

Transcription factor Sp2: Molecular characterization and generation of Sp2 gene targeted mice

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften
(**Dr. rer. nat.**)

dem Fachbereich Biologie
der Philipps Universität Marburg
vorgelegt von

Frank Baur
aus Kirchheim/Teck

Marburg, 2005

Vom Fachbereich Biologie
der Philipps-Universität Marburg
als Dissertation angenommen am:
15.08.2005

Erstgutachter: Prof. Dr. Renate Renkawitz-Pohl
Zweitgutachter: Prof. Dr. Guntram Suske
Tag der mündlichen Prüfung am:
15.09.2005

Contents

Zusammenfassung	I
Summary	III
1. Introduction	1
1.1 Regulation of gene expression	1
1.2 Transcription factors	3
1.3 The Sp/XKLF super-family of transcription factors	4
1.4 The Sp family of transcription factors: protein structure and function	8
1.5 Generation of transgenic mice by “gene targeting”	15
1.6 Thesis aims	19
2. Materials and methods	21
2.1 Materials	21
2.1.1 Laboratory materials and devices	21
2.1.2 Chemicals	21
2.1.3 General solutions	21
2.1.4 Culture media	22
2.1.4.1 Media to culture bacteria	22
2.1.4.2 Media to culture eukaryotic cells	22
2.1.5 Restriction enzymes and DNA-modifying enzymes	23
2.1.6 Antibiotics	23
2.1.7 Antibodies	23
2.1.8 Radioactive substances	24
2.1.9 Oligonucleotides	24
2.1.9.1 Oligonucleotides to generate Sp2 deletion mutants	24
2.1.9.2 Oligonucleotides for Electrophoretic Mobility Shift Assays	25
2.1.9.3 Oligonucleotides for RT-PCR, Southern Blot probes and cosmid library screen (RZPD)	25
2.1.9.4 Oligonucleotides for loxP site generation and PCR amplifi- cation of Sp2 genomic fragments	26
2.1.9.5 Oligonucleotides for mouse genotyping	26
2.1.10 Plasmids	27
2.1.10.1 Previously described plasmids	27
2.1.10.2 Plasmids generated during this thesis work	29
2.1.10.2.1 Bacterial expression plasmids	29

2.1.10.2.2	Drosophila expression plasmids	30
2.1.10.2.3	Mammalian expression plasmids	31
2.1.10.2.4	Knockout construct and pre-constructs	32
2.1.11	Cosmids	33
2.1.12	Bacterial strains	34
2.1.13	Cell lines	34
2.1.14	Mice	35
2.1.15	Rabbits	35
2.2	Methods	37
2.2.1	Radiation protection and biological safety	37
2.2.2	Molecular biological methods	37
2.2.2.1	RNA experiments	37
2.2.2.1.1	RNA isolation and purification	37
2.2.2.1.2	RNA quantification	38
2.2.2.1.3	Reverse transcriptase polymerase chain reaction	38
2.2.2.1.4	Northern Blot analyses	42
2.2.2.2	DNA experiments	43
2.2.2.2.1	DNA isolation and purification	44
2.2.2.2.2	DNA quantification	45
2.2.2.2.3	Polymerase chain reaction	45
2.2.2.2.4	Hybridization and purification of single-stranded oligo-nucleotides	48
2.2.2.2.5	Radioactive labeling of DNA	49
2.2.2.2.6	Southern Blot analyses	51
2.2.3	Biochemical methods	52
2.2.3.1	Recombinant Sp2 protein expression and purification from bacterial inclusion bodies	52
2.2.3.2	Protein extraction from cells (nuclear extract preparation)	54
2.2.3.3	Determination of protein concentrations	54
2.2.3.4	SDS Polyacrylamide Gel Electrophoresis	55
2.2.3.5	Western Blot analyses	56
2.2.3.6	Rabbit immunization and generation of antiserum	57
2.2.3.7	Electrophoretic Mobility Shift Assay	58
2.2.4	Cellbiological methods and animal works	59
2.2.4.1	Cell handling	59

2.2.4.2	Cell immunostaining	60
2.2.4.3	Cell transfections	60
2.2.4.4	Transactivation assays	62
2.2.4.5	Generation of Sp2-targeted mice and mouse handling	63
2.2.4.6	Body size and weight measurements and fixation of mouse embryos	65
2.2.4.7	X-Gal and BluoGal staining of mouse embryos	65
3.	Results	67
3.1	Molecular characterization of the transcription factor Sp2	67
3.1.1	Generation of rabbit polyclonal Sp2-specific antibodies	67
3.1.1.1	Expression of Sp2 protein in <i>E. coli</i> BL21DE3 bacteria	67
3.1.1.2	Isolation of inclusion bodies containing recombinant Sp2 protein	68
3.1.1.3	Immunization of rabbits	68
3.1.1.4	Characterization of Sp2-specific rabbit antisera	69
3.1.2	Endogenous Sp2 protein expression	70
3.1.2.1	Endogenous Sp2 protein expression in various cell lines	70
3.1.2.2	Endogenous Sp2 protein expression in adult mouse tissues	71
3.1.3	Subcellular localization of Sp2 protein	72
3.1.4	Transactivation properties of Sp2 protein overexpressed in SL2 cells	73
3.1.5	DNA binding capacity of Sp2 protein	74
3.1.5.1	GC box binding capacity of full-length Sp2 protein overexpressed in <i>Drosophila</i> SL2 cells	75
3.1.5.2	GC box binding capacity of endogenous Sp2 protein (MEF, HEK-293 and HeLa cells)	76
3.1.5.3	Binding capacity of SL2-overexpressed full-length Sp2 protein towards GT and CT boxes	78
3.1.5.4	Binding capacity of SL2-overexpressed full-length Sp2 protein towards GC box variants	79
3.1.5.5	DNA binding capacity of the Sp2 DNA-binding domain overexpressed in SL2 cells	80
3.1.5.6	DNA binding capacity of SL2-overexpressed N-terminal Sp2 protein deletion mutants	81

3.1.5.7	DNA binding capacity of Gal4-fused C-terminal Sp2 deletion mutants overexpressed in HEK-293 cells	83
3.1.6	Transactivation properties of Sp2 deletion mutants displaying DNA binding capacity	85
3.1.6.1	Transactivation properties of N-terminal Sp2 deletion mutants overexpressed in SL2 cells	85
3.1.6.2	Transactivation properties of Gal4-Sp2 deletion mutants overexpressed in HEK-293 cells	87
3.2	Generation of Sp2 gene targeted mice	91
3.2.1	Sp2 gene structure	91
3.2.2	Strategy to target the Sp2 gene in the mouse	94
3.2.3	Screening for Sp2 genomic DNA	97
3.2.4	Generation of the Sp2 knockout construct	98
3.2.5	Functional analysis of Cre-driven loxP site recombinase capacity	100
3.2.6	Sp2 gene targeting in mouse embryonic stem cells	102
3.2.7	Genotyping of constitutive Sp2 lzn/lzn knockout mice	103
3.2.8	Sp2 expression in targeted Sp2 lzn/lzn knockout mice	104
3.2.8.1	Detection of a Sp2-lzn fusion mRNA in the targeted mice	104
3.2.8.2	Absence of exon 5-8 in the targeted mice	105
3.2.8.3	Endogenous Sp2 expression in mouse embryos	106
3.2.9	Preliminary characterization of targeted Sp2 lzn/lzn mice	108
3.2.9.1	Post- and pre-natal viability of targeted Sp2 lzn/lzn mice	108
3.2.9.2	Reduced growth in day E18.5 Sp2 lzn/lzn embryos	109
4.	Discussion	111
4.1	Structure comparisons between Sp2 and the glutamine-rich family members Sp1, Sp3 and Sp4	111
4.2	Regulation of Sp2 DNA binding capacity and transactivation properties	113
4.3	Sp2 is essential for normal mouse development	120
5.	References	123
6.	Appendix	135
6.1	Abbreviation index	135
6.2	Mouse Sp2 genomic DNA and cDNA sequences	137
6.3	Acknowledgements	145
6.4	Erklärung	146

Zusammenfassung

Die Familie der Sp-Transkriptionsfaktoren charakterisiert sich durch ihre DNA-Bindungsdomäne, bestehend aus drei C-terminalen Zinkfingern vom Typus C2H2. Als Folge dieses hochkonservierten DNA-Bindungsmotivs erkennen die einzelnen Mitglieder der Sp-Familie jeweils mit gleicher Spezifität und Affinität sogenannte GC- (GGGGCGGGG) und GT- (GGTGTGGGG) Boxen. GC- und GT-Boxen sind von großer Bedeutung für die Expressionsregulation verschiedenster ubiquitärer, gewebespezifischer und viraler Gene. Bis jetzt wurden neun Mitglieder der Sp-Familie identifiziert (Sp1 bis Sp9). Zusätzlich zur charakteristischen DNA-Bindungsdomäne besitzen die Transkriptionsfaktoren Sp1 bis Sp4 weitere strukturelle Gemeinsamkeiten, wie z.B. zwei glutaminreiche Transaktivierungsdomänen oder zwei serin-/threoninreiche Regionen. Molekulare sowie funktionelle Eigenschaften sind für die Faktoren Sp1, Sp3 und Sp4 beschrieben. Entsprechende Maus-Knockouts belegen ihre vielfältige Funktion und essentielle Bedeutung bei der Säugerentwicklung. Seit der Klonierung von Sp2, dem am wenigsten konservierten Mitglied der aus Sp1-4 bestehenden glutaminreichen Unterfamilie, wurden keine weiteren Daten hinsichtlich seiner Funktion *in vivo* oder *in vitro* publiziert. Aus diesem Grunde war das Ziel der vorliegenden Arbeit, die Untersuchung der Sp2-Funktion mittels zwei verschiedener Ansätze: einer funktionellen molekularen Charakterisierung, welche unter anderem Expressions-, Transaktivierungs- und DNA-Bindungsstudien beinhaltete, sowie der Herstellung von Sp2-Knockout-Mäusen.

Zur Untersuchung der Sp2-Funktion auf molekularer Ebene wurden polyklonale, Sp2-spezifische Antikörper hergestellt und für Expressionsstudien eingesetzt. Es zeigte sich, daß Sp2 ausschließlich im Zellkern lokalisiert ist und in allen getesteten Zelllinien und Mausgeweben exprimiert wird, wenn auch in unterschiedlichen Mengen. Dies spricht für eine weitläufige bis ubiquitäre Expression in der Maus. In einem zweiten Schritt wurde die Fähigkeit von Sp2 analysiert, Promotoren zu aktivieren, die GC- bzw. GT-Boxen als Regulationselemente beinhalten. Hierzu wurden Reporterassays mit Wildtyp-Sp2-Protein sowie verschiedenen Sp2-Deletionsmutanten durchgeführt. Im Gegensatz zum Transkriptionsfaktor Sp1, der einen starken Aktivator darstellt, aktivierte keines der untersuchten Sp2-Varianten die Expression der Reportergene. Auch wenn Sp2-Fragmente an eine heterologe DNA-Bindungsdomäne fusioniert waren, konnte keine Aktivierung festgestellt werden.

Zusätzlich wurde mittels Gelretardationsexperimenten die Fähigkeit und Spezifität von Sp2, DNA zu binden, untersucht. Es zeigte sich, daß Wildtyp-Sp2-Protein nicht in der Lage war, an DNA zu binden, weder an GC-Boxen („klassische“ Sp1-Bindungsstelle) noch an GC-Box-Varianten oder andere DNA-Bindungssequenzen, wie GT- oder CT-Boxen. Wurden jedoch die N-terminalen Aminosäuren 1-179 deletiert, erfolgte die Bindung an die DNA. Diese Befunde sprechen dafür, daß der Prozeß der Sp2-DNA-Interaktion *in vivo* reguliert ist. Zur Ermittlung der physiologischen Funktion von Sp2, wurden Sp2-Knockout-Mäuse hergestellt. Diese Mäuse sind nicht lebensfähig; sie sterben kurz vor bzw. nach der Geburt. Während Sp2-mutierte Embryonen bis zum Tag E12,5 keine erkennbaren Abnormalitäten aufweisen, sind die Körpergröße und das -gewicht von Embryonen am Tag E18,5 im Vergleich zum Wildtyp deutlich reduziert, wenn auch mit großer Varianz. Diese Ergebnisse verdeutlichen, daß der Transkriptionsfaktor Sp2 essentiell für eine normale Mausentwicklung ist. In welche Differenzierungsprozesse er im Detail involviert ist, müssen weitere Untersuchungen zeigen.

Summary

The Sp family of transcription factors is characterised by its DNA-binding domain, an array of three conserved C2H2 zinc fingers. As a consequence of the conserved DNA binding motif, Sp members recognize GC (GGGGCGGGG) and GT (GGTGTGGGG) boxes with similar specificity and affinity. GC and GT boxes are important for the expression of many different ubiquitous as well as tissue-specific cellular and viral genes. To date, nine members of the Sp family (Sp1 - Sp9) have been identified. In addition to their DNA-binding domain, Sp1 - Sp4 also share other structural features like two glutamine-rich transactivation domains and two serine/threonine-rich regions. Molecular and functional properties have been described for Sp1, Sp3 and Sp4. Mouse deletion mutants, which have been generated for these factors, demonstrate their manifold function and essential importance for mammalian development. Since the cloning of Sp2, which is the less conserved factor among Sp1 - Sp4, no reports about its function, neither *in vitro* nor *in vivo* have been published. Therefore, the aim of this thesis work was to unravel Sp2 function by two parallel approaches: a functional molecular characterization (including expression, transactivation and DNA binding studies) and the generation of Sp2 gene targeted mice.

To study the Sp2 protein at the molecular level, Sp2-specific rabbit polyclonal antibodies were generated. Sp2 protein, which is exclusively localized to the nucleus, was detected in all analyzed cell lines and adult mouse tissues, although in different amounts. This favours at least a widely expression of the transcription factor Sp2. To explore Sp2 transactivation properties, reporter assays were performed with full-length Sp2 protein as well as various Sp2 deletion mutants using different GC- and GT-box-containing promoters. Unlike transcription factor Sp1, which is a strong activator, Sp2 proteins did not activate reporter gene expression. Also, when fusing Sp2 deletions to a heterologous Gal4 DNA binding domain, no activation was detectable. In addition, the DNA binding capacity and specificity of full-length Sp2 protein and a series of Sp2 deletion mutants were investigated by Electrophoretic Mobility Shift Assays. Full-length Sp2 protein was not able to bind to DNA, neither to GC boxes (the "classical" Sp1 binding site) and GC box variants, nor to other DNA binding sequences like GT and CT boxes. However, when deleting the N-terminal amino acids 1-179, GC box binding was possible. These results suggest that the

DNA binding activity is regulated *in vivo*. To unravel the physiological function of transcription factor Sp2, targeted mice were generated. These mice are not viable; they die shortly before or after birth. Whereas Sp2-targeted embryos develop normal until day E12.5, day E18.5 embryos are characterized by a strongly reduced body size and weight, however with strong variations. These results demonstrate a fundamental role of the transcription factor Sp2 for normal mouse development.

1. Introduction

1.1 Regulation of gene expression

Multicellular organisms like mammals are composed of various cell types with different functions. Although displaying different functions, nearly all nuclear cells contain the same genetic information. Apart from so-called house-keeping genes, which are expressed in all cells, the expression of genes is cell-type-dependent. But also cell development or environmental conditions requiring different gene products in the cell influence gene expression. This differential gene expression is the basis for the formation of complex, multicellular organisms with a great diversity of cell types, each of them characterized by a special set of proteins.

Differential gene expression as well as the formation of a functional protein can be regulated on different levels, like e.g. transcription, mRNA processing, mRNA transport and stability but also translation and stability or activity of the translated protein. The regulation on the transcriptional level hereby plays a crucial role.

In eukaryotic cells, gene transcription is driven by three RNA polymerases, RNA polymerase I, II and III (Roeder and Rutter, 1969). Protein-coding genes are transcribed into mRNA by RNA polymerase II. Transcription of DNA into mRNA by RNA polymerase II is one of the most highly regulated processes in the cell. This regulation depends on a complex molecular machinery (Fig. 1.1) consisting of numerous transcription factors (Lemon and Tjian, 2000).

Eukaryotic promoters generally are composed of a core promoter, a distal promoter region and several enhancer or silencer elements. Core promoters often contain the so-called TATA-box (25-30 nucleotides upstream of the transcription start point), a pyrimidine-rich initiator sequence (Inr) or a so-called downstream core promoter element (DPE; approximately 30 nucleotides downstream of the transcription start point) (Krajewska, 1992; Mitchell and Tjian, 1989). Enhancer and silencer elements, on the other hand, are DNA regions in great distances upstream or downstream to the transcription start point. Transcription factors bind to these regions and can enhance or silence gene transcription (Brand et al., 1985; Dynan 1989; Levine and Manley, 1989; Voss and Pongs, 1986).

To initiate transcription, transcription factors bind to the distal promoter and enhancer elements and by this enable the recruitment of RNA polymerase II to the core promoter region (Orphanides et al., 1996; Roeder, 1996). The first step in transcription initiation is the remodeling of chromatin giving access of the pre-initiation complex to the promoter (Fig. 1.1). The pre-initiation complex is assembled stepwise starting with the attachment of the multi-protein complex TFIID to the TATA-box via its subunit TBP (TATA-box-binding protein; Weinzierl et al., 1993) together with the attachment to the DPE element (*Drosophila*) via a TBP-associated subunit (TAF), followed by the association of further general transcription factors like TFIIA, B, F, E and H as well as the RNA polymerase II. Assembly of the pre-initiation complex is sufficient for a basal promoter activity *in vitro* (Lewin, 1990; Roeder, 1991). *In vivo*, however, other factors like specific transcription factors or different co-regulators are necessary (Faisst and Meyer, 1992; Roeder, 1991; Wingender, 1988).

Many promoters, enhancers and silencers contain characteristic DNA sequence motifs, like e.g. GC or CAAT boxes. These boxes are bound by specific transcription factors like Sp factors and can either be located close to the transcription start point or as enhancer or silencer in great distance to it. Interaction of these factors with factors of the pre-initiation complex can e.g. result in conformational changes and initiation of transcription (Kornberg, 1996; Roeder, 1996).

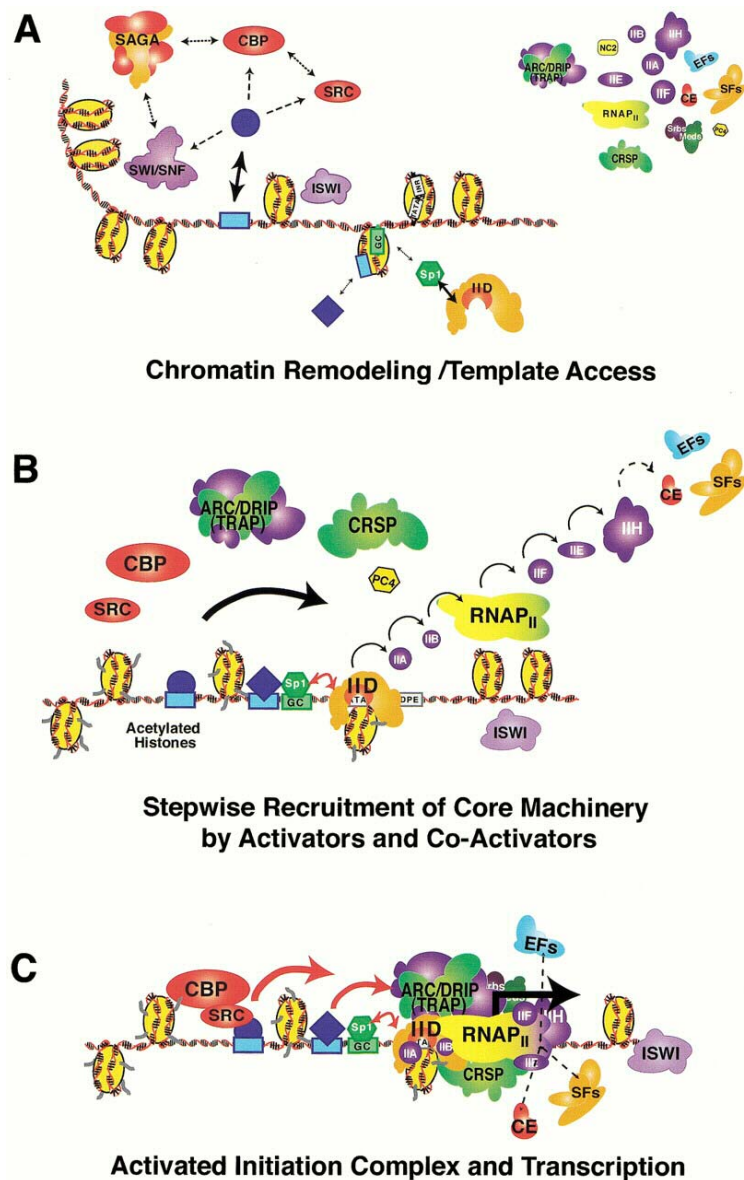


Fig. 1.1. Transcription initiation (Lemon and Tjian, 2000). **A.** Chromatin remodeling to make the promoter accessible for the pre-initiation complex formation. **B.** Stepwise assembly of transcription factors and co-regulators at the core promoter. **C.** Transcription initiation complex.

