demonstrating the binding of rhSP1 to Sp1 sites from the northern and southern *Ldh-B* promoters. Recombinant purified SP1 protein binds and alters the mobility of labeled Sp1, SN, SS1, and SS2 oligonucleotides. SP1-oligo complexes were reduced by competing with unlabeled Sp1 oligonucleotide at increasing

amounts of (0-, 5-, 10-, 20-, 40-fold molar excess; lanes 2–6, 8–12, 14–18, and 20–24). + indicates the presence of rhSP1 protein. \emptyset is a negative control consisting of labeled oligonucleotide in the absence of rhSP1 protein (lanes 1, 7, 13, and 19). The shifted band(s) corresponding to bound rhSP1 are indicated by the arrow.

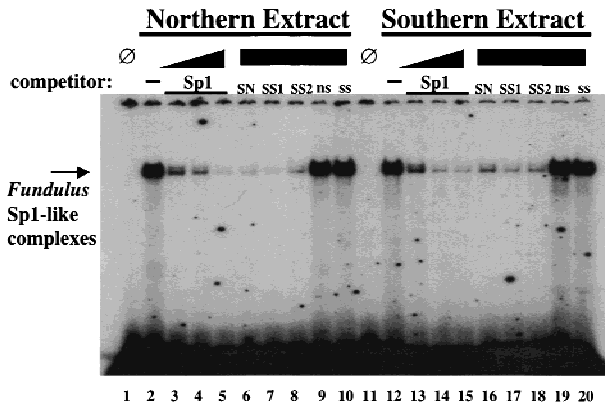


Fig. 4. Northern and southern populations have an SP1-like protein in hepatocyte nuclear extracts. Nuclear extracts from northern (lanes 2–10) or southern (lanes 12–20) populations specifically bind and alter the mobility of the labeled Sp1 oligonucleotide. Unlabeled competitor oligonucleotides were: Sp1 (10-, 25-, 50-fold molar excess; lanes 3, 4, 5 and 13, 14, 15), SN (50-fold molar excess; lanes 6 and 16), SS1 (50-fold molar excess; lanes 7 and 17), SS2 (50-fold molar excess; lanes 8 and 18), ns (a nonspecific binding site [CREB oligonucleotide]; 50-fold molar excess; lanes 9 and 19), and ss (a nonspecific single-stranded oligonucleotide; 50-fold molar excess; lanes 10 and 20). Negative controls, \emptyset , consisting of labeled oligonucleotide in the absence of northern or southern extract (lanes 1 and 11) are included. – indicates lack of competitor. The shifted band(s) corresponding to bound northern or southern SP1-like proteins are indicated by the arrow.

same mobility as the SP1 complex presented in Fig. 4. Interestingly, there are several faster migrating bands that are not outcompeted with the unlabeled Sp1 oligonucleotide. These complexes are specific protein:DNA interactions as they are outcompeted with unlabeled SN (as well as SS1 and SS2), but are not outcompeted with an unlabeled Sp1 oligo, nonspecific competitor, a single-stranded oligonucleotide (compare lanes 3, 7, and 8 to lanes 4–6, 9, and 10, Fig. 5A), poly dI-dC, or with genomic DNA. The results of mobility shifts with SN and the southern extract (lanes 12–20, Fig. 5A) are virtually identical to that seen with the northern extract (lanes

2–10, Fig. 5A). Thus, the SN oligonucleotide binds the *Fundulus* SP1-like protein as well as additional non-SP1 factors that are present in both northern and southern nuclear extracts.

Mobility shifts performed with the SS1 and SS2 oligonucleotides and northern and southern extracts reveal that similar multiple interactions occur (lanes 2, 5, 9 and 12, Fig. 5B). In all cases, the slowest migrating band is the SP1-specific complex (the only complex outcompeted by unlabeled Sp1 oligonucleotide; lanes 4, 7, 11, and 14, Fig. 5B), and the faster migrating complexes are the result of bound non-SP1 factors. Importantly, the degree of non-SP1 binding relative to SP1 protein binding is different among the three *Ldh-B* oligonucleotides. The interaction between the non-SP1 factors and SN are the strongest followed by SS1, then SS2. This can be seen by comparing the intensity of the SP1-specific band to non-SP1 bands for SN (lanes 2 and 12, Fig. 5A) to those for SS1 and SS2 (lanes 2, 5, 9, and 12, Fig. 5B). The SS2 oligonucleotide binds very little non-SP1 factor from either extract. Thus, both the *Fundulus* SP1-like protein and non-SP1 factors bind to the southern Sp1 sites; however, sequence variation between the northern and southern oligonucleotides results in different degrees of non-SP1 factor binding.

Non-SP1 Factors Bind to the 5' GA Subregion

The *Ldh-B* Sp1 sites have two general subregions: the 5' GA region and the 3' GC region. To begin to localize where SP1-like and non-SP1 factors are binding, two oligonucleotides were designed that correspond either to the GA-region of SN (SNga, Fig. 6A) or to the GC-region of SN (SNgc, Fig. 6A). These oligonucleotides were used in mobility shift assays as labeled probes and as unlabeled competitor (Fig. 6B). The SNga oligonucleotide binds both the SP1-like protein and the non-SP1 factors, whereas the SNgc oligonucleotide binds

