Letter to the Editor

5’ Genomic Structure of Human Sp3
Marjorie F. Oleksiak and Douglas L. Crawford
School of Biological Sciences, University of Missouri–Kansas City

Comparative sequence analysis among evolutionarily related genes can provide better insights into gene structures and functions compared with methods that rely solely on canonical rules (e.g., open reading frames, intron positions, or Kozak sequences). Presented here is a comparative sequence analysis of human Sp3 that provides the missing N-terminal portion.

Evolutionarily, the transcription factors Sp1 and Sp3 are related more closely to each other than to other members of the Sp family (Kolell and Crawford 2002), and numerous studies have examined the roles of Sp1 and Sp3 in transcriptional regulation. The Sp transcription factor family members interact with general transcriptional machinery, cell cycle regulators such as retinoblastoma protein, and other transcription factors and Sp1 belong to this family of transcription factors and compete for the same binding sites (Hagen et al. 1992; Kingsley and Winoto 1992). Much research has been published because of the widespread dependence on transcription on Sp transcription factors. For example, almost all mammalian TATA-less promoters have multiple Sp-binding sites that have been found to be required for maximal activity (Azizkhan et al. 1993).

Sp1 and Sp3 are well-studied transcription factors. But virtually all studies on Sp3 regulation of gene expression, its interactions with Sp1, or its repression of transcription have used a truncated form of Sp3, a form lacking the N-terminal domain. Thus, the N-terminal portion of the A domain is missing. This domain is critical for protein-protein interactions (Pugh and Tjian 1990; Seto, Lewis, and Shenk 1993; Gill et al. 1994; Udvardia, Templeton, and Horowitz 1995; Kennett, Moorefield, and Horowitz 2002). Sp3 and Sp1 belong to this family of transcription factors and compete for the same binding sites (Hagen et al. 1992; Kingsley and Winoto 1992). Much research has been published because of the widespread dependence on transcription on Sp transcription factors. For example, almost all mammalian TATA-less promoters have multiple Sp-binding sites that have been found to be required for maximal activity (Azizkhan et al. 1993).

Sp1 and Sp3 are related more closely to each other than to other members of the Sp family (Kolell and Crawford 2002), and numerous studies have examined the roles of Sp1 and Sp3 in transcriptional regulation. The Sp transcription factor family members interact with general transcriptional machinery, cell cycle regulators such as retinoblastoma protein, and other transcription factors (Pugh and Tjian 1990; Seto, Lewis, and Shenk 1993; Gill et al. 1994; Udvardia, Templeton, and Horowitz 1995; Kennett, Moorefield, and Horowitz 2002). Sp3 and Sp1 belong to this family of transcription factors and compete for the same binding sites (Hagen et al. 1992; Kingsley and Winoto 1992). Much research has been published because of the widespread dependence on transcription on Sp transcription factors. For example, almost all mammalian TATA-less promoters have multiple Sp-binding sites that have been found to be required for maximal activity (Azizkhan et al. 1993).

Sp1 and Sp3 are related more closely to each other than to other members of the Sp family (Kolell and Crawford 2002), and numerous studies have examined the roles of Sp1 and Sp3 in transcriptional regulation. The Sp transcription factor family members interact with general transcriptional machinery, cell cycle regulators such as retinoblastoma protein, and other transcription factors (Pugh and Tjian 1990; Seto, Lewis, and Shenk 1993; Gill et al. 1994; Udvardia, Templeton, and Horowitz 1995; Kennett, Moorefield, and Horowitz 2002). Sp3 and Sp1 belong to this family of transcription factors and compete for the same binding sites (Hagen et al. 1992; Kingsley and Winoto 1992). Much research has been published because of the widespread dependence on transcription on Sp transcription factors. For example, almost all mammalian TATA-less promoters have multiple Sp-binding sites that have been found to be required for maximal activity (Azizkhan et al. 1993).

Sp1 and Sp3 are related more closely to each other than to other members of the Sp family (Kolell and Crawford 2002), and numerous studies have examined the roles of Sp1 and Sp3 in transcriptional regulation. The Sp transcription factor family members interact with general transcriptional machinery, cell cycle regulators such as retinoblastoma protein, and other transcription factors (Pugh and Tjian 1990; Seto, Lewis, and Shenk 1993; Gill et al. 1994; Udvardia, Templeton, and Horowitz 1995; Kennett, Moorefield, and Horowitz 2002). Sp3 and Sp1 belong to this family of transcription factors and compete for the same binding sites (Hagen et al. 1992; Kingsley and Winoto 1992). Much research has been published because of the widespread dependence on transcription on Sp transcription factors. For example, almost all mammalian TATA-less promoters have multiple Sp-binding sites that have been found to be required for maximal activity (Azizkhan et al. 1993).

Address for correspondence and reprints: Douglas L. Crawford, School of Biological Sciences, University of Missouri–Kansas City, 5007 Rockhill Road, Kansas City, Missouri 64110. E-mail: crawforddo@umkc.edu.

Key words: human, Sp3, Sp1, genomic structure, transcription.