1.2 Transcription factors

Transcription factors can be divided into general or basal transcription factors (Roeder, 1991; Goodrich and Tjian, 1994) and specific transcription factors (Lewin, 1990; Krajewska, 1992). General or basal transcription factors are associated with the pre-initiation complex at the core promoter as described in 1.1. Specific transcription factors bind to promoter and enhancer or silencer regions, thus directly or indirectly regulating transcription.

According to their regulatory function, specific transcription factors are composed of several functional domains, like e.g. a DNA-binding domain, which enables the factors to sequence-specifically bind to regulatory elements on the DNA, as well as one or more transactivation domain(s), essential for transcriptional activation (Kadonaga, 2004; Mitchell and Tjian, 1989; Ptashne, 1988;). The sequence specificity of DNA-binding domains is achieved by a structural diversity of DNA binding motifs. Such binding motifs are e.g. leucine-zipper, helix-turn-helix, helix-loop-helix or zinc finger motifs and enable the specific recognition of a broad variety of DNA sequences, like e.g. CAAT or GC boxes (Landschulz et al., 1988; Krajewska, 1992; Kadonaga et al., 1987; Pabo and Sauer, 1992). High diversity also exists for transactivation domains, which can be glutamine-rich as in transcription factor Sp1 (Courey and Tjian, 1988), proline-rich as in the factors CTF/NF1 and Jun (Mermod et al., 1989; Struhl, 1988) or rich in acidic amino acids as in the yeast Gal4 transcription factors (Ma and Ptashne, 1987).

In addition, further functional domains, like inhibitory, dimerization or ligand-binding domains (e.g., Kadonaga, 2004) regulate transcription by influencing the transactivation or DNA binding capability of the transcription factor. DNA binding capacity e.g. can be regulated through protein-protein interactions (e.g. Vallian et al., 1998) or posttranslational protein modifications, like phosphorylation or glycosylation (e.g. Rohlff et al., 1997; Armstrong et al., 1997; Han and Kudlow, 1997; Yang et al., 2001).

1.3 The Sp/XKLF super-family of transcription factors

Many different prokaryotic and eukaryotic proteins use zinc-coordinated motifs to bind to DNA. One common type of these so-called zinc fingers consists of two beta sheets and one alpha helix that contain two cysteine and two histidine residues that contact a zinc atom. These C₂H₂ zinc fingers are often found in clusters that allow each of their alpha helices to tightly interact with the major groove of the double-stranded DNA helix (Pavletich and Pabo, 1991). The amino acid composition of the zinc fingers determines their DNA binding specificity and by using them in different arrangements zinc finger proteins can recognize the specific sequences of

nucleotides to which they bind, thus ensuring a highly specific transcriptional regulation.

One particular combination of three conserved C-terminal C2H2 zinc fingers forms the DNA binding motif of the still expanding Sp/XKLF super-family of transcription factors (reviewed e.g. in Bouwman and Philipsen, 2002; Suske et al., 2005; see also Fig. 1.3.1). Sp represents the Sp family of transcription factors, consisting to date of nine family members, Sp1-9, which are described in more detail in chapter 1.4. The Sp family is named after “specificity protein” or “sephacryl and phosphocellulose” columns originally used to purify family founder Sp1 (Kadonaga et al., 1987)). The XKLF family, on the other hand, consists of the numerous so-called Krüppel-like factors (reviewed in Bouwman and Philipsen, 2002; Suske et al., 2005), named after the *Drosophila* segmentation gene Krüppel displaying a similar zinc finger arrangement (Schuh et al., 1986). X hereby represents the main location of expression, like erythrocytes for EKLF (erythroid Krüppel-like factor; Miller and Bieker, 1993) or lung in the case of LKLF (lung Krüppel-like factor; Anderson et al., 1995).

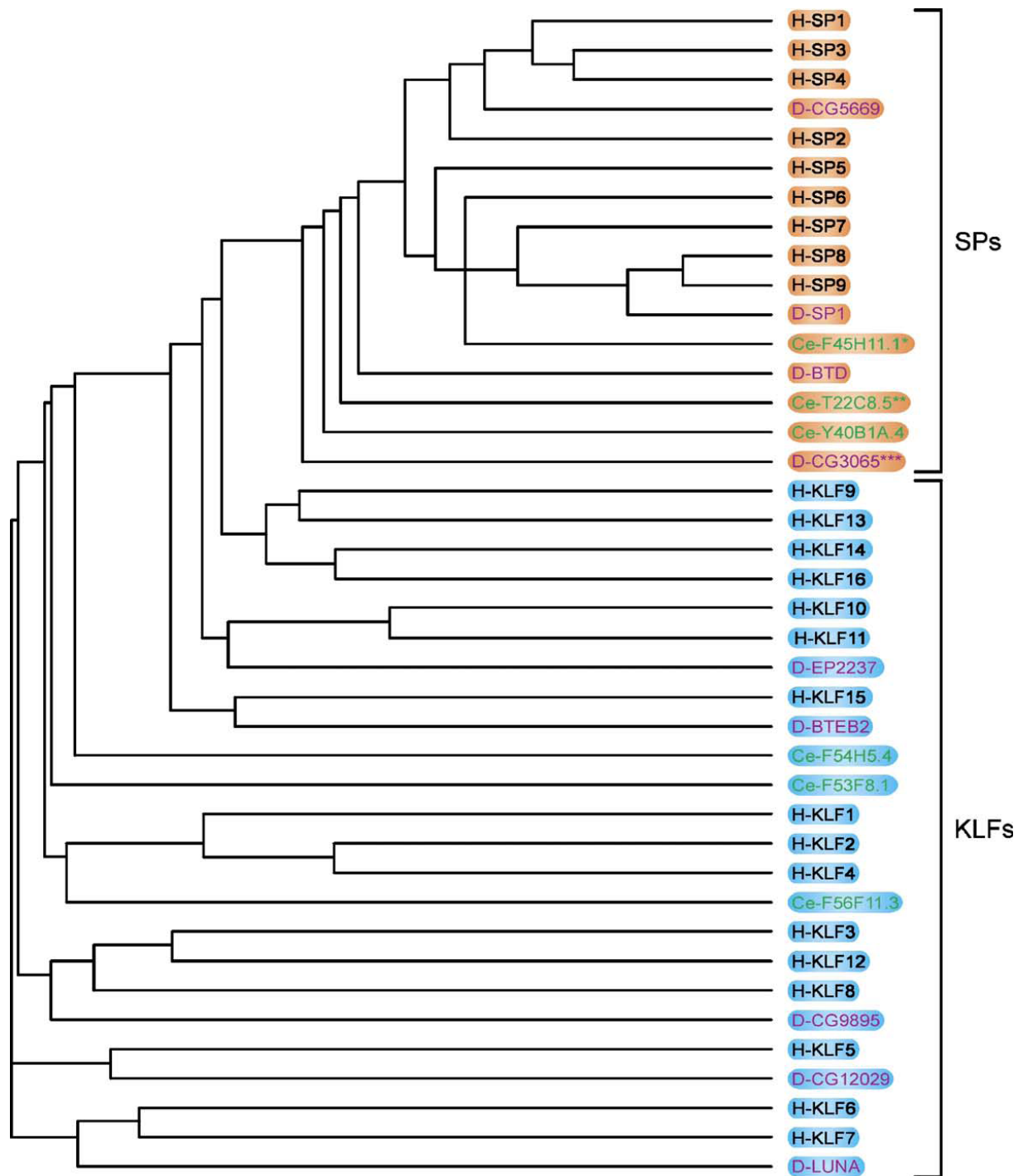


Fig. 1.3.1. Sp/XKLF super-family of transcription factors (Suske et al., 2005). Relationships between the Sp factors and KLFs of human, *Drosophila*, and *C. elegans*. Of each factor, the 110-aa domain containing the Btd/zinc finger motifs was used for the multiple alignment with ClustalW (<http://www.ebi.ac.uk/clustalw/>). This alignment was used to construct the cladogram. *The current sequence of *C. elegans* F45H11.1 contains only the Btd motif and the first finger. **One amino acid was removed from finger 3 of Ce-T22C8.5 (HXXXH instead of HXXXXH), because the extra amino acid is not handled appropriately in the multiple alignment generated with ClustalW. ***The sequence of D-CG3065 was deduced from the *Drosophila* genome sequence; the current annotation does not contain the complete Btd/zinc finger motif.

The zinc fingers of the Sp/XKLF super-family are structurally related to those of the transcription factor ZIF268. Therefore Sp/XKLF factors are likely to contact the DNA in the same fashion as has been determined for this protein. Following a so-called zinc finger code, each zinc finger of ZIF268 contributes with (at least) three DNA-contacting amino acids to its DNA binding properties (Jamieson et al., 2003; Fairall et al., 1993; Kriwacki et al., 1992; Kuwahara et al., 1993; Narayan et al., 1997; Pavletich and Pabo, 1991). According to that, Sp family founder Sp1 is thought to contact the DNA with the amino acids KHA in the first, RER in the second and RHK in the third zinc finger as depicted in Fig. 1.3.2 (Dyana and Tjian, 1983; Philipsen and Suske, 1999).

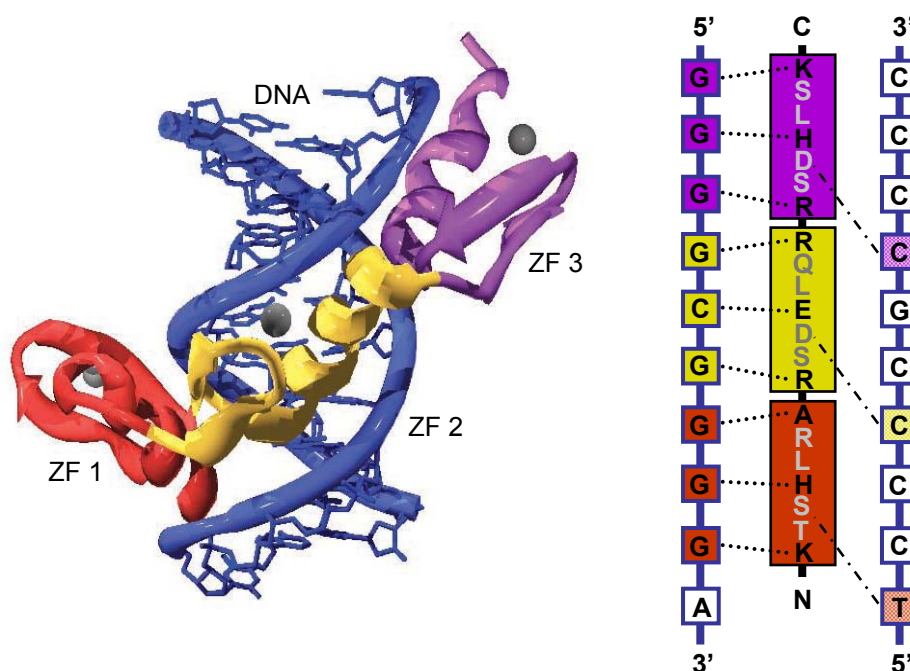


Fig. 1.3.2. Scheme of the Sp1 zinc fingers interacting with a classical GC box (modified after Jamieson et al., 2003). The three individual fingers consist of each one alpha helix and two beta sheets and are connected by a zinc ion (depicted in grey). At least three amino acids in each alpha helix (the critical residues are depicted in black) contact each one nucleotide on the DNA (dotted lines), as indicated on the right. Sp1 zinc fingers (as well as the bases which they contact) are depicted as follows: finger 1 in red, finger 2 in yellow and finger 3 in violet. DNA is depicted in blue.

As a consequence of the conserved DNA binding motif, Sp/XKLF members all recognize the same GC (GGGGCGGGG) and GT/CACC (GGTGTGGGG) boxes albeit with different affinities due to substitutions of critical amino acids in the first (H→L in Sp2) or in the third (L→K in e.g. BKLF, EKLF, UKLF) finger of some of the factors (Bouwman and Philipsen, 2002; Gidoni et al., 1985; Giglioni et al., 1989;

Kingsley and Winoto, 1992; Letovsky and Dynan, 1989). GC (commonly referred to as “Sp1-binding site”) and GT boxes are frequently found in promoters and enhancers/ silencers of many different ubiquitous (e.g. house-keeping genes) as well as specifically regulated cellular and viral genes. In addition, these motifs are required for the maintenance of the methylation-free status of CpG islands (Brandeis et al., 1994; Macleod et al., 1994).

A third subgroup belonging to the Sp/XKLF super-family comprises the transcription factors BTEB1/KLF9, BTEB4/KLF16 and BTEB5/KLF14 (basic transcription element binding protein 1; Imataka et al., 1992), RFLAT-1/KLF13 (RANTES factor of late activated T-lymphocytes 1) as well as TIEG1/KLF10 and TIEG2/KLF11 (TGF β -inducible early gene 1 and 2; Cook et al., 1998; Subramaniam et al., 1995). Based on their zinc finger structure, these factors also primarily bind to the classical GC boxes (Bouwman and Philipsen, 2002; Philipsen and Suske, 1999; Suske et al., 2005).

Whereas the DNA binding domain is highly conserved among the Sp/XKLF super-family members, also eminent structural differences can be observed like the presence or absence of inhibitory domains or various transactivation domain structures. Together with differences in posttranslational protein modifications, protein-protein interactions, expression patterns, etc., this contributes to enhance specificity of transcriptional regulation (Bieker and Southwood, 1995; Gillemans et al., 1998; Bouwman and Philipsen, 2002; Philipsen and Suske, 1999; Suske et al., 2005).

1.4 The Sp family of transcription factors: protein structure and function

To date, the Sp family of transcription factors consists of nine mammalian members, Sp1-9 (Fig. 1.4.1; e.g. Bouwman and Philipsen, 2002; Suske et al. 2005). As described in 1.3, Sp transcription factors bind to GC and GT/CACC boxes found in a variety of promoters and enhancers/ silencers through three characteristic zinc fingers of the C2H2 type. The zinc fingers are located at the C terminus of the proteins and are formed by a stretch of 81 highly conserved amino acids (Fig. 1.4.2).

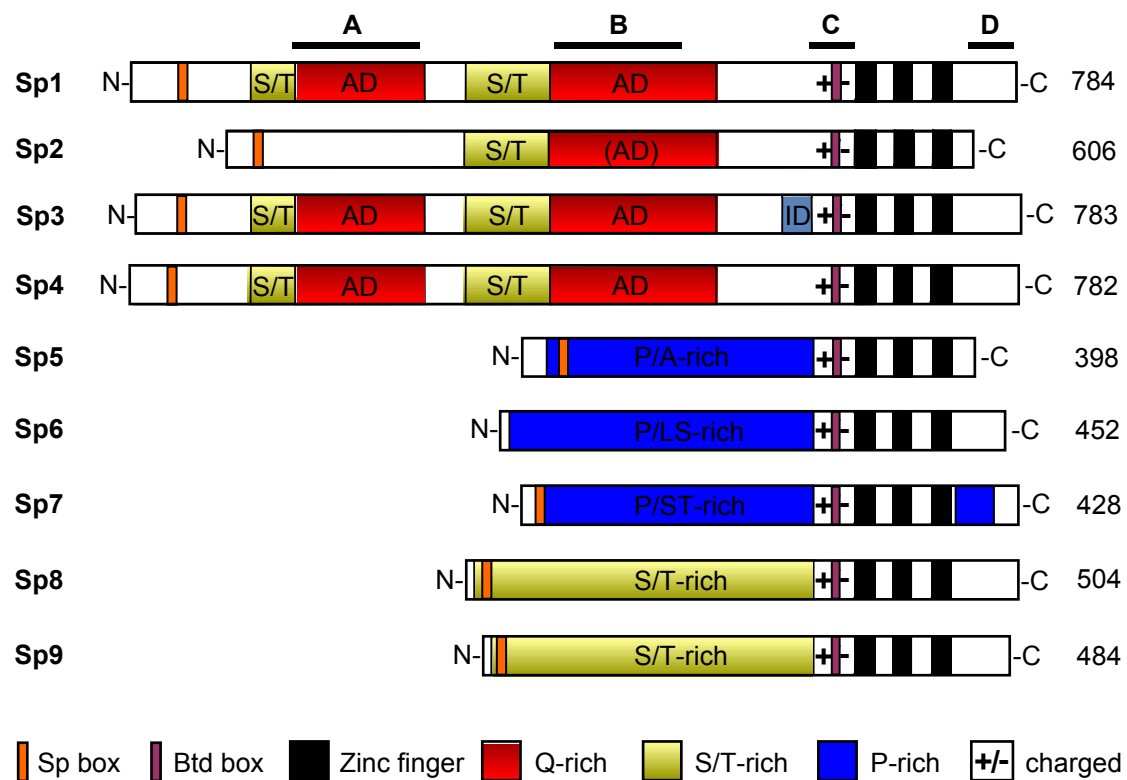


Fig. 1.4.1. Transcription factors Sp1-9 (modified after Bouwman and Philipson, 2002 and Suske et al., 2005). Structural motifs like Sp and Btd boxes, zinc finger, glutamine-rich, serine/threonine-rich, proline-rich and highly charged regions, as well as transactivation (AD) and inhibitory (ID) domains are indicated following the colour code given above. A, B, C, and D modules of Sp1 (Courey and Tjian, 1988) are marked with black bars. On the right: lengths in amino acids according to accession numbers NM_013672.1 (Sp1), NM_030220.2 (Sp2), BC079874.1 (Sp3), NM_009239.1 (Sp4), NM_022435.2 (Sp5), XP_064386 (Sp6/KLF14), NM_130458.1 (Sp7), NM_177082.3 (Sp8), AY591908 (Sp9).

In addition to their DNA-binding region, Sp transcription factors also share further structural features, like the Sp box and the Buttonhead box, of which as yet the function is unclear. The Sp box is located at the N terminus of the proteins (Harrison et al., 2000) and is characterized by the highly conserved sequence SPLALLAATCSR/KI (Bouwman and Philipson, 2002). It contains a potential endoproteolytic cleavage site and is situated close to a region at the N terminus of Sp1 that targets proteasome-dependent degradation *in vitro* (Su et al., 1999). Although not required to direct cleavage, the fact that the Sp box is highly conserved indicates that it has a putative function in the regulation of Sp factor proteolysis. Another possible role for the Sp box may lie in the control of the transactivation capacity through interaction with a putative repressor (Murata et al., 1994).

The Buttonhead box (Btd) is situated at the C terminus of the proteins, directly N-terminally to the zinc finger domain. It consists of 11 conserved amino acids and was originally described in the *Drosophila* Sp1 homologue Buttonhead (Btd; Wimmer et al., 1993). A deletion of the highly charged C domain (see Fig., 1.4.2) comprising the Btd box results in a reduced transactivation potential of Sp1 *in vitro* (Courey and Tjian, 1988). Furthermore, the Btd box appears to be involved in synergistic activation by Sp1 or Sp3 with sterol regulatory element-binding proteins (SREBPs; Athanikar et al., 1997).

Although the functions of the Sp and Btd boxes are not clear at the moment, their absence in the XKLF family confirms the relationship between the Sp transcription factors. Besides these structural similarities, the close relationship between these factors is also demonstrated in human by their co-localization with the four homeobox gene clusters (HOX) on chromosome 12.q13.13 (Sp1 and 7 / HOX C), 17q21.31/32 (Sp2 and 6 / HOX B), 2q31.1 (Sp3, 5 and 9 / HOX D), 7p21.2 (Sp4 and 8 / HOX A) (Bouwman and Philipsen, 2002; Kalff-Suske et al., 1995 and 1996; Matera and Ward, 1993; Scohy et al., 1998; Suske et al. 2005).

Apart from the Sp and Btd boxes, the N-terminal regions of Sp5-9 are completely different from those of Sp1-4 and more closely related to each other (Bouwman and Philipsen, 2002; Suske et al., 2005). Since this thesis is focused on transcription factor Sp2, the factors Sp5-9 will not be discussed in the following, especially as they have not been studied yet in detail.