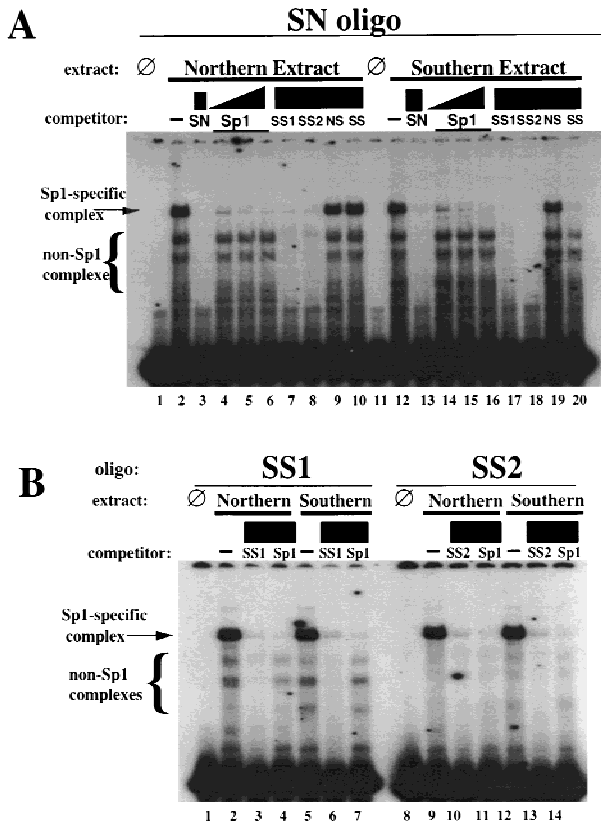


Fig. 5. *Fundulus* SP-1 like Protein. *Fundulus* nuclear extracts have SP1-like proteins that bind the northern and southern Sp1 sites and contain additional non-SP1 factors. **A** Nuclear extracts from northern (lanes 2–10) and southern (lanes 12–20) *Fundulus* populations bind and alter the mobility of a labeled SN oligonucleotide. The specificity of binding was determined by competing with unlabeled competitors. Competitor oligonucleotides were: SN (50-fold molar excess; lanes 3 and 13); Sp1 (10, 25, 50-fold molar excess; lanes 4, 5, 6 and 14, 15, 16), SS1 (50-fold molar excess; lanes 7 and 17), SS2 (50-fold molar excess; lanes 8 and 18), NS (a nonspecific binding site [CREB oligonucleotide]; 50-fold molar excess; lanes 9 and 19), and SS (a nonspecific single-stranded oligonucleotide; 50-fold molar excess; lanes 10 and 20). Negative controls, ∅, consisting of labeled oligonucleotide in the absence of northern or southern extracts (lanes 1 and 11) are included. – indicates no competitor. The shifted band(s) corresponding to bound northern or southern SP1-like protein are indicated by the arrow, while shifted bands corresponding to bound, non-SP1 factor are indicated by the { symbol. **B** Labeled SS1 or SS2 oligonucleotide was incubated with either northern nuclear extract (lanes 2, 3, 4 and 9, 10, 11) or southern nuclear extract (lanes 5, 6, 7 and 12, 13, 14). Unlabeled competitor oligonucleotides were: SS1 (50-fold molar excess, lanes 3 and 6); SS2 (50-fold molar excess, lanes 10 and 13); and Sp1 (50-fold molar excess, lanes 4, 7, 11, and 14). Negative controls consisting of labeled oligonucleotide in the absence of northern or southern extract (lanes 1 and 8) are included. The shifted band(s) corresponding to bound northern or southern SP1-like protein are indicated by the arrow, while shifted bands corresponding to bound, non-SP1 factor are indicated by the { symbol.

only the SP1-like protein (compare lanes 1 and 2, Fig. 6B). This is further demonstrated by using these oligonucleotides as unlabeled competitors against the full-length labeled SN oligonucleotide. The SNga oligonucleotide outcompetes all the SN complexes (lanes 4

and 5, Fig. 6B). However, the SNgc oligonucleotide outcompetes only the SP1-like complex (lanes 6 and 7, Fig. 6B), which is similar to the competition with the consensus Sp1 oligonucleotide (lane 8, Fig. 6B).

The Northern and Southern Sp1 Sites Have Different Affinities for rhSP1

Disassociation constants (K_d) were determined to investigate whether the sequence variability in the Sp1 sites affects the affinity of SP1 protein binding. The multiple interactions that occur between *F. heteroclitus* nuclear extracts and the Sp1 sites precluded an accurate determination of the K_d between the *Fundulus* SP1-like protein and the *Ldh-B* Sp1 sites. Instead, the effect of *Ldh-B* sequence variability was addressed by measuring the affinity of rhSP1 to north or south *Fundulus* Sp1 sites. To determine if there was a difference between north and south *Fundulus* SP1 proteins, K_d s were determined for these proteins binding to the consensus Sp1 site.

The mobility shift assay was used to separate bound and free oligonucleotide. In each lane the shifted band represents a 1:1 stoichiometry of protein to oligo (i.e., only a single protein is bound per oligo; Fig. 7A). The bound and free oligo were quantified and, along with nonlinear regression analysis, were used to determine K_d of the different *Ldh-B* Sp1 sites for the rhSP1 protein. A representative mobility shift autoradiogram is shown (Fig. 7A) along with a plot of bound versus total oligonucleotide (Fig. 7B). The nonlinear regression of bound versus total is used to calculate the K_d . In this way, relative affinities were generated for the binding of rhSP1 to the oligonucleotides of interest (Fig. 7C). Each oligonucleotide was measured at least three times, and the reported values are the average of these separate experiments for the indicated oligonucleotide.

The consensus Sp1 oligonucleotide had the lowest affinity for rhSP1 (i.e., highest K_d ; $K_d = 9.88$ nM), followed by SS2 ($K_d = 8.67$ nM), SS1 ($K_d = 7.0$ nM), and SN with the highest affinity ($K_d = 5.51$ nM). The affinity of the SN oligonucleotide for rhSP1 was significantly greater than both Sp1 and SS2 (approximately twofold greater affinity; $p = 0.01$ and 0.05 , respectively, after using Fisher PLSD to correct for multiple comparison). Although SN had a higher affinity than SS1, this difference was not significant ($p = 0.32$), primarily due to the greater variance for the SS1 (see standard error of the mean, SEM, in Fig. 7C). Thus, sequence variability in the *Ldh-B* Sp1 sites (SN, SS1, and SS2) results in different affinities for the rhSP1 protein.