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
**Fig. 1.**—Sp3 and Sp1 amino-terminal alignments. A, Sp3 deduced protein alignment: human (hum), chicken (chk), rat, and mouse (mou). Boxes enclose similar amino acids. Shaded areas denote identical amino acids. Wedges indicate intron positions. The filled cross indicates the originally (Hagen et al. 1992) and the plain cross indicates the recently (Hernandez et al. 2002) defined 5' ends of human Sp3. B, Sp3, Sp1, and Sp4 deduced protein alignment. Marks are the same as in (A).
performed in Clustal (Higgins, Bleasby, and Fuchs et al. 1996). The third intron in Sp3 is not present in human Sp1 and Sp4 proteins, which are 80 and 82 kDa, respectively. The positions of the first and second introns are identical to these positions in mouse Sp4 (Supp et al. 1996). The third intron in Sp3 is not present in Sp4, but the fourth intron in Sp3 is in the same position as the third intron in Sp4 (data not shown). The first methionine (coding nucleotides 1–3) has an A in the position −3, which is sufficient for most ribosomes to select it irrespective of the rest of the surrounding sequence (Kozak 1991). The second methionine (coding nucleotides 37–39) has a G three base pairs upstream (position 43), suggesting that this AUG could be used to initiate translation.

Support that this is the correct identification of the 5' end of human Sp3 is based on (1) an initiation methionine codon, (2) the similarity to the genomic structure of Sp4, (3) sequence similarities among human and rat ESTs (fig. 1A), (4) similarities among the deduced amino acid sequences of chicken and mouse Sp3's, (5) similarities in protein length, and (6) regions of sequence similarity to other members of the Sp family (fig. 1B: amino acids 1–15, glycine-rich regions in exon 2, and exon 3 between amino acids 65 and 83).

Since they were first described, Sp3 and Sp1 have been shown to affect a wide variety of mammalian genes in transfection assays. Much can be learned about transcriptional control using transfection assays: they address binding site strength, protein-protein interactions, and transcriptional activation and repression. But although the full-length Sp1 sequence has been published for mammals, the full-length Sp3 sequence has not been identified. Incomplete Sp3 sequences were first published in 1992 (Hagen et al. 1992; Kingsley and Winoto 1992). Kingsley and Winoto (1992) identified putative non-AUG start sites, and Kennett, Udvardia, and Horowitz (1997) have identified two truncated Sp3 isoforms produced by internal translational initiation that, in contrast to full-length Sp3, act as potent transcriptional repressors. A recent publication suggests another non-AUG start site initiated in vitro (Hernandez et al. 2002). However, gel-shift assays with mammalian nuclear extracts suggest that endogenous Sp3 is similar in size to Sp1 (though this is uncertain due to known posttranslational modifications of Sp1), and Northern analyses show a transcript of 4.2 kb in a variety of mammalian cell lines (Kingsley and Winoto 1992). These data and the fact that the closely related Sp1 and Sp4 (Kolell and Crawford 2002) both initiate with a methionine and have roughly 70–90 more amino acids at their 5' ends suggest that full-length Sp3 transcripts also would code for a larger protein and initiate with a methionine. Thus, it is likely that no transient transfection studies using human Sp3 sequence to examine transcriptional regulation of human genes have used the full-length Sp3 sequence. In fact, they have been completed without the first three exons of Sp3. The 5' end of Sp1 is thought to be important in protein-protein interactions that affect transcription (Pascal and Tjian 1991; Murata et al. 1994). Because Sp3 belongs to the same family of proteins, the 5' end of Sp3 might have similar transcriptional importance.

Although no Sp3 sequence has been published with a 5' methionine, the Sp3 gene has been mapped to chromosome 2 (Kalff-Suske et al. 1996). Blast searches with putative intron and exon sequences identified by genome walking from known Sp3 sequence match H. sapiens BAC clone RP11-394I13 from chromosome 2. These are contiguous with published Sp3 sequences. Blast searches also match unidentiﬁed rat and human ESTs in the database and a chicken Sp3 (accession number AJ317961). All these code for similar amino-terminal ends containing methionines (fig. 1A). The sequence also is similar to an incomplete mouse Sp3 (Supp et al. 1996). Mouse Sp3 sequence lacks approximately fourteen 5' amino acids and the entire third exon. But it does have the fourth exon, suggesting that mouse Sp3 might be alternatively spliced. These sequences are similar to Sp1 and Sp4 proteins (fig. 1B) with respect to length and protein sequences.

In summary, the identification of this sequence as the full-length Sp3 sequence is supported for the following reasons. First, genome walking and database searching has identiﬁed contiguous sequence upstream of the published Sp3 sequence. Exons identiﬁed in this sequence code for a protein more similar in size to Sp1, as is expected from gel-shift assays with endogenous Sp1 and Sp3 proteins. This protein has a calculated molecular weight of 82 kDa, similar to the calculated molecular weight of human Sp1 (80 kDa), and the deduced transcript (3.9 kb) is more similar in size to the Sp3 transcript (4.2 kb), based on Northern analyses. Sequence similarities with amino termini for Sp1 and Sp4 sequences also suggest that exons identiﬁed in this sequence code for the amino terminus of Sp3. In particular, this sequence has an initiating methionine with a favorable Kozak sequence. This sequence also has an intron-exon structure, similar to that of Sp4. Notably,
the first exons in both proteins are only seven nucleotides long, coding for only two amino acids. Finally, matches in the database to chicken Sp3 and unidentified human and rat ESTs suggest that this transcript is expressed in vivo.

The intron and exon sequence for the 5' end of human Sp3 is shown in the supplementary material on the SMBE website (www.molbiolevol.org). This sequence has been deposited in GenBank and assigned the accession number AF494280.

Acknowledgments

This research benefited from a discussion with Kevin Kolell and was supported by NSF BioInformatics postdoctoral fellowship 0074520 to M.F.O. and NSF IBN grant # 9986602 to D.L.C.

LITERATURE CITED


CLAUDIA KAPPEN, reviewing editor

Accepted July 17, 2002