The protein structure of the transcription factors Sp1-4 is characterized by several domains located N-terminally to the highly conserved zinc finger region (Fig. 1.4.2; Bouwman and Philipsen, 2002; Philipsen and Suske, 1999). These are two glutamine-rich domains (A and B), two serine/threonine-rich regions and a region of highly charged amino acids (domain C) directly N-terminally to the zinc finger domain. The existence of the first transactivation domain and serine/threonine region in Sp2 was unclear when starting this thesis. In addition, the D domain is absent in Sp2. The glutamine-rich domains of Sp1, Sp3 and Sp4 harbour the transactivating function of these factors, the serine/threonine regions are possible targets of posttranslational modifications like phosphorylation. In addition to that, transcription

factor Sp3 is characterized by an inhibitory domain located near the C domain (Dennig et al., 1996).

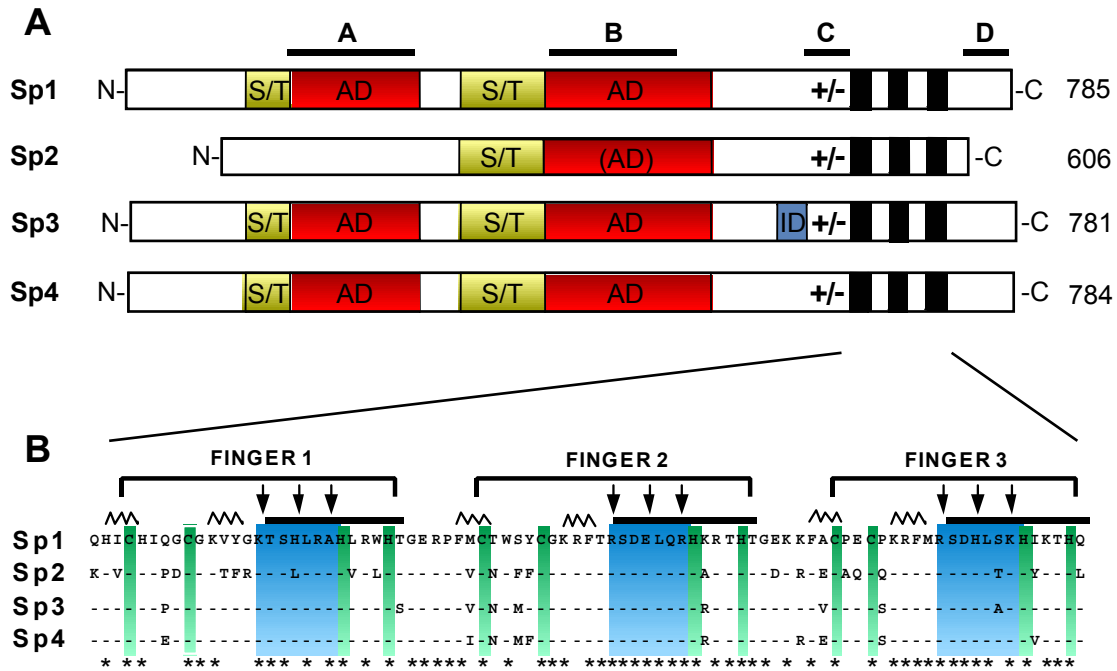


Fig. 1.4.2. Structural features of Sp proteins (Suske, 1999). **A.** Scheme of the transcription factors Sp1-4. Their length in amino acids is indicated on the right. Coloured boxes indicate the zinc finger region (black) as well as glutamine- (red) and serine/threonine-rich (yellow) protein domains; (+/-) represents a region of highly charged amino acids. Lines above the draw of the Sp1 protein indicate the extent of four regions (A, B, C and D) which contribute to the transcriptional properties of Sp1 as defined by Courey and Tjian (1988). Known activation (AD) and inhibitory domains (ID) are indicated. **B.** Protein sequence alignment of the zinc finger domains. Stars below the sequence indicate sequence identity. Cysteine and histidine residues which coordinate zinc ions are underlayed in green, protein regions which contact the DNA in blue. Arrows point to the amino acids which determine the recognition specificity by contacting specific bases of the DNA (adapted from Fairall et al., 1993; Pavletich and Pabo, 1991). Black lines and the zig-zag lines indicate alpha-helical and beta sheet regions, respectively.

Transcription factor Sp1

As first member of the Sp family, Sp1 was isolated from HeLa cells and described as transcription factor binding to the GC boxes of the Simian Virus 40 (SV40) promoter, thus leading to transcriptional activation *in vitro* (Dyran and Tjian, 1983) caused by two glutamine-rich domains (A and B, Fig. 1.4.2) (Courey and Tjian, 1988; Gill et al., 1994; Kadonaga et al., 1988). Human Sp1 consists of 785 amino acids with a calculated molecular weight of 80.6 kDa and is ubiquitously expressed.

It was discovered that Sp1 can directly interact with itself which has important implications for its transactivation capacity (Pascal and Tjian, 1991). Sp1 stimulates transcription from promoters as well as from enhancers (Courey et al., 1989). *In vitro* experiments suggest that this synergistic activation is mediated by Sp1 molecules bound to proximal and distal sites, which interact with each other thus forming stacked tetramers (Mastrangelo et al., 1991), thereby looping out the intervening DNA (Li et al., 1991; Mastrangelo et al., 1991; Su et al., 1991). For the multimerization, activation domain B appeared to be of critical importance (Pascal and Tjian, 1991). Together with domain A, domain B also mediates superactivation of Sp1-dependent transcription which can be achieved by non-DNA-binding mutants in case of multiple binding sites (Courey et al., 1989; Hagen et al., 1995). For synergistic activation by binding to multiple sites, domain D is required in addition to both transactivation domains (Pascal and Tjian, 1991).

Moreover, Sp1 interacts with components of the basal transcription machinery (e.g. TBP (Emili et al., 1994), several TAFs (Hoey et al., 1993; Tanese et al., 1996)) as well as with transcription factors like E2F (Karlsreder et al., 1996) and YY1 (Lee et al., 1993). Sp1 is able to recruit the co-activator complex CRSP (cofactor required for Sp1 activation), thus stimulating transcription of the respective genes (Ryu et al., 1999). In addition to that, Sp1 is target of posttranslational modifications like glycosylation (Jackson and Tjian, 1988) and phosphorylation (Jackson et al., 1990), both being able to influence e.g. its DNA binding or transactivation properties (e.g. Armstrong et al., 1997; Merchant et al., 1999; Roos et al., 1997; Yang et al., 2001).

To investigate the physiological function of Sp1, knockout mice were generated exhibiting deletion of the exons encoding the zinc finger domain (Marin et al., 1997). Sp1-deficient mice display a broad spectrum of abnormalities and die early during embryonic development (day E10.5 at the latest).

Transcription factor Sp2

Among Sp1-4, transcription factor Sp2 is the less explored factor. Sp2 was cloned from a T-cell library, which has been screened with a Sp1 zinc finger probe for homologous DNA sequences (Kingsley and Winoto, 1992). Due to the exchange of a critical amino acid in the first zinc finger necessary for contacting the DNA (Fig. 1.4.2 and Fig. 1.3.2), Kingsley and Winoto claim a higher binding affinity towards GT than

GC boxes. However, investigation of Sp2 binding capacity with a GT box oligonucleotide derived from the T-cell antigen receptor α (TCR α) only results in a low binding affinity (Kingsley and Winoto, 1992). In comparison to Sp1, Sp3 and Sp4, Sp2 exhibits the most structural differences among these factors, e.g. the existence of the first transactivation domain and serine/threonine region in Sp2 was unclear when starting the thesis. In addition, the D domain is absent. Data from the only report of a promoter that is affected by co-transfected N-terminally truncated Sp2 protein suggest that this factor may function in a cell-type-dependent manner (Bacovic et al., 2000). Sp2 repressed Sp1- or Sp3-driven activation of a construct containing the murine CTP:phosphocholine-cytidylyltransferase- α promoter in *Drosophila* SL2 cells but activated the same construct in C3H10T1/2 mammalian cells. Apart from that, no further reports about Sp2 were available when initiating the thesis.

Transcription factor Sp3

Transcription factor Sp3 was parallelly cloned in 1992 by Hagen et al. and Kingsley and Winoto. Sp3 is ubiquitously expressed and exists in four different isoforms, two of them with a molecular weight of ca. 97-115 kDa and two of them in a range of ca. 58-70 kDa. Whereas the long isoforms exhibit both transactivation domains (A and B; see Fig. 1.4.2), the two short isoforms only contain the B domain (Kenneth et al., 1997). Mutation analyses suggest that all four isoforms derive from alternative translational start sites. Moreover, an upstream open reading frame seems to regulate expression of the two long isoforms (Sapetschnig et al., 2004).

Concerning transactivation properties *in vivo*, Sp3 functions as a transcriptional activator (Ihn and Trochanowska, 1997; Liang et al., 1996; Udvadia et al., 1995; Zhao et al., 1997), but also displays no or only weak activity (Dennig et al., 1995; Majello et al., 1994), depending on the investigated promoter. In addition, the number of Sp3 binding sites inside a promoter seems to influence Sp3 activation capacity (Dennig, 1996) as well as the respective cell type (Hansen et al., 1999; Sjøttem et al., 1996). Moreover, Sp3 binds to GC boxes with similar affinity as transcription factor Sp1, thus being able to repress Sp1-mediated activation by competition (Birnbaum et al., 1995; Hagen et al., 1994).

Responsible for Sp3 transcriptional inactivity is a so-called inhibitory domain (Dennig et al., 1996), located near the C region of the Sp3 protein (Fig. 1.4.2). The inhibitory domain is characterized by the amino acid motif IKEE, which can be SUMOylated (Sapetschnig et al., 2002). In addition to that, yeast-two-hybrid screens led to the identification of PIAS1 (Doll, diploma work, 1998; Liu et al., 1998), an E3 ligase, which is involved in the SUMOylation process of proteins and binds to the Sp3 inhibitory domain (Sapetschnig et al., 2002).

To investigate Sp3 physiological function, knockout mice were generated exhibiting a deletion of exon 4, which encodes the two glutamine-rich transactivation domains (see Fig. 1.4.2). Heterozygous Sp3 knockout mice develop and reproduce normal and display no obvious phenotype apart from a slight growth retardation. In contrast to this, homozygous Sp3 knockout mice immediately die after birth due to a lung failure. However, the molecular mechanisms behind this are unclear. In addition, they are characterized by abnormalities in tooth (lack of amelogenin and ameloblastin) and bone development (reduced ossification due to the lack of osteocalcin) (Bouwman et al., 2000; Göllner et al., 2001b). Moreover, homozygous Sp3 knockout mice exhibit an impaired hematopoiesis (Van Loo et al., 2003).

Transcription factor Sp4

Transcription factor Sp4 was identified and cloned along with Sp3 due to its capacity to bind to GT boxes (Hagen et al., 1992). In contrast to Sp1 and Sp3, Sp4 is predominantly expressed in brain, heart, testicles and the epithelial tissue (Hagen et al., 1992; Supp et al., 1996). Sp4 also exhibits two glutamine-rich transactivation domains but in contrast to Sp1, Sp4 is unable to generate synergistic effects on multiple binding sites (Hagen et al., 1995). However, Sp4 can be super-activated by zinc-finger-less Sp1 and repressed by Sp3 (Hagen et al., 1995).

Heterozygous Sp4 knockout mice based on the deletion of exon 2 and 3 encoding the two transactivation domains exhibit no obvious phenotype. In contrast to this, Sp4-deficient mice are characterized by a general growth and weight reduction as well as a high post-natal mortality. Moreover, Sp4 knockout males display an abnormal reproduction behaviour and females a pronounced delay in sexual maturation (Göllner et al., 2001a).

1.5 Generation of transgenic mice by “gene targeting”

The principle utilized to target genes of an organism and to generate knockout mice is the process of so-called homologous recombination (Smithies et al., 1985). Homologous recombination occurs in eukaryotes during meiosis to preserve genomic diversity and is characterized by an exchange of identical or similar (= homologous) DNA sequences.

To generate knockout mice via “gene targeting”, totipotent embryonic stem (ES) cells are isolated from the inner cell mass of an agouti blastocyst (day E3.5) and transfected with a linearized knockout construct harbouring the desired mutation (e.g. point mutation, deletion, insertion) flanked by sequences that are homologous to the genomic region of interest (Fig. 1.5.1). Nowadays, different ES cell lines are available, which are mostly transfected by electroporation (e.g., Evans and Kaufman, 1981; Martin, 1981). By homologous recombination, the genomic regions being homologous to the vector sequences are exchanged and the mutation integrated into the genome.

As the homologous integration of vector sequences into the genome is a rare event, the targeted ES cells have to be selected. In general, this is done by two selection markers, a positive and a negative one (Mansour et al., 1988). Positive selection normally occurs by antibiotic resistance of the targeted ES cells. Commonly, a resistance gene for neomycin or its analogon G418 is integrated into the mouse genome together with the homologous genomic DNA sequences (Mansour et al., 1988; Southern and Berg, 1982).

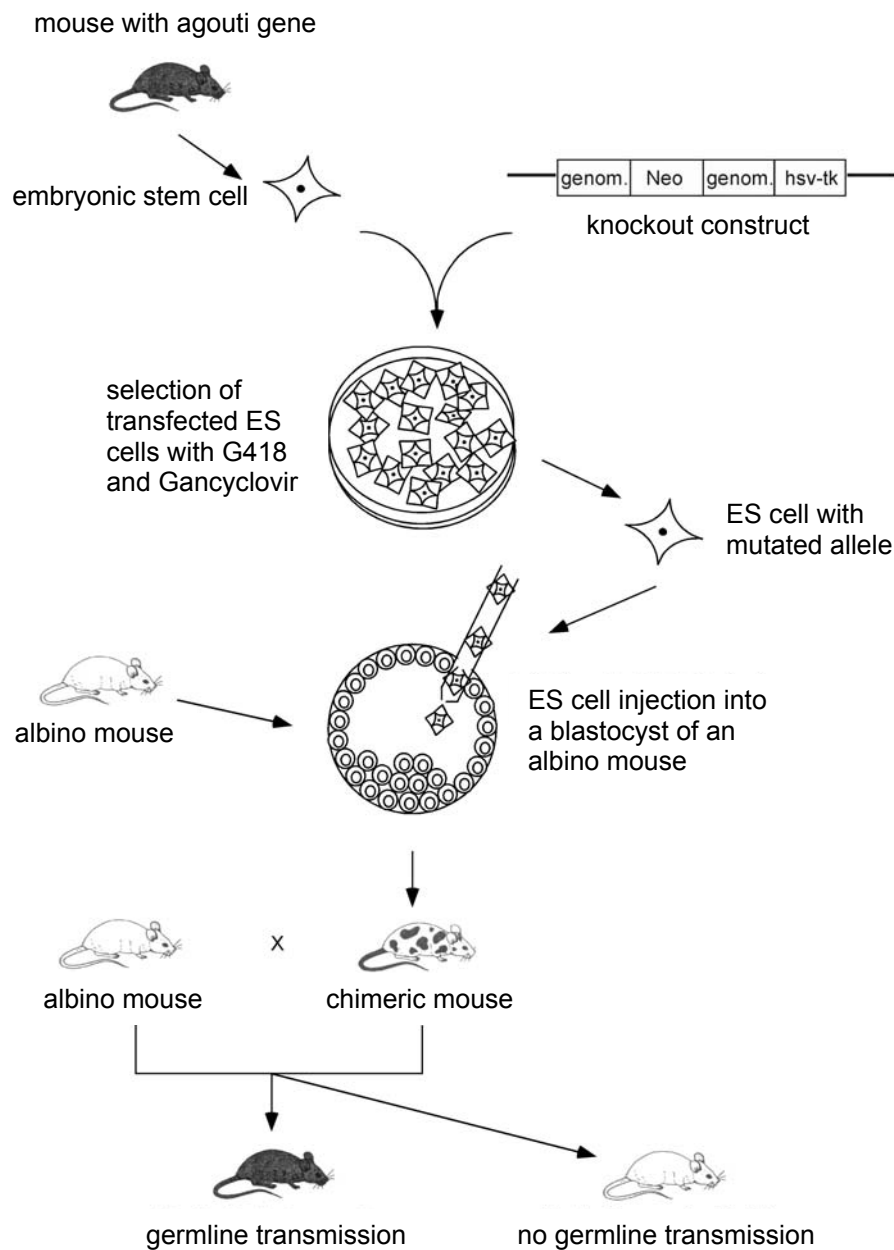


Fig. 1.5.1. Generation of transgenic mice by “gene targeting”. Embryonic stem (ES) cells derived from an agouti mouse are transfected with the linearized knockout construct and cultured on selection medium containing G418 and/or Gancyclovir. G418 and Gancyclovir enable the selection for the integration of the knockout construct into the ES cell genome based on homologous recombination. ES cells, which have been integrated the knockout construct by homologous recombination, are G418- and Gancyclovir-resistant (by Neo gene but not hsv-TK gene integration, the latter located outside the homologous region). These cells are injected into a blastocyst of an albino mouse and implanted into a pseudo-pregnant foster mouse, resulting in chimeric mice. Chimeric males are crossed with albino mice. In the case of germline transmission of the targeting mutation, mice being heterozygous for the entire construct in all body cells can be generated. In contrast to mice without germ line transmission, these mice are characterized by a brown coloured coat. If the knockout construct contains loxP sites, Cre-driven recombination can be performed either on the ES cell as well as on the mouse level enabling the generation of conditional knockout mice.

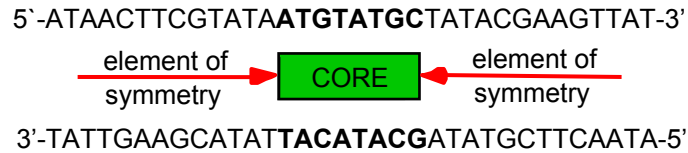
To test, whether the vector DNA integration is locus-dependent or spontaneous, which also occurs, a negative selection marker, located outside the homologous regions is used. Commonly, the gene coding for the herpes simplex Thymidine Kinase (hsv-TK) is used (Mansour et al., 1988). ES cells exhibiting Acyclovir or Gancyclovir sensitivity are characterized by spontaneous integration (i.e. also integrate the hsv-TK gene), whereas ES cells being resistant display vector DNA integration by homologous recombination (i.e. without the hsv-TK gene, located outside the homologous region). Thymidine Kinase expression in the cell results in phosphorylation of the selective reagent Gancyclovir. Phosphorylated Gancyclovir can be integrated during recombination into the DNA instead of thymidine, thus leading to replication incapability and cell death.

ES cells selected by G418 and Gancyclovir, thus exhibiting homologous integration of the entire knockout construct at one allele, are injected into an albino blastocyst, which is then implanted into a pseudo-pregnant foster mouse. Resulting embryos are chimeric, i.e. consist of wildtype and targeted cells (Bradley et al., 1984). As wildtype cells derived from an albino and targeted cells from an agouti mouse, they can be distinguished by their coat colours: wildtype cells display a white, targeted cells a brown coloured coat. To select mice with targeted germ cells, chimeric males are crossed with albino females. In the case of germline transmission, brown coloured embryos are obtained. In these embryos, all cells are heterozygous for the targeted gene locus. By intercrossing heterozygous animals, homozygous knockout mice can be generated (Bradley et al., Doetschman et al., 1987; Robertson et al., 1986).

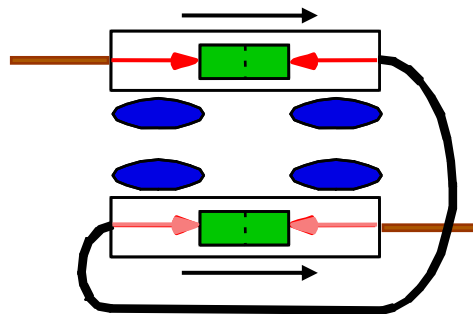
An option in gene targeting is the generation of conditional ES cells and knockout mice. Conditional gene targeting is characterized by the replacement of a genomic sequence by the same sequence that is flanked by two so-called loxP sites. LoxP sites (loci of crossing over derived from the recombination machinery of the P1 phage) consists of 34 highly conserved basepairs (Fig. 1.5.2). Eight central basepairs form the core sequence, which is flanked on each side by a 13 bp symmetry element in reverse orientation. Four monomers of the P1 phage enzyme Cre recombinase (causes recombination) bind to each two symmetry elements of two loxP sites (Hoess et al., 1982), thus forming a so-called Holliday intermediate complex (Fig. 1.5.2). The Cre recombinase monomers cut in the middle of the core sequence and either delete (if both loxP sites display the same orientation) or invert (if both loxP

sites display an opposite orientation) the DNA sequence between the two loxP sites (Stark et al., 1992).