To determine if the affinity difference between the SN and Sp1 oligonucleotides was due to multiple sites in the SN oligonucleotide, a new oligonucleotide was constructed containing multiple consensus Sp1 sites. This oligonucleotide was named Sp1_N and consisted of the SN oligonucleotide with GC sequences replacing all the GA

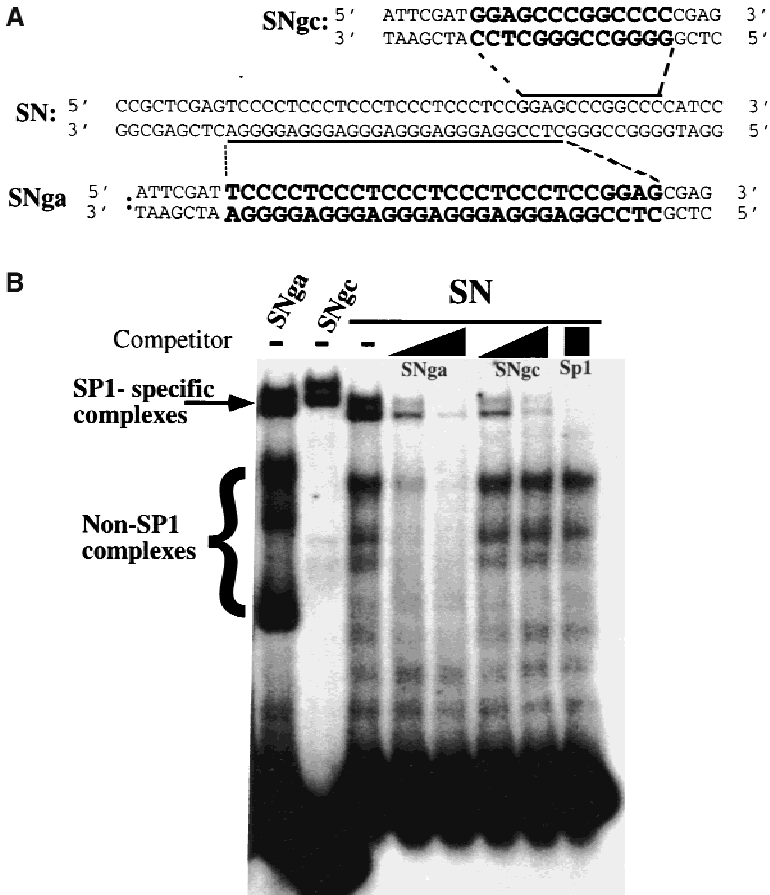


Fig. 6. Protein interacts with GA and GC subregions. SP1-like protein binds both GC and GA subregions, but non-SP1 proteins only bind to the GA subregion. **A** The sequences for oligonucleotides SNGa and SNGc are indicated and their relationship to the original SN oligonucleotide. Bold sequences are derived from SN. Nonbolded sequences are the same in both SNGa and SNGc. **B** Protein:DNA interactions using northern nuclear extract and SN oligo or subregions of SN: SNGa and SNGc. Labeled SNGa (lane 1), SNGc (lane 2), or SN (lanes 3–8) were incubated with northern nuclear extracts. Unlabeled competitor oligonucleotides were: SNGa (10-, 50-fold molar excess; lanes 4 and 5, respectively); SNGc (10-, 50-fold molar excess; lanes 6 and 7, respectively); and Sp1 (50-fold molar excess; lane 8). – indicates lack of competitor. The shifted band(s) corresponding to bound SP1-like protein are indicated by the arrow, while shifted bands corresponding to bound, non-SP1 factor are indicated by the { symbol.

sequences (Sp1_N sequence: 5' CCGCTCGAGTCCCCGCCCCGCCCGCCCCGCCCGCCGGAGCCCGGCCCATCC 3'). These replacements create an oligo with five consensus Sp1 binding sites. Analyses of the Sp1_N indicates that this oligonucleotide had a $K_d = 9.77$ nm for rhSP1. This corresponds to a significantly lower affinity than SN ($p = 0.01$ after correction for multiple comparisons). In fact, the Sp1_N oligonucleotide has a lower affinity than any of the other *Fundulus* Sp1 oligonucleotides tested and is essentially the same as the consensus Sp1 oligo.

To determine the relative affinities of the *Fundulus* northern and southern SP1-like proteins, K_d s were determined with the consensus Sp1 oligonucleotide and northern or southern nuclear extracts. There is only a single protein:DNA interaction when using these *Fundulus* nuclear extracts with the consensus Sp1 oligo nucleotide (Fig. 4). A comparison of the K_d s of rhSP1 and northern or southern SP1-like protein for the Sp1 oligonucleotide revealed that the *Fundulus* SP1-like proteins have a considerably higher affinity than rhSP1 (approximately two-fold higher; $p = 0.005$, Fig. 7C). Although the northern SP1-like protein has a greater affinity (lower K_d) than the southern SP1-like protein for the Sp1 oligonucleotide, they are not significantly difference ($p = 0.06$, Fig. 7C).