A. loxP sites



B. complex of four recombinase monomers and two loxP sites



C. sequence between the two loxP sites is deleted



Fig. 1.5.2. The Cre-loxP system. **A.** Scheme of a loxP site. A loxP site consists of a 34 bp double strand. The 8 bp core sequence (green) is located in the center of the loxP site and is cut in the middle by the Cre recombinase. The core sequence is flanked on each side by a stretch of 13 bp, which are symmetrical (red arrows). **B.** Process of Cre-driven recombination. Four Cre recombinase monomers form a complex with two loxP sites by binding to their symmetry elements (red and pink arrows), thus resulting in a so-called Holliday intermediate. If the loxP sites display the same orientation (black arrows), Cre recombinase cuts in the middle (dashed) of the core sequence (green) and deletes the DNA sequence between the two loxP sites (black cord). **C.** Situation after Cre recombination. After sequence deletion, one loxP site remains consisting of two halves of each loxP site depicted in (B).

To enable the generation of conditional knockout mice, it has to be ensured that exon-intron structures of the targeted locus remain like in wildtype. For this, it is often necessary to delete at first the positive selection cassette via the Cre-loxP system, which can be done on the ES cell as well as on the mouse level. Depending on the research interests, these mice then can be crossed with mice expressing Cre

recombinase in a tissue-specific manner or at specific developmental time points, leading to a knockout of the gene.

1.6 Thesis aims

As mentioned above, no data concerning Sp2 protein expression as well as Sp2 function *in vitro* and *in vivo* were available when starting this thesis work. Therefore, the aims were the following:

To enable the analysis of Sp2 protein expression in various cell lines and adult mouse tissues as well as Sp2 subcellular localization, a first aim of this thesis was the generation of Sp2-specific rabbit polyclonal antibodies.

To characterize Sp2 at the molecular level, Sp2 transactivation function and DNA binding properties were investigated.

To study Sp2 function *in vivo*, the generation of Sp2 gene targeted mice was a further major aim of this thesis.

2. Materials and methods

2.1. Materials

2.1.1 Laboratory materials and devices

Unless stated otherwise, all laboratory materials and devices used for this work were obtained from: Amersham (Freiburg), Biometra (Göttingen), Biorad (München), Eppendorf (Hamburg), Falcon (Hamburg), Greiner (Frickenhausen), Heidolph (Schwabach), Heraeus/Kendro (Hanau), NanoDrop (USA: Wilmington), Nunc (Denmark: Roskilde) and Kodak (Stuttgart).

2.1.2 Chemicals

All chemicals used for this work were obtained in p.A. quality from Amersham (Freiburg), BD Becton & Dickinson (France: Le Pont de Claix), Calbiochem (Canada: La Jolla), Difco (USA: Sparks), Invitrogen (Karlsruhe), Gibco (Karlsruhe), Merck (Darmstadt), Perbio (Bonn), Riedel-de-Haen (Seelze), Roche (Mannheim), Roth (Karlsruhe), Serva (Heidelberg) and Sigma (München).

2.1.3 General solutions

50x Denhardts: 1% (w/v) Ficoll; 1% (w/v) Polyvinylpyrrolidon and 1% (w/v) BSA in H₂O (bidest.).

DEPC-H₂O: 1 ml Diethylpyrocarbonat (DEPC); 9 ml 100% EtOH; 990 ml H₂O (bidest.). Dissolving of DEPC by stirring o/n at 37°C, followed by autoclaving of the solution.

6x DNA loading buffer: 0.25% Bromphenol blue; 0.25% Xylene-Cyanol FF; 30% Glycerol H₂O (bidest.).

Orange G: 1 g Orange G; 20 g Ficoll ad 100 ml H₂O (bidest.).

1x PBS pH 7.4: 8 g NaCl; 0.2 g KCl; 1.15 g Na₂HPO₄ x 7 H₂O; 0.2 g KH₂PO₄ in 1 l H₂O (bidest.).

1x TBE: 89 mM Tris/HCl pH 8.0; 89 mM Boric acid; 2.5 mM EDTA in H₂O (bidest.).

10/1 TE: 10 mM Tris/HCl pH 8.0; 1 mM EDTA pH 8.0 in H₂O (bidest.).

1x TFB pH 6.3: 10 mM K-Mes; 100 mM KCl; 45 mM MnCl₂ x 4 H₂O; 10 mM CaCl₂ x 2 H₂O; 3 mM Hexamine cobalt chloride in H₂O (bidest.).

20x SSC pH 7.0: 3 M NaCl; 0.3 M Sodium citrate in H₂O (bidest.).

2.1.4 Culture media

2.1.4.1 Media to culture bacteria

Luria Bertani medium (LB): 10 g/l Select Peptone 140 (Invitrogen, Karlsruhe); 5 g/l Bacto™ Yeast Extract (BD, USA: Sparks); 10 g/l NaCl.

LB agar plates (with antibiotics): 15 g Select Agar (Invitrogen, Karlsruhe) per 1 l LB-Medium (with 50 mg/l of each antibiotic).

SOB medium: 20 g/l Select Peptone 140 (Invitrogen, Karlsruhe); 5 g/l Bacto™ Yeast Extract (BD, USA: Sparks); 0.548 g/l NaCl; 0.186 g/l KCl.

SOC medium: 10 ml SOB-Medium; 100 µl 2 M Mg²⁺ (1 M MgCl₂; 1 M MgSO₄); 100 µl Glucose.

2.1.4.2 Media to culture eukaryotic cells

Schneider medium: 500 ml Schneider's Drosophila medium (Gibco, Karlsruhe); 55 ml FBS (Sigma, München; suitable for insect cells); 5.5 ml 100x Penicillin/ Streptomycin (Cambrex, Belgium: Verviers); 5.5 ml L-Glutamine (Gibco, Karlsruhe). Culture medium for Drosophila SL2 cells (Schneider, 1972).

Fibroblast medium: 200 ml DMEM with GlutaMAX™ I (Gibco, Karlsruhe); 200 ml Nutrimix F10 (HAM) (Gibco, Karlsruhe); 45 ml FBS (PAA Laboratories, Austria: Parsching); 5 ml 100x Penicillin/ Streptomycin (Cambrex, Belgium: Verviers). Culture medium for mouse embryonic fibroblasts (MEFs).

HEK medium: 500 ml Nutrimix F10 (HAM) with GlutaMAX™ I (Gibco, Karlsruhe); 55 ml FBS (PAA Laboratories, Austria: Parsching); 5.5 ml 100x Penicillin/ Streptomycin (Cambrex, Belgium: Verviers). Culture medium for human embryonic kidney 293 cells (HEK-293).

Ishikawa medium: 500 ml 10x MEM-E (Gibco, Karlsruhe); 55 ml FBS (Gibco, Karlsruhe); 5.5 ml 100x Penicillin/ Streptomycin (Gibco, Karlsruhe); 5.5 ml 200 mM L-Glutamine (Gibco, Karlsruhe); 5.5 ml non-essential amino acids; 20 ml 7.5% NaHCO₃. Culture medium for Ishikawa cells.

2.1.5 Restriction enzymes and DNA-modifying enzymes

Restriction enzymes and DNA-modifying enzymes were obtained from Boehringer (Mannheim), Gibco (Karlsruhe), Invitrogen (Karlsruhe), New England Biolabs (Frankfurt), Promega (Mannheim) and Roche (Mannheim).

2.1.6 Antibiotics

Ampicillin	Bayer (Leverkusen)
Kanamycin	Boehringer (Mannheim)
G418	Gibco (Karlsruhe)
Penicillin/ Streptomycin	Gibco (Karlsruhe)

2.1.7. Antibodies

anti-Sp1; rabbit p.cl. antibody	Hagen et al. 1994 (Marburg)
anti-Sp2; rabbit p.cl. antibody (no. 193, "Zwick"; 10 days after 1. boost)	generated during this thesis
anti-Sp3; rabbit p.cl. antibody; sc-644	Santa Cruz (USA: St. Cruz)
anti-Gal4-DBD; rabbit m.cl. antibody; sc-577	Santa Cruz (USA: St. Cruz)
anti-rabbit; HRP-conj.; m.cl. sec. antib.; NA934V	Amersham Biosciences (Freiburg)
anti-rabbit; FITC-conj.; m.cl. sec.; 111-095-003	Jackson IR (USA: West Grove)

2.1.8 Radioactive substances

[α - ³² P] dCTP (3000 Ci/ mmol)	Amersham Biosciences (Freiburg)
[α - ³² P] dATP (3000 Ci/ mmol)	Amersham Biosciences (Freiburg)

2.1.9 Oligonucleotides

Oligonucleotides were synthesized by Eurogentec Deutschland (Köln), Invitrogen (Karlsruhe) and MWG AG Biotch (Ebersberg). The purpose of each oligonucleotide is also described in the chapters 2.2.2-4.

2.1.9.1 Oligonucleotides to generate Sp2 deletion mutants

Sp2- <i>Coli</i> -Expr-fw	5'-AGA-TGA-CCA-TGG-CCG-CCA-CTG-CTG-CT-3'
Sp2- <i>Coli</i> -Expr-rev	5'-GCC-TGG-ATC-CGC-ACC-TGT-CCA-TCA-TG-3'
Del1-Sp2-fw	5'-TTG-AAA-GGA-TCC-GCT-GTT-GAA-GCT-GCA-GTG-AC-3'
Del1-1-Sp2-fw	5'-TTG-AAA-GGA-TCC-CCT-ATC-AAA-CCC-GCT-CCT-CT-3'
Del1-2-Sp2-fw	5'-TTG-AAA-GGA-TCC-CTG-ATC-AAC-AAA-GGG-AGC-CG-3'
Del1-3-Sp2-fw	5'-TTG-AAA-GGA-TCC-ACC-CCG-TCA-ACA-TCT-GGT-CA-3'
Del1-4-Sp2-fw	5'-TTG-AAA-GGA-TCC-AGT-ACG-ACT-ACC-ACC-CCT-GT-3'
Del2-Sp2-fw	5'-TTG-AAA-GGA-TCC-CCG-CTC-AAC-AAC-CTG-GTG-AA-3'
Del3-Sp2-fw	5'-TTG-AAA-GGA-TCC-CGT-ACA-CCT-TCT-GGT-GAG-GT-3'
Del4-Sp2-fw	5'-TTG-AAA-GGA-TCC-GTG-CCT-GTC-ACC-ATC-ACC-AA-3'
Del1-Sp2-rev-sh	5'-AAC-TTT-GGA-TCC-GTC-CTG-GGT-TAG-AAC-GTC-TC-3'
Sp2FL- <i>Xho</i> I-fw	5'-TTG-AAA-CTC-GAG-GAG-CGA-TCCACA-GAT-GAG-CA-3'
Del1-2- <i>Bam</i> HI-rev	5'-AAC-TTT-GGA-TCC-GGC-TCC-CTT-TGT-TGA-TCA-GG-3'
Del2- <i>Bam</i> HI-fw	5'-TTG-AAA-GGA-TCC-CGC-TCA-ACA-ACC-TGG-TGA-AC-3'
Del1- <i>Bgl</i> II-rev-sh	5'-AAC-TTT-AGA-TCT-GTC-CTG-GGT-TAG-AAC-GTC-TC-3'
Gal4-Sp2-pLALL-fw	5'-TTG-AAA-GGA-TCC-CTG-TCA-GTC-CCA-GTG-ACT-AC-3'
Gal4-Sp2-pE-rev	5'-TCC-TTT-CTA-GAT-CAG-CCC-ACT-GAT-AGT-CAG-G-3'
Gal4-Sp2-mE-rev	5'-CCC-TTT-CTA-GAG-TAC-CTG-ACC-AGA-AGG-TGT-A-3'
Gal4-Sp2-mG-rev	5'-CCC-TTT-CTA-GAT-GTT-CAC-CAG-GTT-GTT-GAG-C-3'

2.1.9.2 Oligonucleotides for Electrophoretic Mobility Shift Assays (EMSAs)

Sp1-1-fw	5'-AGC-TCT- <u>CCC-CGC-CCC</u> -CCG-3'
Sp1-1-rev	5'-TCG-ACG- <u>GGG-GGC-GGG</u> -GAG-3'
Sp2-A-fw	5'-TCG-ACC-GTT- <u>GGG-GCG-GAG</u> -CTT-CAC-3'
Sp2-A-rev	5'-TCG-AGT-GAA- <u>GCT-CCG-CCC</u> -CAA-CGG-3'
Sp2-C-fw	5'-TCG-ACC-GTT- <u>GGG-GCG-GCG</u> -CTT-CAC-3'
Sp2-C-rev	5'-TCG-AGT-GAA- <u>GCG-CCG-CCC</u> -CAA-CGG-3'
Sp2-G-fw	5'-TCG-ACC-GTT- <u>GGG-GCG-GGG</u> -CTT-CAC-3'
Sp2-G-rev	5'-TCG-AGT-GAA- <u>GCC-CCG-CCC</u> -CAA-CGG-3'
Sp2-T-fw	5'-TCG-ACC-GTT- <u>GGG-GCG-GTG</u> -CTT-CAC-3'
Sp2-T-rev	5'-TCG-AGT-GAA- <u>GCA-CCG-CCC</u> -CAA-CGG-3'
GT-1-fw	5'-AGC-TTC-CGT- <u>TGG-GGT-GTG</u> -GCT-TCA-CG-3'
GT-1-rev	5'-TCG-ACG-TGA-AGC- <u>CAC-ACC-CCA</u> -ACG-GA-3'
CT-box-5	5'-AGC-TGG- <u>CGC-CTC-CCC</u> -TGA-3'
CT-box-3	5'-TCG-ATC- <u>AGG-GGA-GGC</u> -GCC-3'
WT-1-fw	5'-AGC-TTC- <u>TCC-CTC-CCC</u> -CTT-3'
WT-1-rev	5'-TCG-AAA- <u>GGG-GGA-GGG</u> -AGA-3'
Gal4-C	5'-GCT-TAG-CGG-AGT-ACT-GTC-CTC-CGA-T-3'
Gal4-D	5'-GGG-ATC-GGA-GGA-CAG-TAC-TCC-GCT-A-3'

Underlined nucleotides represent GC, GT and CT boxes as well as mutated versions of these boxes and the Gal4 binding site. Point mutations in the GC box sequence are depicted in red.

2.1.9.3 Oligonucleotides for RT-PCR, Southern Blot probes and cosmid library screen (RZPD)

MSp2-ES-RT-fw	5'-ATG-AGC-GAT-CCA-CAG-ATG-AGC-A-3'
MSp2-ES-RT-rev	5'-ACT-TGA-CTT-CTG-GAC-AGG-AGC-T-3'
Sp2-Ex5-RT-fw	5'-ATT-CAG-CTG-CCA-TTC-TCC-GA-3'
Sp2-Ex6-RT-rev	5'-AGC-CCA-CTG-ATA-GTC-AGG-TT-3'
Sp2-ES-a-SB-fw	5'-GCT-GGC-TCC-ACA-TCA-CTT-TG-3'
Sp2-ES-a-SB-rev	5'-AGG-TGA-CGG-TTG-ATG-TGG-TT-3'

Sp2-Ex4-RZPD-fw 5'-CCG-TCA-ACA-TCT-GGT-CAT-AAG-C-3'
 Sp2-Ex4-RZPD-rev 5'-CTG-AAA-GTT-CTG-AGA-GGG-CTT-C-3'
 Cyclophilin-fw 5'-TCA-CCA-TTT-CCG-ACT-GTG-GAC-3'
 Cyclophilin-rev 5'-ACA-GGA-CAT-TGC-GAG-CAG-ATG-3'

2.1.9.4 Oligonucleotides for loxP site generation and PCR amplification of Sp2 genomic fragments

Sp2-loxP1-fw 5'-CTA-GAA-CTG-TGG-TCG-ACA-TAA-CTT-CGT-ATA-ATG-TAT-GCT-ATA-CGA-AGT-TAT-GGT-AC-3'
 Sp2-loxP1-rev 5'-CAT-AAC-TTC-GTA-TAG-CAT-ACA-TTA-TAC-GAA-GTT-ATG-TCG-ACC-ACA-GTT-3'
 Sp2-loxP2-fw 5'-GGC-CGC-ACT-GTG-CTC-GAG-ATA-ACT-TCG-TAT-AAT-GTA-TGC-TAT-ACG-AAG-TTA-TGT-CGA-CAC-TGT-G-3'
 Sp2-loxP2-rev 5'-TCG-ACA-CAG-TGT-CGA-CAT-AAC-TTC-GTA-TAG-CAT-ACA-TTA-TAC-GAA-GTT-ATC-TCG-AGC-ACA-GTG-C-3'
 Sp2-loxP3-fw 5'-GGC-CGC-ACT-GTG-GGA-TCC-ATA-ACT-TCG-TAT-AAT-GTA-TGC-TAT-ACG-AAG-TTA-TC-3'
 Sp2-loxP3-rev 5'-GGC-CGA-TAA-CTT-CGT-ATA-GCA-TAC-ATT-ATA-CGA-AGT-TAT-GGA-TCC-CAC-AGT-GC-3'
 MSp2-In1-fw 5'-ATT-ATA-GCG-GCC-GCG-TGT-CTG-TAA-TAT-C-3'
 MSp2-In1-rev 5'-TGA-CTC-GGA-TCC-GCA-GGG-ATA-ACA-ACA-CAC-CCA-GGT-3'
 MSp2-In2-fw 5'-GAA-TGA-GAA-TTC-GAA-CTA-GAA-CC-3'
 MSp2-In2-rev 5'-TGC-GAG-GAA-TTC-TTT-CTG-AAA-GA-3'
 MSp2-Ex4-fw 5'-TAT-TAT-TGC-GGC-CGC-CGA-CAC-TAG-GCC-AGC-AGG-GC-3'
 MSp2-Ex4-rev 5'-CGG-CAC-TCG-AGT-CAT-TCT-TGC-TGT-TAT-ATA-TAC-TC-3'

2.1.9.5. Oligonucleotides for mouse genotyping

MSp2Ex/13Seq3 5'-CCC-TCT-CAG-AAC-TTT-CAG-ATC-3'
 MSp2In2Seq4 5'-CTT-AGG-AGG-GAT-CTA-GAC-TAG-3'

Neo

5'-CAT-CGC-CTT-CTA-TCG-CCT-TCT-TGA-3'

2.1.10 Plasmids

2.1.10.1 Previously described plasmids

pET3a-d: Plasmid for protein expression in *E. coli* BL21DE3 bacteria. One characteristic feature of the BL21DE3 strain is an IPTG-inducible T7 RNA polymerase; a-d represent different reading frames. For further details, see Studier et al. (1990).

pPacHD: Drosophila expression plasmid (Bond and Davidson, 1986).

pPacHD-Flag: Drosophila expression plasmid, containing a Flag/ HA epitope.

pPacUbx: Drosophila expression plasmid, containing an Ubx leader sequence. Plasmid was obtained as a *Bam*HI fragment of pPacSp1-DBD.

pPac-Sp1: Plasmid for the expression of full-length Sp1 protein in Drosophila SL2 cells. For detailed information, the reader is referred to Courey and Tjian (1988) and Pascal and Tjian (1991).

pPac-Sp3FL-new: Plasmid for the expression of full-length Sp3 protein (containing both upstream AUGs) in Drosophila SL2 cells.

pPacUbx-Sp1ZF: Plasmid for the expression of the Sp1 zinc finger region (DNA-binding domain) in Drosophila SL2 cells.

pSG424Gal4: Plasmid for the expression of the Gal4 DNA-binding domain in mammalian cells (HEK-293). Plasmid was obtained as a *Bam*HI/*Xba*I fragment from pSG424Gal4-Sp1A.

pSG424Gal4-Sp1A: Mammalian expression plasmid encoding the approx. 186 amino acid Sp1A fragment N-terminally fused to a Gal4 DNA binding domain (Southgate et al., unpublished).

pCMVSPORT6: Mammalian expression plasmid. Also usable for in-vitro transcription/translation. Invitrogen (Karlsruhe).

pCMVSPORT6-mSp2cDNA: Plasmid for the expression of mouse full-length Sp2 (aa1-606) in mammalian cells. The plasmid was obtained from the RZPD Deutsches Ressourcenzentrum für Genomforschung GmbH (Heidelberg), clone ID: IRAKp961P0541. Sp2 cDNA can be isolated as a *Sall/NotI* fragment.

pN3: Mammalian expression plasmid. Plasmid was obtained as a *Sall/NotI* fragment from pEGFP-N3. Clontech (USA: Palo Alto).

pEGFP-C1: Enhanced green fluorescent protein (EGFP) expression plasmid to optically determine transfection efficiencies in mammalian cells. Clontech (USA: Palo Alto).

p97b: B-galactosidase expression plasmid to normalize transfection efficiencies in Drosophila SL2 cells. For detailed information, the reader is referred to Di Nocera and Dawid (1983) as well as Hagen et al. (1994).

pRSV- β -Gal: B-galactosidase expression plasmid to normalize transfection efficiencies in mammalian cells. For detailed information, the reader is referred to Hagen et al. (1995).

pGL3-Prom: Luciferase reporter plasmid to measure luciferase activity in Drosophila SL2 cells. The luciferase gene is driven by the Simian Virus 40 (SV40) promoter containing five Sp1 binding sites (GC boxes). For detailed information, the reader is referred to Dynan and Tjian (1983). Promega (USA: Madison).

pBCAT-2: Plasmid for transfections in Drosophila SL2 cells expressing a chloramphenicol acetyl transferase (CAT) reporter gene driven by an artificial promoter containing two Sp1 binding sites of the HTLV promoter and an E1b-TATA

box. For detailed information, the reader is referred to Hagen et al. (1994) and Pascal and Tjian (1991).

p(GC)₂-CAT: Plasmid for transfections in *Drosophila* SL2 cells expressing a chloramphenicol acetyl transferase (CAT) reporter gene driven by an artificial promoter containing two GC boxes and the E1b-TATA box.

pGAWG5E1b: Plasmid for transfections in mammalian cells expressing a Luciferase reporter gene driven by an artificial promoter containing five Gal4 binding sites and the E1b-TATA box.

pPNT: Plasmid containing a pgk-driven neomycin resistance (Neo) and herpes simplex Thymidine Kinase (hsv-TK) gene (Tybulewicz et al., 1991; see also Mangold, diploma work, 1995).

pGT1,8-IRES-βGeo: Plasmid containing the En2-SA-IRES-LacZ-Neo-SVpA cassette, which can be isolated as a *SalI* fragment (Mountford and Skarnes, unpublished; see also Mangold, diploma work, 1995). Obtained from S. Philipsen (Rotterdam). En2: intron of the engrailed gene; SA: splice acceptor site; IRES: internal ribosomal entry site; LacZ: β-galactosidase gene; Neo: neomycin resistance gene; SVpA: poly-adenylation signal derived from Simian Virus.