Discussion

The hallmark of most promoter sequences is their conservation among different species (Azizkhan et al. 1993; Gumucio et al. 1993, 1996). Consistent with this, there are not many differences in *Ldh-B* Sp1 sites among *Fundulus* species that diverged millions of years ago. However, this is not the case for the northern population of *F. heteroclitus* where there are many changes in *Ldh-B* Sp1 sites: three unique fixed differences that cause the northern Sp1 region to segregate phylogenetically from the southern population and all other species (Fig. 1C). These changes eliminate a consensus Sp1 binding site and increase the number of GA Sp1-like binding motifs. Nonetheless, these Sp1 sites from both populations bind protein in vivo, in vitro and affect transcription in three different fish cell lines (Segal et al. 1996; Crawford et al. 1999b). Additionally, these changes are associated with the observed difference in *Ldh-B* gene expression: the northern population expresses more *Ldh-B* than these other taxa. Importantly, these sequence changes (together with sequence differences in the other functional regions in the proximal promoter) are the result of directional selection (Crawford et al. 1999b). Thus, the relatively large number of changes in the *Ldh-B* Sp1 region

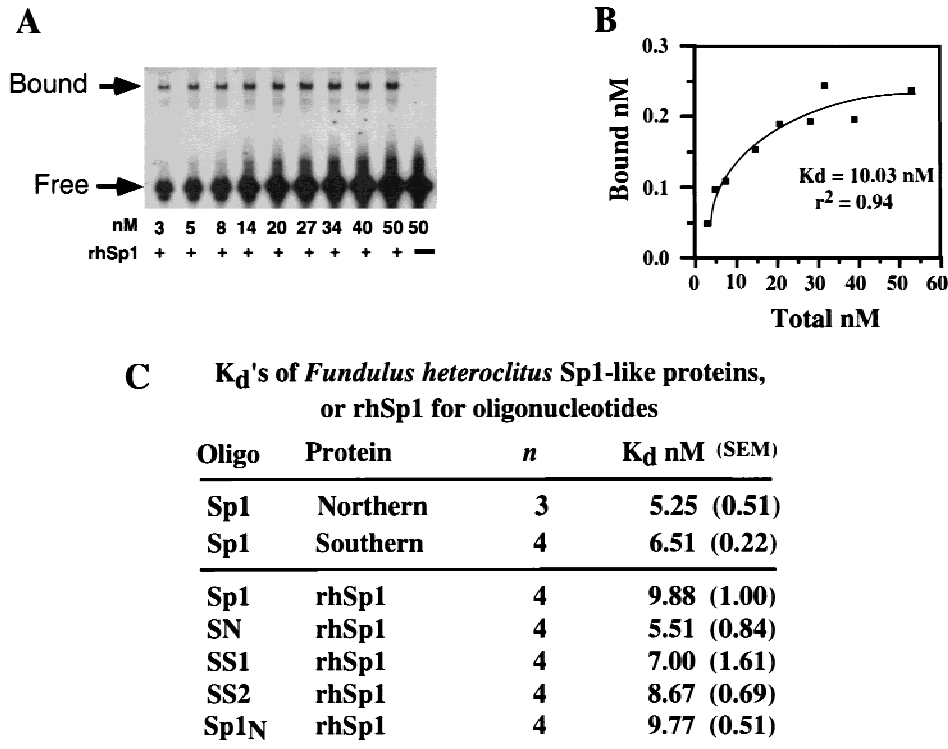


Fig. 7. Analysis of binding affinity. **A** A representative mobility shift with increasing amounts of labeled SS2 oligonucleotide (lanes 1–9 with 3, 5, 7, 14, 20.5, 27, 33.5, 40, and 50 nM of oligo respectively) and a constant amount of purified recombinant human SP1 (rhSP1; 0.013 footprinting units). A negative control consisting of 50 nM labeled oligonucleotide in the absence of rhSP1 (lane 10) is included. **B** Binding curves for SS2 oligo with rhSP1. Amounts of bound oligonucleotide were quantified using Phosphorimager. The amount of bound versus total oligonucleotide per lane is plotted. **C** Disassociation constants (K_d) for different oligos with rhSP1 or nuclear extracts with the consensus Sp1 oligo. Oligos are shown in Fig. 1. Protein refers either to

nuclear extract from northern or southern populations of *F. heteroclitus* or rhSP1. *n* refers to the number of experimental determinations. K_d s are in nM with standard error of mean in parentheses. Nonlinear analyses of binding affinities were generated from the data for each experimental gel, and the K_d determined with the use of GraphPad Prism 2.0 software (GraphPad). Each oligonucleotide:protein interaction was tested in at least three separate experiments and the mean K_d was calculated from these separate measures. Statistical significance of differences between different oligonucleotide K_d s were determined by analysis of variance (ANOVA, with correction for multiple comparisons) using Minitab 10.1.

from northern populations of *F. heteroclitus* are not due to random genetic effects and most likely effect a change in transcription.

Functional Consequence of LDH-B Sequence Variation

The sequence variation in the *Ldh-B* Sp1 sites affects transcription in hepatoma, cardiac, and embryonic fish cell lines (Segal et al. 1996; Crawford et al. 1999a). This likely reflects binding and transactivation of a *Fundulus* SP1-like factor because: (1) *Ldh-B* proximal promoter activity is dependant on expression of SP1 (Fig. 2); (2) *Ldh-B* Sp1 sites bind rhSP1 (Fig. 3); and (3) *Fundulus* hepatocyte nuclear extracts express a protein that binds to the *Ldh-B* Sp1 sites similarly to rhSP1 (Fig. 4; see further discussion below: *Fundulus* SP1 Protein). The *Ldh-B* Sp1 sequence variation affects rhSP1 binding affinity (Fig. 7C) and the apparent binding of non-SP1 proteins (compare Fig. 5A to 5B). Thus, these differences between northern and southern sequences both

within and between cell lines demonstrate that sequence divergence in this region is capable of effecting modest differences in transcription factor binding. The implication is that the natural sequence variation in the Sp1 sites produce these observed differences between populations in the *Ldh-B* transcription rate. Studies on other genes also suggest that Sp1 sequence variation may play a role in modulating transcription. For example, a single mutation in the long terminal repeat (LTR) of the Moloney murine sarcoma virus (Mo-MSV) creates a perfect Sp1 site that allows for efficient transcription in embryonic carcinoma cells that other wise does not occur (Prince and Rigby 1991). Additionally, sequence variation in a proximal Sp1 site of the human low-density lipoprotein receptor gene promoter has been linked to familial hypercholesterolemia (Koivisto et al. 1994). These studies indicate that naturally occurring modifications of proximal Sp1 sites may be an important mechanism for changing transcription rate.

The sequence variation in the *Ldh-B* Sp1 sites affects

SP1 protein binding affinities. The dissociation constant (K_d) for rhSP1 binding to the Sp1 oligo is approximately 9.9 nM (Fig. 7C). This is similar to the reported values for K_d s of SP1, which range from 0.1 nM to 10 nM (reviewed in Sogawa et al., [1993]). The rhSP1 K_d reported here is at the higher range (thus lower affinity) of these values, but is comparable to that determined by Kriwacki et al. (1992). Differences among these studies are likely to be due to both different assay systems (i.e., mobility shift versus footprinting) and different assay conditions (i.e., salt concentration in binding buffer, temperature of binding reaction, etc.). The results presented here indicate that the northern Sp1 (SN) has an approximate twofold higher affinity than either the consensus Sp1 or southern (SS2) sequences (Fig. 7C). These results were unexpected because the northern sequences lack the consensus Sp1 site (GGGCGG) and the southern sequences and the Sp1 oligo all contain it (Fig. 1; Letovsky and Dynan 1989; Kutoh et al. 1993; Chang et al. 1996). Additionally, the order of K_d s among *Fundulus* sequences (SN, SS1, SS2; by decreasing affinity) correlates with the number of GA motifs (GGGAGG). Thus, northern Sp1 sites have a greater affinity for SP1 by virtue of two unique changes: the substitution of a G for a C in the consensus Sp1 site (#5, Fig. 1) and the insertion of a C, which creates an additional GA motif (#1, Fig. 1).

It may be that the K_d s reported here reflect binding to the GA motifs. The *Fundulus* SP1 protein does bind to the GA motifs of the sequence as well as the GC motifs (Fig. 6). Similarly, mammalian SP1 is known to bind to variations of the canonical sequence in other promoters (Jones et al. 1986; Evans et al. 1988; Kriwacki et al. 1992). The binding of SP1 to the GA motif is transcriptionally important in the promoters of the rat transforming growth factor- α gene, the human c-myc gene, and the insulin-like growth factor binding protein-2 gene (Chen et al. 1992; DesJardins and Hay 1993; Kutoh et al. 1993). Kriwacki et al. (1992) report that a GA sequence binds SP1 protein with high affinity; however, several other studies have indicated that the GA sequence binds SP1 protein with a lower affinity than the GC sequence (Letovsky and Dynan 1989; Kutoh et al. 1993; Chang et al. 1996). Importantly, these studies have measured binding to single binding sites, whereas in our study the *Ldh-B* oligonucleotides each contain multiple binding sites. Synergy in binding cannot explain the results here because in these assays, only one protein molecule is bound per DNA molecule (despite the presence of multiple sites per DNA molecule). Furthermore, Pascal and Tjian (1991) determined that SP1 protein does not bind synergistically, although they did detect synergy in transactivation. It may be that the presence of multiple sites increases stability of the protein: DNA complex or that multiple sites allow for sliding of the protein from site to site. However, when the GA motifs in the SN oligo were

converted to consensus Sp1 sites (creating the Sp1_N oligonucleotide), this Sp1_N also had a significantly lower affinity than the SN oligonucleotide. It is interesting to speculate that, although a single GA motif has a lower affinity for SP1 than a single consensus Sp1 site (Letovsky and Dynan 1989; Kutoh et al. 1993; Chang et al. 1996), multiple overlapping GA motifs may result in an apparent higher affinity than either a single or multiple consensus Sp1 sites. This is supported by the fact that K_d s correlate with the total number of GA motifs in the *Fundulus Ldh-B* northern and southern sequences. Regardless of the mechanism, it is clear that sequence variability in the *Ldh-B* Sp1 sites results in significantly different affinities for the rhSP1 protein.

In addition to binding SP1 protein, the *Ldh-B* Sp1 sites bind several non-SP1 factors from *Fundulus* hepatocyte nuclear extracts (Fig. 5). Many naturally occurring Sp1 binding regions bind non-SP1 factors (e.g., ETF [Kageyama et al. 1989], Krox-24 and Krox-20 [Faisst and Meyer 1992], BTEB [Imataka et al. 1992], GCF [Kageyama et al. 1989], Wnt-1 [St-Arnaud and Moir 1993], and MAZ [Bossone et al. 1992]). These results indicate that non-SP1 factors binding to Sp1 sites can contribute to differential transcriptional regulation. For the *Ldh-B* Sp1 region there is variation in the extent of non-SP1 binding to the *Ldh-B* northern and southern sequences (Fig. 5), with SN binding the most, followed by SS1 and SS2. This non-SP1 binding was localized exclusively to the GA motif region of the oligonucleotides (Fig. 6). An examination of these GA sequences among oligonucleotides (Fig. 1 and 3) reveals four GA motifs in SN because of a single C insertion (#1 Fig. 1C), while the southern sequences contain either three in SS1 or, because of variation in the length of GA region, only two in SS2. Thus, the extent of non-SP1 binding is correlated with sequence variation in the region of binding.