2.1.10.2 Plasmids generated during this thesis work

For those fragments, which have been amplified by PCR, the reader should also consult chapter 2.1.9 for the oligonucleotide sequences and chapter 2.2.2.3 for the resulting PCR fragments.

2.1.10.2.1 Bacterial expression plasmids

pET3d-mSp2: Plasmid for the expression of full-length mouse Sp2 in *E. coli* BL21DE3 bacteria. The Sp2 fragment was amplified by PCR with primer Sp2-*Coli*-

Expr-fw and Sp2-*Coli*-Expr-rev, the first one containing a *NcoI* site, the second one containing a *Bam*HI site. Sp2 fragment was subcloned into pET3d via *NcoI/Bam*HI.

2.1.10.2.2 *Drosophila* expression plasmids

pPac-mSp2(1-606): Plasmid for the expression of mouse Sp2 (aa1-606) in *Drosophila* SL2 cells. Sp2 was subcloned into pPac as a *Sa*II/*Not*I fragment with *Bam*HI linkers obtained from pCMVSPORT6-mSp2cDNA.

pPacUbx-mSp2(62-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa62-606) in *Drosophila* SL2 cells. Sp2 fragment was amplified by PCR with primer Del1-1-Sp2-fw and Del1-Sp2-rev-sh, both containing a *Bam*HI site, and subcloned into pPacUbx via *Bam*HI.

pPacUbx-mSp2(112-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa112-606) in *Drosophila* SL2 cells. Sp2 fragment was amplified by PCR with primer Del1-2-Sp2-fw and Del1-Sp2-rev-sh, both containing a *Bam*HI site, and subcloned into pPacUbx via *Bam*HI.

pPacUbx-mSp2(160-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa160-606) in *Drosophila* SL2 cells. Sp2 fragment was amplified by PCR with primer Del1-3-Sp2-fw and Del1-Sp2-rev-sh, both containing a *Bam*HI site, and subcloned into pPacUbx via *Bam*HI.

pPacUbx-mSp2(180-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa180-606) in *Drosophila* SL2 cells. Sp2 fragment was amplified by PCR with primer Del1-4-Sp2-fw and Del1-Sp2-rev-sh, both containing a *Bam*HI site, and subcloned into pPacUbx via *Bam*HI.

pPacUbx-mSp2(207-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa207-606) in *Drosophila* SL2 cells. Sp2 fragment was amplified by PCR with primer Del2-Sp2-fw and Del1-Sp2-rev-sh, both containing a *Bam*HI site, and subcloned into pPacUbx via *Bam*HI.

pPacUbx-mSp2(349-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa349-606) in *Drosophila* SL2 cells. Sp2 fragment was amplified by PCR with primer Del3-Sp2-fw and Del1-Sp2-rev-sh, both containing a *Bam*HI site, and subcloned into pPacUbx via *Bam*HI.

pPacUbx-mSp2(441-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa441-606) in *Drosophila* SL2 cells. Sp2 fragment was amplified by PCR with primer Del4-Sp2-fw and Del1-Sp2-rev-sh, both containing a *Bam*HI site, and subcloned into pPacUbx via *Bam*HI.

pPacUbx-mSp2(478-606): Plasmid for the expression of the N-terminal Sp2 deletion fragment Sp2 (aa478-606), the DNA binding domain (Sp2 zinc finger region), in *Drosophila* SL2 cells. Sp2 fragment was subcloned as a *Pst*I fragment of full-length mouse Sp2 cDNA with *Bam*HI linker (Klenow filling) into pPacUbx via *Bam*HI.

pPacFlag/HA-mSp2(Δ 112-207): *Drosophila* expression plasmid encoding a Sp2 deletion fragment lacking aa112-207. Both Sp2 fragments, Sp2 (aa1-111) and Sp2 (aa208-606), were amplified by PCR. Primers for the amplification were Sp2FL-*Xho*I-fw and Del1-2-*Bam*HI-rev for Sp2 (aa1-111), Del2-*Bam*HI-fw and Del1-*Bgl*II-rev-sh for Sp2 (aa208-606). Sp2 (aa1-111) was subcloned into pPacHD-Flag via *Xho*I/*Bam*HI, Sp2 (aa208-606) via *Bam*HI, leaving a non-functional *Bam*HI/*Bgl*II site at the 3'-end of the fragment.

2.1.10.2.3 Mammalian expression plasmids

pN3-mSp2cDNA: Plasmid for the expression of mouse full-length Sp2 (aa1-606) in mammalian cells. Sp2 cDNA was subcloned into pN3 via *Sal*I/*Not*I obtained from pCMVSPORT6-mSp2cDNA.

pSG424Gal4-mSp2(6-471): Plasmid for the expression of the fused Gal4-Sp2 deletion fragment Sp2 (aa6-471) in HEK-293 cells. Sp2 fragment was amplified by PCR with primer Gal4-Sp2-pLALL-fw and Gal4-Sp2-pE-rev, the first one containing a *Bam*HI site, the second one a *Xba*I site. The Sp2 fragment was subcloned into pSG424Gal4 via *Bam*HI/*Xba*I.

pSG424Gal4-mSp2(6-357): Plasmid for the expression of the fused Gal4-Sp2 deletion fragment Sp2 (aa6-357) in HEK-293 cells. Sp2 fragment was amplified by PCR with primer Gal4-Sp2-pLALL-fw and Gal4-Sp2-mE-rev, the first one containing a *Bam*HI site, the second one a *Xba*I site. The Sp2 fragment was subcloned into pSG424Gal4 via *Bam*HI/*Xba*I.

pSG424Gal4-mSp2(6-215): Plasmid for the expression of the fused Gal4-Sp2 deletion fragment Sp2 (aa6-215) in HEK-293 cells. Sp2 fragment was amplified by PCR with primer Gal4-Sp2-pLALL-fw and Gal4-Sp2-mG-rev, the first one containing a *Bam*HI site, the second one a *Xba*I site. The Sp2 fragment was subcloned into pSG424Gal4 via *Bam*HI/*Xba*I.

pSG424Gal4-mSp2(207-471): Plasmid for the expression of the fused Gal4-Sp2 deletion fragment Sp2 (aa207-471) in HEK-293 cells. Sp2 fragment was amplified by PCR with primer Del2-*Bam*HI-fw and Gal4-Sp2-pE-rev, the first one containing a *Bam*HI site, the second one a *Xba*I site. The Sp2 fragment was subcloned into pSG424Gal4 via *Bam*HI/*Xba*I.

2.1.10.2.4 Knockout construct and pre-constructs

pPNT-loxP1: Plasmid used to clone the conditional Sp2 knockout construct. First pre-construct. A *Xba*I-*Sal*I-LoxP-*Kpn*I fragment was synthesized by MWG AG Biotech (Ebersberg), hybridized and subcloned into the pPNT vector (see 2.1.10.1).

pPNT-loxP2: Plasmid used to clone the conditional Sp2 knockout construct. Second pre-construct. A *Not*I-*Xho*I-LoxP-*Sal*I-[*Xho*I] fragment was synthesized by MWG AG Biotech (Ebersberg), hybridized and subcloned into pPNT-loxP1.

pPNT-Intron4: Plasmid used to clone the conditional Sp2 knockout construct. Third pre-construct. A ca. 1.8 kb Sp2 genomic fragment downstream of exon 4 (= intron 4; genomic DNA: nt32067-nt33907; see 6.2) was amplified by PCR with primer MSp2-In2-fw and MSp2-In2-rev and subcloned into pPNT-loxP2 via *Eco*RI.

pPNT-Exon4: Plasmid used to clone the conditional Sp2 knockout construct. Fourth pre-construct. Sp2 exon 4 flanked by intronic regions (genomic DNA: nt30460-nt32066; see 6.2) was amplified by PCR with primer MSp2-Ex4-fw and MSp2-Ex4-rev, the first one containing a *NotI*, the second one a *XhoI* linker. The ca. 1.6 kb fragment was subcloned into pPNT-Intron5 via *NotI/XhoI*.

pPNT-loxP3: Plasmid used to clone the conditional Sp2 knockout construct. Fifth pre-construct. A *NotI-BamHI-LoxP-[NotI]* fragment was synthesized by MWG AG Biotech (Ebersberg), hybridized and subcloned into pPNT-Exon4.

pPNT-Intron3: Plasmid used to clone the conditional Sp2 knockout construct. Sixth pre-construct. A ca. 3.2 kb Sp2 genomic fragment upstream of exon 4 (= intron 3; genomic DNA: nt27217-nt30459; see 6.2) was amplified by PCR with primer MSp2-In1-fw and MSp2-In1-rev, the first one containing a *NotI*, the second one a *BamHI* linker. The fragment was subcloned into pPNT-loxP3 via *NotI/BamHI*.

pPNT-cSp2ko: Final construct to target the Sp2 gene in the mouse. A ca. 7.4 kb En2-SA-IRES-LacZ-Neo-SVpA fragment derived from pGT1,8-IRES- β Geo was subcloned into pPNT-Intron4 via *SaI*.

For further details concerning the generation of the final knockout construct and all pre-constructs, the reader is also referred to chapter 3.2.4.

2.1.11 Cosmids

To receive genomic fragments of murine Sp2, a 129/ola mouse cosmid library (library no. 121) was screened at the RZPD Deutsches Ressourcenzentrum für Genomforschung GmbH (Heidelberg) with a DNA probe against mouse Sp2 exon 4. The probe (called “Ex4-RZPD”) was amplified by PCR with primer MSp2-Ex4-RZPD-fw and MSp2-Ex4-RZPD-rev (see 2.1.9.3). In total, 15 clones were identified and four of them verified by Southern Blot analysis. Clone number MPMGc121L17390Q2 was used to amplify the genomic fragments for the conditional knockout construct by PCR. Cosmids consist of a 8.7 kb Lawrist 7 vector (ampicillin resistance) usable in *E. coli* and a 38 kb genomic insert.

2.1.12 Bacterial strains

***Escherichia coli* DH5 α** : Bacterial strain used for plasmid amplification. For detailed information, see Hanahan (1983).

***Escherichia coli* BL21DE3**: Bacterial strain used for protein overexpression. This strain constitutively expresses the Lac repressor. Moreover, expression of a T7 polymerase gene is regulated by the IPTG-inducible Lac promoter. Consequently, target gene expression is regulated by an IPTG-induced T7 polymerase gene expression. In addition, the Lon and OmpT protease genes are absent in BL21DE3, thus leading to a decrease in protease activity.

***Escherichia coli* 294-Cre**: Bacterial strain expressing Cre recombinase, thus providing a simple assay to test for the recombination competence of constructs that are designed for use in Cre-mediated genomic manipulations. Cre recombinase expression is regulated by the λ -P_R promoter and the temperature-sensitive cI857 repressor. Recombination competence of loxP sites engineered into a plasmid can be tested by transformation into the 294-Cre strain and overnight growth at 37°C followed by restriction digestion checks of the plasmid DNA. For further details, the reader is referred to Buchholz et al. (1996).

2.1.13 Cell lines

***Drosophila* SL2 cells**: Insect cells lacking Sp factors. For a detailed description, see Schneider (1972) and Suske (2000).

Mouse embryonic fibroblasts (MEF): Wildtype cell line derived from day E13.5 mouse embryos of Sp3 heterozygous crossings (Göllner; diploma work, 1998).

Mouse embryonic stem cells (ES): Mouse embryonic stem cells used for Sp2 gene targeting derived from the ES cell line C88Bl/6 TdW and were cultured and handled by Nynke Gillemans at the Erasmus University of Rotterdam.

Human embryonic kidney 293 cells (HEK-293): Cell line derived from human embryonic kidney epithelial cells, transformed with adenovirus 5 DNA.

HEK-293-pGAW5E1b cells: A HEK-293 cell line, which has been stably transfected with plasmid pGAW5E1b, thus constitutively expressing a Luciferase reporter gene regulated by an artificial promoter containing five Gal4 binding sites and an E1b-TATA box.

HeLa (Henrietta Lacks' cells): Cell line derived from human epithelial cells of a cervix adeno-carcinoma. HeLa cells are aneuploid and steroid hormone receptor negative. For Western Blot analyses or Electrophoretic Mobility Shift Assays (EMSAs), nuclear extracts of these cells were kindly provided by Alexandra Sapetschnig.

Ishikawa cells: Cell line derived from human endometrial cells of an adeno-carcinoma (Nishida et al., 1985).

2.1.14 Mice

The European house mouse *Mus musculus*, L. (Mammalia: Rodentia) was used as model organism for this thesis. To generate knockout mice, embryonic stem cells, derived from 129/ola mouse blastocysts, were transfected with the knockout construct, injected into blastocysts and the blastocysts implanted into pseudo-pregnant foster mice (see also 2.2.4.5). For more detailed information, see Bishop et al. (1985). Additionally, all data concerning mouse strains and breeding are available at the mouse data bank of the Erasmus University of Rotterdam (EUR).

2.1.15 Rabbits

Two New Zealand White rabbits (rabbit no. 193 (“Zwick”) and no. 194 (“Zwack”); European wild rabbit *Oryctolagus cuniculus*, L. (Mammalia: Lagomorpha)) were used to generate polyclonal Sp2-specific antibodies. The detailed procedure of Sp2 antibody generation is described in chapter 2.2.3.6.

2.2 Methods

2.2.1 Radiation protection and biological safety

Genetic engineering experiments were performed in accordance with the guidelines for genetic engineering (Gentechnikgesetz vom 16. Dezember 1993). Contaminated materials and solutions were collected separately and autoclaved. Experiments with radioactive substances, including waste disposal, followed the guidelines for radiation protection (Strahlenschutzverordnung vom 20. Juli 2001).

2.2.2 Molecular biological methods

2.2.2.1 RNA experiments

To protect RNA from ribonucleases (RNases), all experiments were performed as sterile as possible: All solutions were freshly prepared, autoclaved and used only for RNA experiments. As water, fresh Braun's aqua ad iniectabilia (Braun, Melsungen) or water treated with Diethylpyrocarbonat (DEPC-H₂O) was used. Plastic tubes and pipette tips were used freshly and autoclaved; glass ware was cleaned with HCl (conc.), followed by 2 M NaOH and DEPC-H₂O. For all experiments, vinyl gloves were used. Unless stated otherwise, all experiments were performed at 4°C.

2.2.2.1.1 RNA isolation and purification

To isolate RNA from mouse embryonic stem cells and embryonic or adult organs, two different methods were applied:

RNeasy Mini Kit (Qiagen, Hilden)

To isolate RNA from mouse embryonic stem cells, confluent cells were used. To isolate RNA from adult mouse organs, organs were frozen in liquid nitrogen immediately after mouse dissection. Organs were ground with a mortar on dry ice and dissolved in RNeasy Mini Kit RLT buffer (600 µl buffer per 30 mg pulverized organ). To remove insoluble tissue particles, the solution was shortly centrifuged and

the supernatant dissolved in the same volume of 70% ethanol. The further isolation procedure for embryonic stem cells and organs followed the RNeasy Mini Kit user manual (Qiagen, Hilden). For more detailed information, the reader is referred to the manual and/or to Göllner (PhD thesis, 2002), who performed the isolation.

LiCl/ Urea protocol

To isolate RNA from mouse embryonic organs, organs derived from day E18.5 embryos were frozen in liquid nitrogen immediately after embryo dissection. Frozen organs were pulverized in 2-3 ml 3 M LiCl + 6 M urea by high pressure mortaring (Ultrathorax; Heidolph, Schwabach). The cell suspension was centrifuged at 4°C for 25 min and the resulting pellet dissolved in 500 µl RB buffer + 500 µl phenol/ chloroform after washing with 3 M LiCl + 6 M urea for a second time. To facilitate dissolving, the solution was shaken at room temperature for 45 min. After 5 min of centrifugation, the water phase, containing the RNA, was dissolved in DEPC-H₂O or precipitated in 3 volumes of ethanol + 1/10 volumes of 2 M sodium acetate pH 5.6 to store (in solution) at -20°C.

RB buffer: 10 mM Tris pH 7.5; 1mM EDTA; 0.5 % SDS.

2.2.2.1.2 RNA quantification

RNA was quantified by UV photometry (Amersham Biosciences, Freiburg; or NanoDrop Technologies, USA: Wilmington). The UV absorption by nuclear acids were measured at a wavelength of $\lambda = 260$ nm. At this wavelength the side chains of the aromatic amino acids display maximal absorption. The absorption value allows to calculate RNA concentration of a solution (1 OD₂₆₀ = 40 µg/ml). In addition, RNA was also quantified by agarose gel electrophoresis.

2.2.2.1.3 Reverse transcriptase polymerase chain reaction (RT-PCR)

Two different methods were applied to perform reverse transcriptase polymerase chain reactions (RT-PCR):

RT-PCR (standard protocol)

To investigate Sp2 mRNA expression in mouse embryos, RNA from day E18.5 mouse embryos were prepared (LiCl/ urea method). A SuperScriptII reverse transcriptase and a temperature stable DNA polymerase from *Thermus aquaticus* (*Taq*) (Invitrogen, Karlsruhe) were utilized for cDNA synthesis (RT) and amplification (PCR), respectively. The single reaction steps were as follows:

DNase treatment:

	<u>10 µl reaction</u>
DEPC-H ₂ O	ad 10 µl
RNA (E18.5 embryos)	2 µg
10x DNase reaction buffer	1 µl
DNase (1 U/µl) (Promega, Mannheim)	1 µl

Reaction was performed at 37°C for 30 min. To stop the reaction, 1 µl of 25 mM EDTA was added to the reaction mix followed by 10 min incubation at 65°C and cooling down on ice.

Reverse transcriptase:

To denature RNA secondary structures, RNA mix (after DNase treatment) was incubated at 90°C for 5 min before preparing the reverse transcription reaction. Components of the reaction mix were obtained from Invitrogen (Karlsruhe).

	<u>20 µl reaction</u>
DEPC-H ₂ O	ad 20 µl
RNA mix (after DNase treatment)	8 µl
5x RT reaction buffer	4 µl
10 mM dNTP mix	1 µl
100 mM DTT	2 µl
Oligo dT (0.5 µg/µl)	1 µl
RNase inhibitor (5 U/µl)	1 µl
SuperScriptII RT (1 U/µl)	1 µl

Reaction was performed at 42°C for 1 hour. After a 15 min incubation at 70°C and cooling-down at 4°C, 0.5 µl of RNase H (1 U/µl) were added to the reaction mix and incubated at 37°C for 30 min to remove RNA. Thus, the synthesized cDNA could be used for PCR amplification (reaction components: Invitrogen, Karlsruhe).