The *Fundulus* SP1-like protein also binds to the GA motif of these sequences (Fig. 6). The data presented here does not distinguish whether SP1 and non-SP1 factors bind in a mutually exclusive manner, cooperatively, or regardless of the other. However, the regulation of many genes is achieved in part by interactions between SP1 and other factors binding to overlapping sites. The Wilms tumor protein WT1 represses the colony-stimulating factor-1 gene when WT1 outcompetes either SP1 or SP3 for their respective overlapping binding sites (Harrington et al. 1993). Similarly, gastrin gene transcription is repressed in islet cells when AT-binding factor binds to its binding site adjacent to a Sp1 binding sites (Chung et al. 1995), putatively by interfering with SP1 transactivation. Upregulation of platelet-derived growth factor β -chain (PDGF- β) expression is achieved when the early-growth-response gene product (Egr-1) binds to its binding site in the PDGF- β proximal promoter and displaces SP1 bound to overlapping sequences (Khachigian et al. 1996). Finally, rat growth hormone promoter activity is dependent on the integrity of over-

lapping pituitary-specific transcription factor (PIT-1) and Sp1 binding sites, even though both proteins cannot be bound at the same time (Schaufele et al. 1990). These examples demonstrate the wide variety of regulatory mechanisms involving SP1 and non-SP1 factors binding to overlapping or adjacent sequences. If similar interactions occur in the *Fundulus Ldh-B* proximal promoter, the different extent of non-SP1 factor binding between northern and southern sequences would affect transcription rate. Furthermore, the differential strength of northern and southern promoter sequences across different fish cell lines (Segal et al. 1996; Crawford et al. 1999b) may be directly related to the degree or type of non-SP1 factor expressed in those cell types.

Fundulus Sp1-Like Protein

Whether the *Fundulus* nuclear protein that binds Sp1 site is homologous to mammalian SP1 protein remains to be determined. Nonetheless, it is similar to mammalian SP1 in that (1) it binds specifically to the Sp1 consensus sequence (GGGCGG); (2) when bound to the Sp1 consensus sequence it appears as two closely migrating bands (presumably due to post-translational modifications); and (3) the absolute mobility of this doublets is equivalent to the human SP1. This *Fundulus* protein has a higher affinity to the consensus Sp1 sequence than the rhSP1 protein (expressed and purified from HeLa cells). Additionally, the *Fundulus* nuclear protein binds in vivo at the same position as rhSP1 protein bind in vitro (Crawford et al. 1999b).

The affinity of the northern and southern *Fundulus* SP1-like protein for the consensus Sp1 site are similar to each other and approximately twice that of rhSP1 protein (5.2 and 6.5 versus 9.9 nm, Fig. 7C). The high affinity of the *Fundulus* SP1-like protein to the consensus Sp1 sequence strongly suggests that this nonmammalian vertebrate has an SP1 homologue. However, an antibody raised against human SP1 (Santa Cruz Biotechnology) does not cross-react with these *Fundulus* SP1-like proteins or any other *Fundulus* nuclear proteins. The antibody does interact with a low molecular weight cytoplasmic protein from *Fundulus* (data not shown). The antibody used here is polyclonal and directed toward the epitope consisting of amino acid residues 436–454 of the human SP1 protein. This epitope is distinct from the COOH-terminal zinc fingers that are exclusively responsible for sequence specific binding (Kadonaga et al. 1988). Furthermore, these amino acids do not appear to be responsible for transcriptional activation, but may be important for the affinity of SP1 binding to its specific binding sequence (Kadonaga et al. 1988). Amino acid replacements in this region could occur without significantly altering SP1 protein function (i.e., binding specificity and transactivational properties). Thus, the differences in the amino acid sequence of the *Fundulus* SP1-like proteins that prevent antibody recognition may

well be responsible for the higher affinity. An alternative explanation for the higher affinity of the *Fundulus* SP1-like proteins is that they have undergone post-translational modification that is not present in the rhSP1 expressed in HeLa cells. Glycosylation of mammalian SP1 does not appear to have any effect on binding affinity (Jackson and Tjian 1988); however, phosphorylation has been reported to affect the binding affinity of both mouse and rat SP1 (Leggett et al. 1995; Daniel et al. 1996). An alternative explanation is that there is no *Fundulus* homologue to SP1. Without primary sequence of the *Fundulus* protein that binds Sp1 sites it is uncertain if it is SP1, another member of the SP1 family or a unique GC binding protein. Regardless of the eventual designation of this *Fundulus* protein, it nonetheless binds Sp1 sites with a greater affinity than rhSP1 and should be similarly affected by the sequence variation. More extensive characterization of the *Fundulus* SP1-like protein should provide information about this matter and contribute to our understanding of the evolution of the SP1 protein.

Conclusion

The molecular mechanisms responsible for variation in constitutive transcription rates and the functional significance of intraspecific sequence diversity in proximal promoters are largely unknown. The results presented here demonstrate that naturally occurring divergence in Sp1 binding sites of the *F. heteroclitus Ldh-B* proximal promoter has functional consequences for the extent and affinity with which specific proteins (both SP1 and non-SP1) bind and is correlated with differential transcription rate in several cell lines. These data suggest that the teleost fish *Fundulus* has a nuclear protein that binds to Sp1 sites with similar specificity and greater affinities than human SP1 protein. Together, these data demonstrate that natural variation in Sp1 sites may be an important mechanism for the subtle modulation of transcription rate.

Acknowledgment. This research benefited from helpful discussions with Drs. Mark Q. Martindale, Steve Hand, Radovan Zak, and two anonymous reviewers. Kevin Kolell assisted in the analysis of SP1 binding affinities and Margie Oleksiak was responsible for western analyses and cross-reactivity of SP1 antibodies. This research was supported by IBN 9419781 to DLC. JAS was supported by NIH predoctoral training grant GM 07183.

References

- Arkhipova IR (1995) Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* 139:1359–1369
- Azizkhan JC, Jensen DE, Pierce AJ, Wade M (1993) Transcription from TATA-less promoters: dihydrofolate reductase as a model. *Crit Rev Eukar Gene Exp* 3:229–254
- Bernardi G, Powers DA (1995) Phylogenetic relationships among nine species from the genus *Fundulus* (Cyprinodontiformes, Fundulidae) inferred from sequences of the cytochrome B. *Copeia* 1995: 469–473

- Boam DS, Davidson I, Chambon P (1995) A TATA-less promoter containing binding sites for ubiquitous transcription factors mediates cell type-specific regulation of the gene for transcription enhancer factor-1 (TEF-1). *J Biol Chem* 270:19487–19494
- Bossone SA, Asselin C, Patel AJ, Marcu KB (1992) MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci USA* 89:7452–7456
- Cashner RC, Rogers JS, Grady JM (1992) Phylogenetic studies of the genus *Fundulus*. In: Mayden RL (ed) Systematic, historical ecology and North American freshwater fishes. Stanford University Press, Stanford, CA, pp 421–437
- Chang R, Yang E, Chamblis D, Kumar A, Wise J, Mehta KD (1996) In vivo role of the Sp1 site neighboring sterol-responsive element-1 in controlling low-density lipoprotein receptor gene expression. *Biochem Biophys Res Commun* 218:733–739
- Chen X, Azizkhan JC, Lee DC (1992) The binding of transcription factor Sp1 to multiple sites is required for maximal expression from the rat transforming growth factor alpha promoter. *Oncogene* 7:1805–1815
- Chung KC, Huang D, Chen Y, Short S, Short ML, Zhang Z, Jungmann RA (1995) Identification of a silencer module which selectively represses cyclic AMP-responsive element-dependent gene expression. *Mol Cell Biol* 15:6139–6149
- Colgan J, Manley JL (1995) Cooperation between core promoter elements influences transcriptional activity in vivo. *Proc Natl Acad Sci USA* 92:1955–1959
- Courey AJ, Tjian R (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* 55:887–898
- Crawford DL, Powers DA (1989) Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*. *Proc Natl Acad Sci USA* 86:9365–9369
- Crawford DL, Powers DA (1992) Evolutionary adaptation to different thermal environments via transcriptional regulation. *Mol Biol Evol* 9:806–813
- Crawford DL, Pierce VA, Segal JA (1999a) Evolutionary physiology of closely related taxa: analyses of enzyme expression. *Am Zool* 32:389–400
- Crawford DL, Segal JA, Barnett JL (1999b) Evolutionary analysis of TATA-less proximal promoter function. *Mol Biol Evol* 16:194–207
- Daniel S, Zhang SY, Depauloiraach AA, Kim KH (1996) Dephosphorylation of Sp1 by protein phosphatase 1 is involved in the glucose-mediated activation of the acetyl-coa carboxylase gene. *J Biol Chem* 271:14692–14697
- DesJardins E, Hay N (1993) Repeated CT elements bound by zinc finger proteins control the absolute and relative activities of the two principal human c-myc promoters. *Mol Cell Biol* 13:5710–5724
- Dynan WS, Tjian R (1983) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* 35:79–87
- Endler JA (1986) Natural selection in the wild. Princeton University Press, Princeton, NJ
- Ernst P, Smale ST (1995) Combinatorial regulation of transcription II: the immunoglobulin mu heavy chain gene. *Immunity* 2:427–438
- Evans T, DeChiara T, Efstratiadis A (1988) A promoter of the rat insulin-like growth factor II gene consists of minimal control elements. *J Mol Biol* 199:61–81
- Faisst S, Meyer S (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res* 20:3–26
- Goodrich JA, Cutler G, Tjian R (1996) Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell* 84:825–830
- Gorski K, Carneiro M, Schibler U (1986) Tissue-specific in vitro transcription from the mouse albumin promoter. *Cell* 47:767–776
- Gumucio DL, Shelton DA, Bailey WJ, Slightom JL, Goodman M (1993) Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the epsilon-globin gene. *Proc Natl Acad Sci USA* 90:6018–6022
- Gumucio DL, Shelton DA, Blanchard-McQuate K, Gray T, Tarle S, Heilstedt-Williamson H, Slightom JL, Collins F, Goodman M (1994) Differential phylogenetic footprinting as a means to identify base changes responsible for recruitment of the anthropoid gamma gene to a fetal expression pattern. *J Biol Chem* 269:15371–15380
- Gumucio DL, Shelton DA, Zhu W, Millinoff D, Gray T, Bock JH, Slightom JL, Goodman M (1996) Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the beta-like globin genes. *Mol Phylogen Evol* 5:18–32
- Hagen G, Muller S, Beato M, Suske G (1994) Sp1-mediated transcriptional activation is repressed by Sp3. *EMBO J* 13:3843–3851
- Harrington MA, Konicek B, Song A, Xia XL, Fredericks WJ, Rauscher FJ (1993) Inhibition of colony-stimulating factor-1 promoter activity by the product of the Wilms' tumor locus. *J Biol Chem* 268:21271–21275
- Hu J, Qazzaz H, Brennan MD (1995) A transcriptional role for conserved footprinting sequences within the larval promoter of a *Drosophila* alcohol dehydrogenase gene. *J Mol Biol* 249:259–269
- Huelsenbeck JP, Hillis DM (1992) Signal, noise, reliability in molecular phylogenetic analyses. *J Heredity* 83:189–195
- Imataka H, Sogawa K, Yasumoto K, Kikuchi Y, Sasano K, Kobayashi A, Hayami M, Fujii-Kuriyama Y (1992) Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. *EMBO J* 11:3663–3671
- Jackson SP, Tjian R (1988) O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* 55:125–133
- Jackson S, Gottlieb T, Hartley K (1993) Phosphorylation of transcription factor Sp1 by the DNA-dependent protein kinase. *Adv Second Mess Phosphop Res* 28:279–286
- Jolliff K, Li Y, Johnson LF (1991) Multiple protein-DNA interactions in the TATAA-less mouse thymidylate synthase promoter. *Nuc Acids Res* 19:2267–2274
- Jones KA, Kadonaga JT, Luciw PA, Tjian R (1986) Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* 232:755–759
- Kadonaga JT, Carner KR, Masiarz FR, Tjian R (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* 51:1079–1090
- Kadonaga JT, Courey AJ, Ladika J, Tjian R (1988) Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* 242:1566–1570
- Kageyama R, Merlino GT, Pastan I (1989) Nuclear factor ETF specifically stimulates transcription from promoters without a TATA box. *J Biol Chem* 264:15508–15514
- Khachigian LM, Lindner V, Williams AJ, Collins T (1996) Egr-1-induced endothelial gene expression: a common theme in vascular injury. *Science* 271:1427–1431
- Koivisto UM, Palvimo J, Janne OA, Kontula K (1994) A single-base substitution in the proximal Sp1 site of the human low density lipoprotein receptor promoter as a cause of heterozygous familial hypercholesterolemia. *Proc Natl Acad Sci USA* 91:10526–10530
- Kollmar R, Sukow KA, Sponagle SK, Farnham PJ (1994) Start site selection at the TATA-less carbamoyl-phosphate synthase (glutamine-hydrolyzing)/aspartate carbamoyltransferase/dihydroorotase promoter. *J Biol Chem* 269:2252–2257
- Kriwacki RW, Schultz SC, Steitz TA, Caradonna JP (1992) Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1. *Proc Natl Acad Sci USA* 89:9759–9763
- Kutuh E, Margot JB, Schwander J (1993) Genomic structure and regulation of the promoter of the rat insulin-like growth factor binding protein-2 gene. *Mol Endocrin* 7:1205–1216
- Leggett RW, Armstrong SA, Barry D, Mueller CR (1995) Sp1 is phos-