Polymerase chain reaction (PCR):

	<u>25 µl reaction</u>
H ₂ O (Braun, Melsungen)	ad 25 µl
cDNA mix (after RT reaction)	1 µl
5'-Primer (see below)	50 ng
3'-Primer (see below)	50 ng
10x PCR reaction buffer	2.5 µl
10 mM dNTP mix	0.5 µl
50 mM MgCl ₂	0.75 µl
DMSO (5%)	1.25 µl
<i>Taq</i> polymerase (1 U/µl)	1 µl

To detect Sp2 mRNA expression in day E18.5 mouse embryos, primers against exon 5 and 6 (Sp2-Ex5-RT-fw; Sp2-Ex6-RT-rev) were used. As internal control for equal RNA amounts, primers against the cyclophilin gene (Cyclophilin-fw; Cyclophilin-rev) were utilized. Expected fragment lengths are: 228 bp for Sp2 and 130 bp for the cyclophilin gene.

Reaction temperatures and times:

1. primary denaturing	94°C	4 min
2. denaturing	94°C	30 sec
3. primer annealing	60°C	30 sec
4. cDNA amplification	72°C	30 sec
5. 35 cycles (step 2-4)		
6. final extension	72°C	7 min
7. cooling	4°C	for ever

To analyze the result, 10 µl of each reaction was loaded on an agarose gel.

Access RT-PCR System Kit (Promega, USA: Madison)

To investigate Sp2 mRNA expression in mouse embryonic stem cells, 1 µg of RNA derived from Sp3 +/- ES cells were used for RT-PCR (RNA preparation followed the RNeasy Mini Kit (Qiagen, Hilden)). These cells represent Sp2 wildtype situation, and total RNA from these cells was kindly provided by Heike Göllner. RT-PCR was performed following the Access RT-PCR System Kit user manual (Promega, USA: Madison). A reverse transcriptase from *Avian Myeloblastose Virus (AMV)* and a temperature stable DNA polymerase from *Thermus flavus (Tfl)* were utilized.

Reaction mix:

	<u>50 µl reaction</u>
H ₂ O (Braun, Melsungen)	ad 50 µl
5x Reaction buffer	10 µl
5'-Primer (MSp2-ES-RT-fw)	50 pmol
3'-Primer (MSp2-ES-RT-rev)	50 pmol
10 mM dNTP mix	1 µl
RNA (ES cells)	1 µg

To eliminate RNA secondary structures, the reaction mix was denatured at 95°C for 5 min. After cooling down on ice, the reaction mix was completed by adding the following components:

	<u>50 µl reaction</u>
25 mM MgSO ₄	2 µl
AMV Reverse transcriptase (5 U/µl)	1 µl
Tfl DNA polymerase (5 U/µl)	1 µl

Reaction temperatures and times:

Reverse transcription

1. cDNA synthesis	48°C	45 min
2. transcriptase inactivation	94°C	2 min

Polymerase chain reaction (PCR)

1. denaturing	94°C	30 sec
2. primer annealing	57°C	30 sec

3. cDNA amplification	72°C	30 sec
4. 30 cycles (step 1-3)		
5. final extension	68°C	7 min
6. cooling	4°C	for ever

To detect the expected Sp2 mRNA fragment (fragment length: 560 bp), 5 µl of each reaction were loaded on an agarose gel.

2.2.2.1.4 Northern Blot analyses

To detect Sp2 mRNA expression in mouse embryos, RNA from day E18.5 mouse embryos were prepared (LiCl/ urea method). 20 µg RNA were dissolved in 5 µl DEPC-H₂O, added to 20 µl loading buffer, incubated at 65°C for 15 min and loaded with 5 µl Orange G (see 2.1.3) on a 1% denaturing formaldehyde agarose gel. To receive adequate separation of RNA, the gel was run in 1x MOPS at 20-30 V over night. After gel electrophoresis, the gel was stained with ethidiumbromide (1 µl ethidiumbromide (25 mg/ml) in 100 ml 1x MOPS) for 30 min, photographed under UV light and washed twice with 10x SSC for each 30 min. RNA was blotted in 10x SSC at 4°C over night to a Hyobond-N⁺ nylon membrane (Amersham Biosciences, Freiburg). For details the reader is referred to Braun (PhD thesis, 2001), e.g. After blotting, RNA was UV-crosslinked to the membrane (0.3 J/cm²) and the membrane washed with 3x SSC. The washed membrane was prehybridized at 42°C for 2 hours, followed by hybridization in a 6 ml hybridization mix at 42°C over night together with the radioactive labeled DNA probe (probe: Ex 3-6 NB; end concentration: 1-2 x 10⁶ cpm/ml, see also 2.2.2.2.5). The labeled probe was directed against exon 3-6 (cDNA: nt 71-1500; see 6.2) of murine Sp2 and obtained by *Bam*HI/*Bgl*II digestion of pPac-mSp2(1-606). Before autoradiography, membrane was washed twice with 2x SSC + 0.1% SDS at room temperature for each 5 min, once at 50°C for 30 min and finally in 0.1x SSC + 0.1% SDS at 60-65°C for 1-2 hours to reduce background radiation. Autoradiography was performed for 24 hours.

Buffers and solutions:

- **Loading buffer:** 375 µl deionized formamide; 75 µl 10x MOPS; 120 µl 37% formaldehyde in 50 µl DEPC-H₂O.

- **10x MOPS pH 7.0:** 0.4 M 3-N-Morpholinopropansulfonic acid (MOPS); 0.1 M sodium acetate; 0.01 M EDTA in DEPC-H₂O.
- **Denaturing formaldehyde agarose gel (1%):** 1.5 g agarose; 15 ml 10x MOPS; 7.5 ml 37% formaldehyde in 127.5 ml DEPC-H₂O. Agarose and water were autoclaved before adding MOPS and formaldehyde.
- **(Pre-) Hybridization mix:** 500 ml formamide; 100 ml 100x Denhardt's; 50 ml 1 M Tris pH 7.5; 10 ml 10 mg/ml denatured salmon sperm DNA; 90 g dextran sulfate (in 250 ml H₂O (bidest.)); 58 g NaCl; 5 ml 20% SDS (filtered sterile); ad 1 l H₂O (bidest.). In contrast to the prehybridization mix, the hybridization mix additionally contains the labeled DNA probe.

2.2.2.2 DNA experiments

All experimental steps were performed at 4°C, unless stated otherwise. Buffers, solutions and media were prepared with bidistilled water and autoclaved or filtered sterile before use. The following standard methods were performed as described by Ausubel et al. (1987) or Sambrook et al. (1989):

- DNA digestion by restriction endonucleases,
- DNA treatment with alkaline phosphatase,
- Klenow filling of sticky DNA ends,
- agarose gel electrophoresis,
- DNA preparation from agarose gels,
- ligation of DNA fragments,
- ethanol or isopropanol precipitation,
- phenol/ chloroform extraction,
- DNA purification by cesium chloride ultra centrifugation.

Purification of DNA from agarose gels or PCR fragments followed the QIAquick user manual (Qiagen, Hilden).

Transformation of bacteria was carried out following two different protocols, either the heatshock protocol of Hanahan (1983) for standard transfections of *E. coli* DH5 α

bacteria with plasmid DNA, or the CaCl₂ method for mouse Sp2 protein expression in *E. coli* BL21DE3. For detailed information, see Stielow (diploma work, 2003).

To test recombinase capacity of loxP sites in the conditional Sp2 knockout construct, *E. coli* 294-Cre bacteria were transformed with either the final knockout construct pPNT-cSp2ko (containing three loxP sites) or the precursor plasmid pPNT-loxP2 (containing two loxP sites). Transformations were performed following Hanahan (1983). After incubation of transformed bacteria at 37°C over night, plasmid DNA was isolated, linearized by *NotI* or *HindIII* restriction digestion and analyzed by agarose gel electrophoresis. For further details, the reader is referred to 2.1.12 or Buchholz et al. (1996).

2.2.2.2.1 DNA isolation and purification

Isolation and purification of recombinant DNA from bacteria

Analytical (small scale) and preparative (large scale) isolation of recombinant DNA from bacteria (*E. coli*) followed the NucleoBond® AX plasmid purification manual (Macherey and Nagel, Düren). The protocol for analytical DNA preparation was modified as described by Krüger (diploma work, 2002).

Isolation and purification of mouse genomic DNA

Yolk sac or mouse tail tip DNA was used for PCR genotyping of mouse embryos or new born mice, respectively. To isolate DNA, yolk sacs or tail tips were taken up in 500 µl mouse mix buffer (MMB) with 10 µl proteinase K (10 mg/ml) and incubated for 1-2 h or over night on a 55°C shaker. After 20 min of centrifugation, the pellet was removed and 300 µl isopropanol were added to the supernatant, followed by 30 min centrifugation to spin down the precipitated DNA. The DNA pellet was washed twice with 75% ethanol and finally dissolved in 100 µl 10/1 TE. For PCR genotyping, 1 µl of DNA was used per reaction.

Mouse mix buffer (MMB): 50 ml 1 M Tris/HCl pH 7.5; 5 ml 0.5 M EDTA; 20 ml 5 M NaCl; 5 ml 2% (w/v) SDS ad 500 ml H₂O (bidest.).

2.2.2.2.2 DNA quantification

DNA was quantified by UV photometry (Amersham Pharmacia, Freiburg; or NanoDrop Technologies, USA: Wilmington). UV absorption by nuclear acids were measured at a wavelength of $\lambda = 260$ nm. At this wavelength the side chains of the aromatic amino acids display maximal absorption. The absorption value allows to calculate DNA concentration of a solution (double strand DNA: $1 \text{ OD}_{260} = 50 \mu\text{g/ml}$). In addition, DNA was also quantified by agarose gel electrophoresis.

2.2.2.2.3 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used for:

- amplification of Sp2 genomic fragments to generate a conditional knockout construct for the Sp2 gene in the mouse;
- genotyping of mouse embryos;
- generation of Sp2 DNA fragments for bacterial expression (*E. coli* BL21DE3);
- generation of Sp2 deletion fragments;
- generation of DNA probes to screen a cosmid library at the RZPD for Sp2 genomic fragments (see also 2.1.11);
- generation of DNA probes for Southern Blot analyses.

Polymerase chain reaction (PCR) followed three different protocols:

puRe Taq Ready To Go™ Kit (Amersham Biosciences, Freiburg)

The Ready To Go™ system was used for the amplification of Sp2 genomic fragments, Sp2 DNA fragments for bacterial expression, DNA probes for cosmid screening as well as Southern Blot analyses. Per reaction, 50 pg of DNA, 15 pmol of each primer and H₂O (Braun, Melsungen) ad 25 μl were added to the beads. For more detailed information, the reader is referred to the Ready To Go™ user manual.

Reaction temperatures and times:

1. primary denaturing	94°C	4 min
2. denaturing	94°C	30 sec
3. primer annealing	55-60°C*	30 sec
4. DNA amplification	72°C	30 sec
5. 30 cycles (step 2-4)		
6. final extension	72°C	7 min
7. cooling	4°C	for ever

* Annealing temperatures vary from 55-60°C depending on primer length and the relation of G/C to A/T contents in the primers. The annealing temperature was determined by the following equation: $T_m = 4 \times (G+C) + 2 \times (A+T)$.

<u>primer</u>	<u>fragment</u>	<u>T_m</u>	<u>PCR fragment length</u>
<u>Sp2 genomic fragments:</u>			
MSp2-In1-fw and -rev	genomic fragment	60°C	3.2 kb
MSp2-In2-fw and -rev	genomic fragment	60°C	1.8 kb
MSp2-Ex4-fw and -rev	genomic fragment	60°C	1.6 kb
<u>Sp2 DNA fragments:</u>			
Sp2- <i>Coli</i> -Expr-fw and -rev	bacterial expression	60°C	1.9 kb
<u>Sp2 DNA probes:</u>			
Sp2-ES-a-SB-fw and -rev	ES Southern Blotting	57°C	387 bp
Sp2-Ex4-RZPD-fw and -rev	RZPD screening	57°C	525 bp

RZPD cosmid clone no. MPMGc121L17390Q2 (genomic fragments; ES cell Southern Blotting) or pCMVSPORT6-mSp2cDNA (bacterial expression; RZPD screening) was taken as DNA template.

To analyze the result, 5-10 µl of each reaction were loaded on an agarose gel. Genomic fragments were completely sequenced to check exact amplification.

Invitrogen PCR Kit (Invitrogen, Karlsruhe)

To genotype targeted Sp2 mouse embryos, PCR analyses were performed. All reaction components were used from Invitrogen (Karlsruhe).

	<u>25 µl reaction</u>
H ₂ O (bidest.)	ad 25 µl
DNA (isolated from yolk sacs or tail tips)	1 µl
5'-Primer (see below)	50 ng
3'-Primer (see below)	50 ng
10x PCR reaction buffer	2.5 µl
10 mM dNTP mix	0.5 µl
50 mM MgCl ₂	0.75 µl
DMSO (5%)	1.25 µl
<i>Taq</i> polymerase (1 U/µl)	1 µl

Reaction temperatures and times were as indicated above. Primers annealed at 55°C. The primers MSp2Ex/L3Seq3 and MSp2In2Seq4 allowed to detect Sp2 wildtype allele (wt) by a 0.76 kb fragment. Primers Neo and MSp2In2Seq4 led to a 0.9-1 kb fragment representing the targeted Sp2 locus (lzn). DNA templates derived from mouse embryonic yolk sacs or tail tips (see 2.2.2.2.1). To analyze the result, 5-10 µl of each reaction were loaded on an agarose gel.

Phusion™ Kit (Finnzymes Oy, Finland: Espoo)

To generate Sp2 deletion mutants, the DNA fragments were amplified by PCR following the Phusion™ Kit user manual. The kit is based on a proof-reading DNA polymerase with an error rate of 4.4×10^{-7} (according to the manufacturer). For detailed information including the composition of the reaction mix, reaction times and temperatures, the reader is referred to the manual. Annealing temperatures of the primers were 58°C. Plasmid pPac-mSp2 served as DNA template (50 pg per reaction).

<u>primer</u>	<u>PCR fragment length</u>
Del1-1-Sp2-fw and Del1-Sp2-rev-sh	1.89 kb
Del1-2-Sp2-fw and Del1-Sp2-rev-sh	1.74 kb
Del1-3-Sp2-fw and Del1-Sp2-rev-sh	1.60 kb
Del1-4-Sp2-fw and Del1-Sp2-rev-sh	1.54 kb
Del2-Sp2-fw and Del1-Sp2-rev-sh	1.48 kb
Del3-Sp2-fw and Del1-Sp2-rev-sh	1.05 kb
Del4-Sp2-fw and Del1-Sp2-rev-sh	0.78 kb
Sp2FL- <i>Xho</i> I-fw and Del1-2- <i>Bam</i> HI-rev	0.37 kb
Del2- <i>Bam</i> HI-fw and Del1- <i>Bgl</i> II-rev-sh	1.45 kb
Gal4-Sp2-pLALL-fw and Gal4-Sp2-pE-rev	1.39 kb
Gal4-Sp2-pLALL-fw and Gal4-Sp2-mE-rev	1.06 kb
Gal4-Sp2-pLALL-fw and Gal4-Sp2-mG-rev	0.63 kb
Del2- <i>Bam</i> HI-fw and Gal4-Sp2-pE-rev	0.79 kb

To analyze the result, 5-10 μ l of each reaction were loaded on an agarose gel. Sp2 deletion fragments were completely sequenced to check exact amplification.

2.2.2.2.4 Hybridization and purification of single-stranded oligonucleotides

Hybridization of loxP site oligonucleotides

To generate loxP sites, single-stranded loxP site oligonucleotides (see 2.1.9.4) were hybridized in equimolar amounts at 75°C for 20 min followed by cooling-down slowly to room temperature.

Hybridization of oligonucleotides for Electrophoretic Mobility Shift Assays

Equimolar amounts of single-stranded oligonucleotides (5-10 μ g) were hybridized in 10/1 TE ad 30 μ l at 75°C for 20 min and slowly cooled down to room temperature over night. Hybridized oligonucleotides were separated from unhybridized ones by polyacrylamide-TBE gel electrophoresis on a 12% gel. After gel separation, the gel was transferred to a silica plate (Kieselgel 60 F₂₅₄, Merck, Darmstadt) and the silhouette of the hybridized oligonucleotides was detected under UV light ($\lambda = 254$

nm). Oligonucleotides were cut out and eluted from the gel slices by grinding and incubation of the slices in each 500 µl 10/1 TE at 37°C and 250 rpm over night. Eluted DNA was separated from polyacrylamide by centrifugation over glass wool and purified by NAP-10 columns (Amersham Biosciences, Freiburg), followed by ethanol precipitation. Finally, the quantified DNA was radioactively labeled as described in 2.2.2.2.5 and used for Electrophoretic Mobility Shift Assays (EMSAs).

Polyacrylamide-TBE gel (12%): 120 ml 5x TBE; 40 ml 30% acrylamide/ bisacrylamide (37.5 / 1); 40 ml H₂O (bidest.); 1 ml 10% APS; 120 µl TEMED. Running buffer: 1x TBE. Electrophoresis: 80 V for 30 min; 130 V for 5-6 hours.

2.2.2.2.5 Radioactive labeling of DNA

Radioactive labeling of double-stranded oligonucleotides for EMSA

To radioactively label double-stranded oligonucleotides for Electrophoretic Mobility Shift Assays (EMSAs), 5' sticky ends were filled by the Klenow fragment of the DNA polymerase I using the Megaprime DNA Labelling System (Amersham Biosciences, Freiburg).

Reaction mix:

	<u>20 µl reaction</u>
oligonucleotides	20 ng
10x restriction enzyme buffer B (Roche, Mannheim)	2 µl
dATP (Megaprime Kit; Amersham Biosciences)	1 µl
dGTP (Megaprime Kit; Amersham Biosciences)	1 µl
dTTP (Megaprime Kit; Amersham Biosciences)	1 µl
[α- ³² P] dCTP (3000 Ci/ mmol)	2µl
Klenow enzyme (1 U/µl)	1 µl
H ₂ O (Braun, Melsungen)	ad 20 µl

The reaction was incubated at room temperature for 20 min, followed by phenol/ chloroform extraction. The extracted oligonucleotides were purified by ProbeQuant™ G-50 Micro Columns (Amersham Biosciences, Freiburg) and diluted to a concentration of 18000 cpm/µl (Scintillator; Beckman Coulter, Krefeld).

Oligonucleotides used for radioactive labeling and Electrophoretic Mobility Shift Assays are listed in 2.1.9.2.

Radioactive labeling of single-stranded DNA

The single-stranded DNA probe for Northern Blot analyses was labeled by the Nick Translation Kit of Amersham Biosciences (Freiburg): 100 ng DNA together with 4 μ l dNTP mix (dCTP, dGTP and dTTP), 2 μ l of fresh [α - 32 P] dATP (3000 Ci/mmol), 2 μ l enzyme mix (1 U/ μ l) and H₂O (bidest.) ad 20 μ l were incubated at 37°C for 30 min. In contrast to that, single-stranded DNA probes for Southern Blot analyses were labeled following the Megaprime Kit of Amersham Biosciences (Freiburg). Here, 20 ng DNA together with 2.5 μ l primer mix and 10/1 TE ad 14 μ l were denatured at 100°C for 5 min, followed by immediately cooling-down to 4°C. 4 μ l of fresh [α - 32 P] dCTP (3000 Ci/mmol), 5 μ l of 5x buffer containing the other nucleotides and 2 μ l Klenow enzyme (1 U/ μ l) were added to the reaction mix and incubated first at 37°C for 15 min, then at room temperature for at least 45 min. To remove free nucleotides, the reaction mixes were purified by ProbeQuant™ G-50 Micro Columns (Amersham Biosciences, Freiburg).

Probes used for Northern (NB) and Southern Blot (SB) analyses were as follows:

<u>probe</u>	<u>function</u>	<u>hybridization (see 6.2)</u>
Ex 3-6 NB	NB probe against Sp2 exon 3-6	cDNA: nt 71-1500
Ex4-RZPD	SB probe against exon 4 to verify RZPD cosmids	cDNA: nt 533-1057
ES-a-SB	SB probe to detect targeted Sp2 allele in ES cells	DNA: nt 34808-35195

Oligonucleotides to amplify the probes are listed in 2.1.9.3. The Northern Blot probe against exon 3-6 (Ex 3-6 NB) was obtained by *Bam*HI/*Bgl*II digestion of pPac-mSp2(1-606).

2.2.2.2.6 Southern Blot analyses

Southern Blot analyses (Southern, 1975) were performed to verify results obtained from screening a cosmid library at the RZPD Deutsches Ressourcenzentrum für Genomforschung GmbH (Heidelberg) for the Sp2 gene as well as to genotype targeted Sp2 embryonic stem cells and mouse embryos. For cosmid check, 1 µg of each cosmid DNA was digested with *Bgl*II, *Hha*I or *Eco*RI in a total volume of 20 µl. For mouse ES cell and embryo genotyping, 0.5 µg DNA were digested with *Sac*I in a total volume of 50 µl. 10 µl (cosmid DNA) or 50 µl (ES cells, embryos) of each reaction mix were separated by agarose gel electrophoresis for 4-5 hours. After staining the gel for 30 min with ethidiumbromide (1 µl ethidiumbromide (10 mg/ml) in 100 ml H₂O), it was destained with H₂O for 15 min and photographed under UV light. The gel was incubated in a denaturing solution for 30 min, neutralized for 30 min and blotted in 20x SSC over night to a Hyobond-N⁺ nylon membrane (Amersham Biosciences, Freiburg). For methodological details, the reader is referred to Krüger (diploma work, 2002), e.g. After blotting, DNA was UV-crosslinked to the membrane (0.3 J/cm²; Biometra Crosslinker) and the membrane incubated at 80°C for 1.5 hours. The membrane was shortly washed in 5x SSC and prehybridized at 42°C for 2 hours, followed by hybridization at 42°C for 4-5 hours (for cosmid clones) or over night (for mouse ES cell and embryo DNA) together with the denatured, radioactive labeled DNA probe (end concentration: 1-2 x 10⁶ cpm/ml, see also 2.1.9.3 and 2.2.2.2.5). The labeled probes used for Southern Blot analyses are listed below. Previous to autoradiography, the membrane was washed twice with 2x SSC + 0.1% SDS at room temperature for each 1 min, once at 50°C for 30 min and finally in 0.1x SSC + 0.1% SDS at 60°C for 45 min to reduce background radiation. Autoradiography was performed by film exposure (Kodak, Stuttgart) to the membrane at -80°C over night.

<u>probe</u>	<u>function</u>	<u>hybridization (see 6.2)</u>
Ex4-RZPD	SB probe against exon 4 to verify RZPD cosmids	cDNA: nt 533-1057
ES-a-SB	SB probe to detect targeted Sp2 allele in ES cells and mouse embryos	DNA: nt 34808-35195

Resulting fragments after restriction endonuclease digestions are as follows:

- Southern Blot analysis of RZPD cosmids: *Bgl*II: 5.8 kb; *Hha*I: 2.8 kb; *Eco*RI: 1.6 kb;

- Southern Blot analysis of targeted Sp2 embryonic stem cells and mouse embryos: Sac I: 6.9 kb representing wildtype allele, 3.2 kb representing targeted allele.

Solutions:

- **Denaturing solution:** 0.5 M NaOH; 1.5 M NaCl in H₂O (bidest.).
- **Neutralization solution pH 7.4:** 1 M Tris; 1.5 M NaCl in H₂O (bidest.).
- **Prehybridization solution:** 5 ml formamide; 2.5 ml 20x SSC; 1 ml 50x Denhardt's; 250 µl of 10 mg/ml denatured herring sperm DNA; 500 µl 1M sodium phosphate pH 6.4; 100 µl 10% SDS (filtered sterile); H₂O (bidest.) ad 10 ml.
- **Hybridization solution:** 5 ml formamide; 2 ml 20x SSC; 0.5 ml 50x Denhardt's; 100 µl of 10 mg/ml denatured herring sperm DNA; 200 µl 1 mM sodium phosphate pH 6.4; 100 µl 10% SDS (filtered sterile); 2 ml dextran sulfate; H₂O (bidest.) ad 10 ml.

2.2.3 Biochemical methods

Unless stated otherwise, all experimental steps were performed at 4°C to prevent proteins from degradation.

2.2.3.1 Recombinant Sp2 protein expression and purification from bacterial inclusion bodies

Large-scale overexpressed proteins in bacteria accumulate in so-called insoluble inclusion bodies. To generate Sp2-specific polyclonal antibodies in rabbits, mouse Sp2 protein was expressed in four 1 l cultures of *E. coli* BL21DE3 bacteria and purified from bacterial inclusion bodies following a modified protocol from Nagai et al. (1985). Bacterial transformation by CaCl₂ and IPTG-induced protein expression followed the protocols described in Stielow (diploma work, 2003). In the following description of inclusion body isolation and Sp2 protein purification from bacterial inclusion bodies, all amounts and volumes are referred to a 1 l culture.

Two hours after IPTG induction (0.5 mM IPTG), 1 l of cultured bacteria (*E. coli* BL21DE3) was harvested and spun down at 4000 rpm for 10 min. Bacterial pellets were resuspended in 40 ml resuspension buffer and incubated at 4°C for 30 min after adding 2.5 mg/ml lysozyme. To digest DNA, 10 µg/ml DNase I together with 1 mM MnCl₂ and 10 mM MgCl₂ were added and the suspension was incubated at room temperature on a head-over-tail rotator for 1 hour. After adding 100 ml washing buffer 1, the suspension was centrifuged at 13000 rpm and 4°C for 12 min to pellet all insoluble components, like the inclusion bodies. The inclusion body pellet was dissolved in 20 ml washing buffer 1, centrifuged at 4000 rpm and 4°C for 12 min and resuspended in 20 ml washing buffer 2. The pellet was centrifuged again, washed up to 12 times in washing buffer 2 until the pellet did not decrease in size any more and dissolved in 2 ml 2x Laemmli protein loading buffer (see 2.1.2). To purify Sp2 protein from inclusion bodies, the suspension was incubated at 100°C for 10 min and separated by a discontinual, preparative SDS Polyacrylamide Gel Electrophoresis (10% SDS-PAGE; see also 2.2.3.4, 20x volume) at 80 V over night. To detect Sp2 protein, the gel was stained with 300 mM CuCl₂ up to 10 min and repeatedly washed in H₂O. The Sp2 protein band was cut out from gel, washed three times in washing buffer 3 for each 10 min to remove copper ions, and incubated three times in elution buffer for each 10 min. Sp2 protein was eluted by electro-elution at 100 V for 6 hours in a gel chamber (dialysis tubes: Serva, Heidelberg) by changing the running buffer each 45 min. The protein solution was dialyzed against 0.5x PBS at room temperature for 5 hours changing buffer four times. To enrich proteins, the solution was lyophilized over night and the lyophilisate resuspended in 5-10 ml H₂O (Braun, Melsungen). After determination of protein concentration by comparing increasing Sp2 volumes with a BSA standard of known concentration on a 10% SDS polyacrylamide gel, the purified Sp2 protein was used for rabbit immunization (see 2.2.3.6).

Buffers:

- **Resuspension buffer:** 50 mM Tris/HCl pH 8.0; 25% (w/v) succrose; 1 mM EDTA.
- **Washing buffer 1:** 1% (w/v) deoxycholate; 20 mM Tris/HCl pH 7.5; 2 mM EDTA; 1.6% (v/v) NP-40; 200 mM NaCl. The components should be added in this order.
- **Washing buffer 2:** 50 mM Tris/HCl pH 7.5; 150 mM NaCl; 0.5% (v/v) EDTA.
- **Washing buffer 3:** 250 mM Tris/HCl pH 8.8; 250 mM EDTA.

- **Elution buffer:** 250 mM Tris/acetate pH 7.4; 1% (w/v) SDS; 100 mM DTT.
- **Running buffer:** 50 mM Tris/acetate pH 7.4; 0.1% (w/v) SDS.

2.2.3.2 Protein extraction from cells (nuclear extract preparation)

Nuclear proteins derived from cultured cells were prepared for Western Blot analyses or Electrophoretic Mobility Shift Assays (EMSAs).

To isolate nuclear proteins from cultured cells (6 cm culture plates), cells were washed twice in 1x PBS, harvested with a rubber spatula in 1.4 ml 1x PBS and spun down at 13000 rpm for 10 sec. Cell pellets were dissolved in 400 μ l of hypotonic NE buffer B. After 10 min of incubation at 4°C, cells were spun down for 10 sec and the pellets containing the nuclei dissolved in 50 μ l hypertonic NE buffer C. Nuclei were incubated in NE buffer C at 4°C for 20 min, spun down for 2 min, and the supernatant containing the nuclear proteins was split into aliquots. Protein concentration of the solutions was determined by Bradford assay (see below) and the aliquots were stored at -80°C until use.

Buffers:

- **NE buffer B:** 10 mM HEPES/KOH pH 7.9; 1.5 mM MgCl₂; 10 mM KCl; 0.5 mM DTT; 0.2 mM PMSF; 1x PIC.
- **NE buffer C:** 20 mM HEPES/KOH pH 7.9; 1.5 mM MgCl₂; 420 mM NaCl; 0.2 mM EDTA; 25% Glycerol; 0.5 mM DTT; 0.2 mM PMSF; 1x PIC.

2.2.3.3 Determination of protein concentrations

Quantification after Bradford

Photometric determination of protein concentrations after Bradford (1976) is based on a shift of the absorption maximum from $\lambda = 465$ nm to $\lambda = 595$ nm of Coomassie brilliant blue when binding to proteins. This shift can be detected photometrically at a wavelength of $\lambda = 595$ nm: the increase in extinction is proportional to the protein concentration of the measured solution. 5 μ l of bovine serum albumin (BSA) in a concentration of 0.1 to 1.0 mg/ml were used to generate a standard curve. To

measure protein concentrations in a linear range, the measured sample volumes (1-5 μ l) were chosen in that way, that their extinction values did not exceed OD = 1.0. Preparation of the staining solution and measuring procedure followed the instructions of the Biorad Protein Assay Kit (Biorad, München).

Quantification after Lowry

Protein concentrations of solutions containing SDS or other detergences could not be determined by Bradford quantification. In that case, a modified method of Lowry (1951), basis of the Biorad Detergent Compatible DC Kit (Biorad, München), was applied following the kit instructions. 20 μ l of the BSA standards and extinction-dependent sample volumes were used for photometric measurements at a wavelength of $\lambda = 750$ nm.

2.2.3.4 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Based on their molecular weights, proteins were separated by discontinual, denaturing SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) (Laemmli, 1970). Samples were denatured in 1 volume of 2x Laemmli protein loading buffer (see 2.1.2) at 100°C for 5 min before loading on a gel (Mini gel chambers; Biorad, München). Gel consists of two parts, one focussing the proteins (pH 6.8), the other one separating the proteins (pH 8.8). Percentage of the separating gel section ranged from 6-12% depending on the molecular weight of the proteins of interest. Proteins were focussed at 80 V until they reached the separating section, and separated at 120V. After electrophoresis, proteins were stained (Coomassie Brilliant Blue; CuCl_2) or analyzed in Western Blot. For deeper detailed information, the reader is referred to Stielow (diploma work, 2003). Protein staining by CuCl_2 is described in 2.2.3.1. For Coomassie staining, the reader is referred to the Roti-Blue® user manual (Roth, Karlsruhe).

Focussing gel (5%)

	<u>end volume: 3 ml</u>
H ₂ O (Braun, Melsungen)	ad 3 ml
30% acrylamide / bisacrylamide (37.5 / 1)	500 μ l
focussing buffer (0.5 M Tris/HCl pH 6.8)	760 μ l

10 % sodium dodecyl sulfate (SDS)	30 µl
10% ammonium peroxy disulfate (APS)	30 µl
N'-tetramethylethylenediamine (TEMED)	3 µl

Separating gel (6-12%)

	<u>end volume: 7.5 ml</u>
H ₂ O (Braun, Melsungen)	ad 7.5 ml
30% acrylamide / bisacrylamide (37.5 / 1)	1.5 to 3.0 ml
separating buffer (1.5 M Tris/HCl pH 8.8)	1.86 ml
10 % sodium dodecyl sulfate (SDS)	75 µl
10% ammonium peroxydisulfate (APS)	75 µl
N'-tetramethylethylenediamine (TEMED)	10 µl

Solutions:

- **Running buffer:** 25 mM Tris/HCl pH 8.3; 250 mM Glycine; 0.1% (w/v) SDS.
- **Molecular weight standards:** Rainbow® marker (Amersham Biosciences, Freiburg); Prestained protein ladder (MPI Fermentas, St. Leon-Rot); Protein ladder (Biorad, München).

2.2.3.5 Western Blot analyses

To selectively detect proteins in protein solutions, Western Blot analyses were performed. For this, SDS-PAGE-separated proteins (2.2.3.4) were blotted to a polyvinylidene difluoride membrane (PVDF; Immobilon-P transfer membrane; Millipore, Eschborn) at 2.5 mA/cm² for 75 min using a Biometra semi-dry blotting system (Biometra, Göttingen). For details, the reader is referred to Göllner (PhD thesis, 2002), e.g. The immobilized proteins were blocked in a blocking solution at 4°C over night. Immune detection was performed with first and secondary antibodies listed below at room temperature for each one hour. Detection of the antibody-bound proteins was effected by the conjugation of the secondary antibody with horse reddish peroxidase (HRP) and the Enhanced Chemiluminescence Plus (ECL-Plus) system (Amersham Biosciences, Freiburg). To remove unspecifically bound

antibodies, thus reducing background signals, membrane was stringently washed after first and secondary antibody incubation for each 30 min in TBST.

<u>Antibody</u>	<u>dilutions for Western Blot analyses*</u>
anti-Sp1; rabbit p.cl. antibody	1/3000
anti-Sp2; rabbit p.cl. antibody (no. 193, "Zwick"; 10 days after 1. boost)	unless stated otherwise 1/5000
anti-Sp3; rabbit p.cl. antibody; sc-644	1/1000
anti-Gal4; rabbit p.cl. antibody; sc-577	1/1000
anti-rabbit; HRP-conj.; m.cl. sec. antib.; NA934V	1/10000

* = Antibodies were diluted in a 1% blocking solution. Membrane was incubated in 10 ml antibody solution.

Solutions:

- **1x TBS:** 20 mM Tris/HCl pH 7.6; 137 mM NaCl in H₂O (bidest.).
- **1x TBST:** 1x TBS; 0.1% (v/v) Tween 20.
- **5% blocking solution:** 20 mM Tris/HCl pH 7.6; 137 mM NaCl; 0.1% (v/v) Tween 20; 5% (w/v) Difco™ Skim milk powder (BD, USA: Sparks).
- **Anode buffer 1:** 0.3 M Tris/HCl pH 10.4; 10% (v/v) methanol.
- **Anode buffer 2:** 25 mM Tris/HCl pH 10.4; 10% (v/v) methanol.
- **Cathode buffer:** 25 mM Tris/HCl pH 9.4; 40 mM Glycine 10% (v/v) methanol.

2.2.3.6 Rabbit immunization and generation of antiserum

To generate polyclonal Sp2 antisera, two New Zealand White rabbits (no. 193, "Zwick" and no. 194, "Zwack") were immunized with purified mouse Sp2 protein expressed in *E. coli* BL21DE3 bacteria (2.2.3.1). Per each immunization reaction, 1ml of protein solution (= 75 µg Sp2) together with 1 ml Freund's Adjuvant were subcutaneously injected into 3-4 different areas at the flank. In the first three immunization reactions the protein solution was mixed with complete Freund's Adjuvant (CFA; Sigma, München), all further immunizations were performed with incomplete Freund's Adjuvant (IFA; Sigma, München). In contrast to incomplete Freund's Adjuvant, complete adjuvant contains heat-inactivated, dried

Mycobacterium tuberculosis bacteria, which unspecifically stimulate the rabbit immune system.

Immunization reactions were as follows: Day 1, 3 and 6 with each 75 µg Sp2 in CFA; day 36 (1. boost) and 70 (2. boost) with each 75 µg Sp2 in IFA. Before starting the first immunization reaction, 5 ml blood were taken from both rabbits for pre-immune serum extraction. As well, each 10 days after boost immunizations, 5 ml blood were taken to extract the Sp2 antiserum. Rabbit immunization was kindly performed by Bastian Stielow.

To extract pre-immune or Sp2 antisera, blood coagulated at room temperature for 30 min followed by 10 min centrifugation at 4000 rpm. The separated sera were transferred to fresh reaction tubes, mixed for conservation with 0.05% sodium azide and stored as aliquots at -20°C. Fresh aliquots were used for Western Blot analyses as well as immunostaining and Electrophoretic Mobility Shift Assays (EMSAs).

2.2.3.7 Electrophoretic Mobility Shift Assay (EMSA)

To investigate protein-DNA interactions, Electrophoretic Mobility Shift Assays (Fried and Crothers, 1981; Garner and Revzin, 1986) were performed. For this, nuclear extracts containing the protein of interest (either endogenous or overexpressed) were incubated with a radioactively labeled double-stranded oligonucleotide containing a specific sequence of a protein binding site. The oligonucleotides used to investigate Sp2-DNA interactions are listed in 2.1.9.2; the labeling protocol is described in 2.2.2.2.5. Protein-bound oligonucleotides migrate slower in gel (shift) than unbound oligonucleotides. Additional incubation with an antibody directed against the protein of interest increases the slower migration of the protein-DNA complex (supershift).

A reaction mix consisting of 0.2-10 µg nuclear extract, 1x BSB, 0.5 mg/ml BSA, 37.5 ng/µl poly (dIdC), 4 mM DTT and 145 mM KCl in a total volume of 20 µl was incubated at 4°C for 10 min. After incubation, the labeled oligonucleotide (18000 cpm) was added to the reaction mix and incubated at 4°C for 20 min. In case of supershift experiments, this incubation was reduced to 15 min followed by incubation with 1 µl antibody against the protein of interest (alternatively: pre-immune serum) at

4°C for 15 min. Protein-bound oligonucleotide was separated from unbound oligonucleotide by a 4% non-denaturing polyacrylamide gel (100 V for 30 min, 140 V for ca. 2.5 h). After electrophoresis, gel was transferred to a Whatman filter paper (3 mm) and vacuum-dried at 80°C for 2 hours (Gel dryer; Biorad, München) followed by autoradiography at -80°C over night.

Materials:

- **5x Bandshift buffer BSB (GH):** 62.5 mM HEPES pH 7.5; 31.25 mM MgCl₂; 45% Glycerol; 25 μM ZnSO₄; 0.25% NP-40.
- **4% Polyacrylamide gel (PAA):** 10 ml 40% acrylamide/bisacrylamide (40:1); 10 ml 5x TBE; 80 ml H₂O (Braun, Melsungen); 700 μl 10% APS; 70 μl TEMED. Running buffer: 0.5x TBE.

2.2.4 Cellbiological methods and animal works

2.2.4.1 Cell handling

Cells were handled sterile. Frozen cells (liquid nitrogen) were transferred to 10 ml of the required culture medium, centrifuged at 1000 rpm for 5 min to remove DMSO, transferred again to 12 ml of fresh medium and plated for culturing (cell culture flasks; Greiner, Frickenhausen). Drosophila SL2 cells were cultured at 27°C, mammalian cells at 37°C in a 5% CO₂-containing atmosphere. When reaching 80-100% confluency, cells were split and used either for further culturing or were frozen for storage in 1.5 ml medium containing 10% DMSO, first at -80°C for two days, then in liquid nitrogen (Cryo Tube™ Vials; Nunc, Denmark: Roskilde). Splitting of Drosophila SL2 cells occurred after detaching cells from the culture plate by intensive slapping against the plate. Mammalian cells (MEF, HEK-293, Ishikawa) were detached from culture plate by 3 min trypsination (1 ml Trypsin/EDTA; Gibco, Karlsruhe) after washing with 1x PBS and removing the medium. Split cells were dissolved in 12 ml medium and plated again for culturing. Cells having been split for more than two times were frozen in liquid nitrogen for at least some weeks to recover. For mouse embryonic stem (ES) cell handling, see 2.2.4.5.

2.2.4.2 Cell immunostaining

To investigate Sp2 subcellular localization, human embryonic kidney 293 cells (HEK-293) were grown in a 24 well culture plate (Greiner, Frickenhausen) on 13 mm cover slips until 80-90% confluency, washed twice in 1x PBS and fixed at room temperature in a 4% paraformaldehyde (PFA) solution. After fixation cells were washed twice with PBS and permeabilized with 0.2% Triton X-100 (Serva, Heidelberg) in PBS for 25 min. Permeabilized cells were washed again three times in 1x PBS, blocked in a 3% BSA blocking solution for 1 hour followed by first antibody and secondary antibody incubation each for one hour (150 µl per cover slip). To additionally reduce unspecific binding of the antibodies, cells were washed after each antibody incubation three times in 1x PBS each for 10 min. Mounting of cells to glass slides occurred with one drop of Vecta-shield mounting medium containing DAPI-4,6-diamino-2-phenylindole (Vector Laboratories, USA: Burlingame). Pictures were taken by a camera mounted on a Leica DMLB microscope (Leica, Bensheim) using filters for green (ex.: $\lambda = 480$ nm; em.: $\lambda = 520$ nm) and red fluorescence (ex.: $\lambda = 550$ nm; em.: $\lambda = 600$ nm).

<u>antibody</u>	<u>dilutions for immunostaining*</u>
anti-Sp1; rabbit p.cl. antibody	1/1000
anti-Sp2; rabbit p.cl. antibody (no. 193, "Zwick"; 10 days after 1. boost)	1/1000
anti-Sp3; rabbit p.cl. antibody; sc-644	1/1000
anti-rabbit; FITC-conj.; m.cl. sec.; 111-095-003	1/300

* = Antibodies were diluted in a 3% BSA blocking solution. Cells were incubated in 400 µl antibody solution.