- phorylated and its DNA binding activity down-regulated upon terminal differentiation of the liver. *J Biol Chem* 270:25879–25884
- Letovsky J, Dynan WS (1989) Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. *Nuc Acids Res* 17:2639–2653
- Li W-H (1997) *Molecular evolution*. Sinauer Associates, Sunderland, MA
- Lu J, Lee W, Jiang C, Keller EB (1994) Start site selection by Sp1 in the TATA-less human Ha-ras promoter. *J Biol Chem* 269:5391–5402
- O'Shea-Greenfield A, Smale ST (1992) Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J Biol Chem* 267:6450
- Pascal E, Tjian R (1991) Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes & Develop* 5:1646–1656
- Pierce VA, Crawford DL (1997a) Phylogenetic analysis of glycolytic enzyme expression. *Science* 275:256–259
- Pierce VA, Crawford DL (1997b) Phylogenetic analysis of thermal acclimation of the glycolytic enzymes in the genus *Fundulus*. *Physiol Zool* 70:597–609
- Prince VE, Rigby PW (1991) Derivatives of Moloney murine sarcoma virus capable of being transcribed in embryonal carcinoma stem cells have gained a functional Sp1 binding site. *J Virol* 65:1803–1811
- Pugh BF, Tjian R (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes & Develop* 5:1935–1945
- Roeder RG (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem Sci* 16:402–408
- Schaufele F, West BL, Reudelhuber TL (1990) Overlapping Pit-1 and Sp1 binding sites are both essential to full rat growth hormone gene promoter activity despite mutually exclusive Pit-1 and Sp1 binding. *J Biol Chem* 265:17189–17196
- Segal JA, Crawford DL (1994) LDH-B enzyme expression: the mechanisms of altered gene expression in acclimation and evolutionary adaptation. *Am J Physiol* 267:R1150–R1153
- Segal JA, Schulte PM, Powers DA, Crawford DL (1996) Descriptive and functional characterization of variation in the *Fundulus heteroclitus* *Ldh-B* proximal promoter. *J Exp Zool* 275:355–364
- Smale ST (1994) Core promoter architecture for eukaryotic protein-coding genes. In: Conaway RC, Conaway JW (eds) *Transcription: mechanisms and regulation*. Raven Press, New York, NY, pp 63–81
- Sogawa K, Kikuchi Y, Imataka H, Fujii-Kuriyama Y (1993) Comparison of DNA-binding properties between BTEB and Sp1. *J Biochem* 114:605–609
- St-Arnaud R, Moir JM (1993) Wnt-1-inducing factor-1: a novel G/C box-binding transcription factor regulating the expression of Wnt-1 during neuroectodermal differentiation. *Mol Cell Biol* 13:1590–1598
- Swofford DL (1989) PAUP: phylogenetic analysis using parsimony, version 3.0b. Illinois Natural History Survey, Champaign, IL
- Verrijzer CP, Chen JL, Yokomori K, Tjian R (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* 81:1115–1125
- Wiley EO (1986) A study of the evolutionary relationships of *Fundulus* topminnows (Teleostei: Fundulidae). *Am Zool* 26:121–130
- Wiley SR, Kraus RJ, Mertz JE (1992) Functional binding of the “TATA” box binding component of transcription factor TFIID to the –30 region of TATA-less promoters. *Proc Natl Acad Sci USA* 89:5814–5818
- Zenzie-Gregory B, Khachi A, Garraway IP, Smale ST (1993a) Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. *Mol Cell Biol* 13:3841–3849
- Zenzie-Gregory B, Sheridan P, Jones KA, Smale ST (1993b) HIV-1 core promoter lacks a simple initiator element but contains a bipartite activator at the transcription start site. *J Biol Chem* 268:15823–15832