3% BSA blocking solution: 3% BSA; 1x PBS; 0.05% (v/v) Tween; filter sterilized.

2.2.4.3 Cell transfections

To overexpress proteins in insect (*Drosophila* SL2) or mammalian cells (HEK-293), cells were transiently transfected with the respective expression plasmids listed in 2.1.10 following three different protocols:

Transfection by FuGENE 6 (Roche, Mannheim)

Standard transfections were performed by FuGENE 6 (Roche, Mannheim). *Drosophila* SL2 or HEK-293 cells (4×10^6 cells per 6 cm culture plate) were transfected with 1 μ g of cesium-chloride-purified DNA in 5 μ l FuGENE 6 reagent or with 1 μ g of cesium-chloride-purified DNA in 3 μ l FuGENE 6 reagent, respectively. For further details concerning the protocol, the reader is referred to the user manual. To analyze overexpressed proteins in Western Blot or Electrophoretic Mobility Shift Assays (EMSAs), cells were harvested for protein extraction 36 hours after transfection.

Transfection by jetPEI™ (Biomol, Hamburg)

To overexpress Gal4 fusion proteins in HEK-293 cells for transactivation assays measuring luciferase activity (Luc; see 2.2.4.4), cells were transiently transfected following the jetPEI™ transfection protocol (Biomol, Hamburg). Cesium-chloride-purified DNA was used to enhance the transfection efficiency. Transfection mixes for 4×10^6 cells per 6 cm culture plate were as follows:

- HEK-293 cells: 1 μ g Gal4 fusion plasmid, 3 μ g luciferase reporter plasmid pGAWG5E1b and 0.5 μ g control plasmid for transfection efficiency comparisons pRSV- β -Gal in 9 μ l jetPEI™.
- HEK-293-pGAWG5E1b-(stable) cells: 1 μ g Gal4 fusion plasmid and 0.5 μ g control plasmid for transfection efficiency comparisons pRSV- β -Gal in 3 μ l jetPEI™.

Cells were harvested for transactivation assays 48 hours after transfection.

Transfection by Calcium-Phosphate

To overexpress proteins in *Drosophila* SL2 cells for transactivation assays (Luc, CAT; see 2.2.4.4), cells were transiently transfected by the Calcium-phosphate method. Cesium-chloride-purified DNA was used to enhance the transfection efficiency. For both transactivation assays (Luc, CAT), the transfection procedure for 4×10^6 cells per 6 cm culture plate was as follows: 0.02-0.5 μ g Sp expression plasmid, 4 μ g luciferase (pGL3-Prom) or chloramphenicol acetyl transferase (pBCAT-2) reporter plasmid together with 2 μ g p97b control plasmid for transfection efficiency comparisons were mixed with 250 μ l sterile CaCl₂ solution. The solution was added drop-wise within 30 sec to 250 μ l 2x HeBS while continuously vortexing the HeBS

solution. Vortexing, i.e. supplying oxygen to the transfection mix, enables formation of calcium-DNA precipitates. After incubation at room temperature for 30 min, mix was distributed drop-wise and equally to the cells. Harvesting of cells for transactivation assays occurred 48 hours after transfection.

Solutions:

- **CaCl₂ solution:** 250 mM CaCl₂ in 1 mM HEPES pH 7.1; filter sterilized.
- **250 ml 2x HeBS:** 4 g NaCl; 0.175 g KCl; 0.125 g Na₂HPO₄; 0.5 g Dextrose; 2.5 g HEPES pH 7.1; filter sterilized.

2.2.4.4 Transactivation assays

Sp2 transactivation properties were investigated by measuring either Sp2-mediated luciferase (Luc) activity or chloramphenicol acetyl transferase (CAT) activity. To consider variations in transfection efficiency and to calculate relative Luc and CAT activity values, β -galactosidase activity was determined supplementary.

Chloramphenicol acetyl transferase activity assay (CAT assay)

Investigation of Sp2 transactivation properties by measuring CAT activity was performed following the CAT ELISA instruction manual (Roche, Mannheim). Transfection procedure and amounts of transfected DNA are indicated in 2.2.4.3.

Luciferase activity assay (Luc assay)

To determine luciferase activity of transient transfected cells (6 cm culture plates), cells were washed twice in 1x PBS, harvested with a rubber spatula in 1.3 ml 1x PBS and spun down at 13000 rpm for 10 sec. Cell pellets were dissolved in 100 μ l Luc resuspension buffer and frozen in liquid nitrogen followed by immediately re-thawing at 37°C. Freezing and re-thawing was repeated three times to lyse cells. After lysis, cells were centrifuged at 4°C for 10 min and 50 μ l of the supernatant were added to 360 μ l luciferase buffer. Luciferase activity of the cell suspension was measured by light emission (10 sec measuring) by injecting 100 μ l D-luciferin solution to each sample (AutoLumat LB953; Berthold Technologies, Bad Wildbad).

Solutions:

- **Luc resuspension buffer:** 120 µl 100 mM Potassium phosphate; 1.2 µl 1 M DTT; 1.08 ml H₂O (Braun, Melsungen).
- **Luciferase buffer (per reaction):** 180 µl 50 mM Glycylglycine pH 7.8; 5.4 µl 1 M Potassium phosphate pH 7.8; 5.4 µl 1 M MgSO₄; 28.8 µl 50 mM EGTA; 7.2 µl 0.1 M ATP; 3.6 µl 0.1 M DTT; 133.2 µl H₂O (Braun, Melsungen).
- **Luciferin solution:** 900 µl 1 mM D-Luciferin (Sigma, München); 2.25 ml 50 mM Glycylglycine; 9 µl 1 M DTT; 1.35 ml H₂O (Braun, Melsungen).

β-Galactosidase activity assay (β-Gal assay)

To consider variations in transfection efficiency and to calculate relative Luc and CAT activity values, β-galactosidase activity was determined (Hall et al., 1983). Cells treated as described for Luc and CAT assays were mixed with 250 µl buffer Z and 60 µl O-Nitrophenyl-β-D-galactopyranoside (ONPG) and incubated at 30°C until the solution became yellow. Depending on assay and cell line, volumes of added cells were as follows: 80 µl of cells derived from CAT assay preparation, 60 µl of cells derived from Luc assay preparation, whereas in both cases only 5 µl of HEK-293-pGAWG5E1b-(stable) cells were used. Yellow staining of the solution is based on the enzymatic cleavage of ONPG to O-Nitrophenol by β-galactosidase. To stop the enzymatic reaction, 100 µl 1 M Na₂CO₃ was added and the suspension was photometrically measured at λ = 420 nm (Emax Precision Microplate Reader; Molecular Devices, München).

Solutions:

- **Buffer Z pH 7.0:** 60 mM NaHPO₄; 40 mM NaH₂PO₄; 10 mM KCl; 1 mM MgSO₄; 50 mM β-Mercaptoethanol.
- **O-Nitrophenyl-β-D-galactopyranoside (ONPG) pH 7.0:** 4 mg/ml ONPG (Sigma, München) in 100 mM sodium phosphate buffer.

2.2.4.5 Generation of Sp2-targeted mice and mouse handling

Embryonic stem cell works

To target the Sp2 gene, mouse embryonic stem cells (C88Bl/6 TdW; EUR Rotterdam) were grown in BRL-conditioned medium + 10³ U/ml Leukaemia inhibitory

factor (LIF) on gelatinized (0.1% gelatin) culture plates (Greiner, Frickenhausen). To target embryonic stem (ES) cells with the conditional Sp2 knockout construct, cells (1×10^7 per 10 cm^2 culture plate) were transfected with $15 \mu\text{g}$ 2x cesium-chloride-purified and *NotI*-linearized DNA by electroporation (BioRad Gene Pulser; Biorad, München). After transfection, cells were cultured for ca. 10 days in a selection medium containing $200 \mu\text{g/ml}$ G418 (Gibco, Karlsruhe). Single clones were picked and cultured in selection medium until 60-70% confluency on 24-well plates (Greiner, Frickenhausen). Then, cells were split, grown again to 60-70% confluency and either stored until further use in ES cell medium + 10% DMSO at -70°C or used for DNA-isolation. To isolate DNA, cells were lysed in MMB buffer and proteinase K (10 mg/ml) at 55°C over night, followed by phenol/ chloroform extraction and isopropanol (0.6 volumes) precipitation. DNA pellet was washed with 70% ethanol and dissolved in $50\text{-}75 \mu\text{l}$ 10/1 TE. For genotyping, $0.5 \mu\text{g}$ were digested by *SacI* in a total volume of $50 \mu\text{l}$ and analyzed by Southern Blotting and PCR (see 2.2.2.2.6 and 2.2.2.2.3, respectively). All ES cell works were kindly performed by Nynke Gillemans, Southern Blot analyses by Imme Krüger at the Erasmus University of Rotterdam.

Media and buffers:

- **ES cell culture medium:** 500 ml DMEM (high Guucose + Glutamine) (Gibco, Karlsruhe); 200 ml BRL conditioned medium; 50 ml FBS (suitable for embryonic stem cells); 5 ml non-essential amino acids; 5 ml 100x Penicillin/ Streptomycin (Cambrex, Belgium: Verviers); $0.9 \text{ ml } \beta\text{-Mercaptoethanol}$ (0.1 ml in 14 ml medium); 0.9 ml Leukaemia inhibitory factor (LIF of ESGRO) 10^6 U/ml (Gibco, Karlsruhe).
- **Mouse mix buffer (MMB):** $50 \text{ ml } 1 \text{ M Tris/HCl pH } 7.5$; $5 \text{ ml } 0.5 \text{ M EDTA}$; $20 \text{ ml } 5 \text{ M NaCl}$; $5 \text{ ml } 2\% \text{ (w/v) SDS}$ ad $500 \text{ ml H}_2\text{O}$ (bidest.).

Blastocyst injection and implantation into foster mice

Injection of targeted embryonic stem cells into blastocysts as well as blastocyst implantation into pseudo-pregnant mice was kindly performed by John Kong a Sang at the Erasmus University of Rotterdam. Mouse embryos were genotyped either by Southern Blot analysis with probe ES-a-SB (2.2.2.2.6 and 2.1.9.3) or by PCR with primers MSp2Ex/L3Seq3, MSp2In2Seq4 and Neo as described (2.2.2.2.3).

General mouse handling

To avoid infections, mice were kept and handled as sterile as possible. Only up to 3 mice were kept in a 22 x 16 cm cage. Mice were used for breeding when being 8-10 weeks old. To guarantee sterility as well as the accordance with the legal requirements of animal protection (Tierschutzgesetz vom 25.05.1998) and gene technology (Gentechnikgesetz vom 16. Dezember 1993), mice were handled only by the personal of the mouse facilities. All data concerning mouse strains and breeding are available at the mouse data bank of the Erasmus University of Rotterdam (EUR).

2.2.4.6 Body size and weight measurements and fixation of mouse embryos

Body size and weight of mouse embryos (day E9.5, E12.5 and E18.5) were measured immediately after dissection. To determine the body size, the distance between top of the skull and tip base was measured. After measuring, embryos were fixed in 4% paraformaldehyde in PBS for 2 hours until over night under gentle stirring, washed twice in PBS for each 30-60 min and stored in 70% ethanol at 4°C until further procedure. Yolk sacs or tail tips were used for PCR genotyping (see 2.2.2.2.3). To statistically evaluate measurement results, t-tests were performed.

2.2.4.7 X-Gal and BluoGal staining of mouse embryos

X-Gal and BluoGal staining of mouse embryos was performed by Dr. Sjaak Philipsen (Erasmus University of Rotterdam).

X-Gal staining

To investigate Sp2 expression in mouse embryos and to prove functional integration of the knockout construct into the mouse genome, X-Gal and BluoGal stainings based on LacZ expression in Sp2-targeted mice were performed with day E12.5 embryos. Embryos were prepared and genotyped as described above followed by 30 min fixation in 4% paraformaldehyde at 4°C while stirring gently. After fixation, embryos were washed three times in LacZ washing buffer for each 30 min, first at 4°C then at room temperature, and stained in LacZ staining solution at 37°C over night in the dark while stirring gently. In case of precipitate formation, embryos were

shortly washed in DMSO after staining. Pictures were taken by a camera (JVC video camera KY-F55BE) mounted on a stereo microscope (Leica Wild M10). Embryos were stored until further use in 70% ethanol at 4°C.

Solutions:

- **LacZ washing buffer:** 0.2 M Sodium phosphate buffer pH 7.0; 2 mM MgCl₂; 0.02% NP-40; 0.01% Sodium desoxycholate (10% stem solution).
- **LacZ staining solution pH 7.3:** 82.5 mg K₃Fe(CN)₆; 105 mg K₄Fe(CN)₆; 100 µl 0.5 M EGTA in 50 ml LacZ washing buffer; 1 mg/ml 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal solution in DMSO).

BluoGal staining

BluoGal staining of mouse embryos was performed in the same way as described for X-Gal staining. Instead of X-Gal, BluoGal was used at the same concentration.

3. Results

3.1 Molecular characterization of the transcription factor Sp2

3.1.1 Generation of rabbit polyclonal Sp2-specific antibodies

When initiating this thesis work, no Sp2-specific antibodies were available. For this, Sp2-specific polyclonal antibodies were generated by expressing Sp2 protein in *E. coli* BL21DE3, followed by immunizing two New Zealand White rabbits (no. 193, “Zwick” and no. 194, “Zwack”) with SDS-PAGE-purified Sp2 protein (see 2.2.3.1 and 2.2.3.6). The specificity of the obtained antiserum was tested by Western Blot analysis (2.2.3.5).

3.1.1.1 Expression of Sp2 protein in *E. coli* BL21DE3 bacteria

Mouse Sp2 protein was expressed by pET3d-mSp2 in two 1 l cultures of *E. coli* BL21DE3 bacteria. In both cultures IPTG-induced Sp2 expression was detectable after 2 hours of induction.

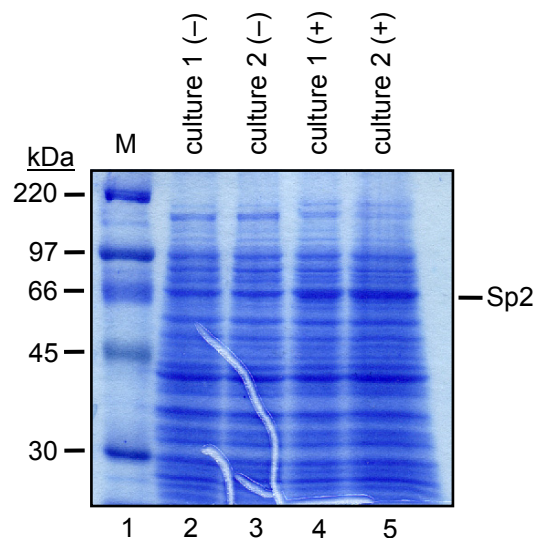


Fig. 3.1.1.1. Expression of mouse Sp2 protein in two 1 l cultures of *E. coli* BL21DE3. 1 ml of each culture (culture 1 and 2) before (-) and after (+) IPTG induction was analyzed for Sp2 expression. Bacterial pellets were resuspended in 2x Laemmli protein loading buffer and 10 μ l of each protein solution separated by 10% SDS-PAGE. M, Rainbow® marker. Coomassie staining.

3.1.1.2 Isolation of inclusion bodies containing recombinant Sp2 protein

Mouse Sp2 protein expressed in *E. coli* BL21DE3 accumulates in inclusion bodies. Inclusion bodies of two IPTG-induced 1 l cultures were isolated according to Nagai et al. (1985; see also 2.2.3.1) and an aliquot of each culture was analyzed by SDS-PAGE. Sp2 protein could be detected in both inclusion body aliquots (Fig. 3.1.1.2).

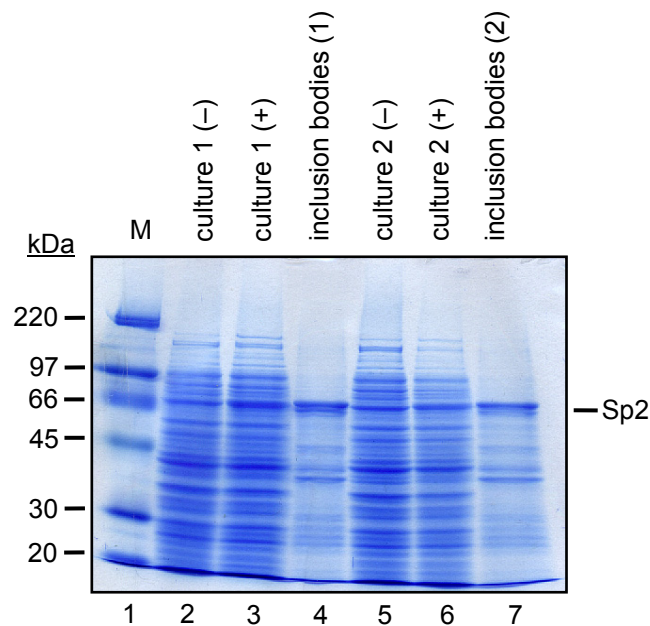


Fig. 3.1.1.2. Purified inclusion bodies containing recombinant Sp2 protein. Inclusion body pellets derived from two 1 l *E. coli* BL21DE3 cultures were resuspended in equal volumes of 2x Laemmli protein loading buffer. 10 μ l of each inclusion body solution were separated by 10% SDS-PAGE (lane 4 and 7). As controls, 10 μ l of protein solution derived from each non-induced (-) and induced bacteria culture (+) were loaded (lane 2, 3, 5 and 6). M, Rainbow® marker (lane 1). Coomassie staining.

3.1.1.3 Immunization of rabbits

Mouse Sp2 protein was purified from inclusion bodies by preparative SDS-PAGE, lyophilized and dissolved in H₂O as described in 2.2.3.1. Protein concentration was determined by comparing increasing Sp2 volumes with increasing amounts of a BSA standard of known concentration by SDS-PAGE, resulting in a Sp2 protein concentration of approximately 100 ng/ μ l and a total amount of 500 μ g of recombinant mouse Sp2 protein.

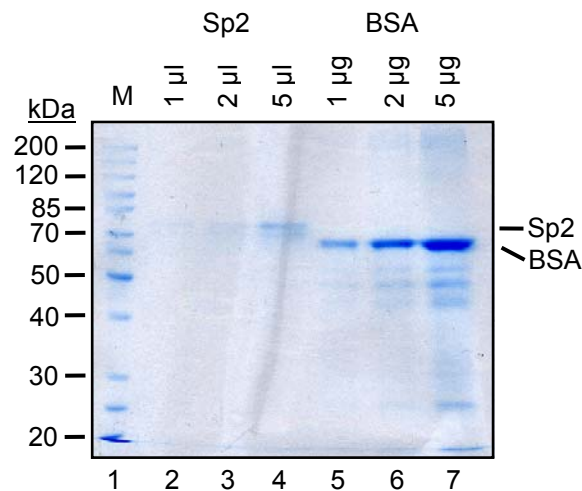


Fig. 3.1.1.3. Rough quantification of purified recombinant mouse Sp2 protein. Quantification of recombinant mouse Sp2 protein after purification by preparative SDS-PAGE and lyophilization. Increasing Sp2 volumes (1-5 µl; lanes 2-4) compared with increasing BSA amounts (1-5 µg; lanes 5-7), each dissolved in equal volumes of 2x Laemmli protein loading buffer. M, Protein-Ladder (MBI Fermentas, St. Leon-Rot) (lane 1). Coomassie staining. 10% SDS-PAGE.

As the total amount of recombinant Sp2 protein was insufficient for rabbit immunization, expression of mouse Sp2 protein in *E. coli* BL21DE3 was repeated under the same conditions as described resulting in additional 250 µg of Sp2 protein (data not shown), thus leading to a final amount of 750 µg recombinant mouse Sp2 protein. To generate Sp2-specific polyclonal antibodies, two rabbits were immunized with each 75 µg of recombinant Sp2 per immunization reaction (see 2.2.3.6).

3.1.1.4 Characterization of Sp2-specific rabbit antisera

Antiserum taken 10 days after the first boost (see 2.2.3.6) was analyzed in Western Blot for Sp2 specificity. In addition, pre-immune serum that has been obtained before initiating immunization, was used as a control for antiserum specificity. As depicted in Fig. 3.1.1.4, Sp2 antiserum but not the pre-immune serum (both derived from rabbit no. 193, “Zwick”) specifically detects recombinant Sp2 protein in Western Blot without cross reactivity towards recombinant Sp1 or Sp3 protein. This specificity could be confirmed by Western Blot analysis of nuclear extracts (*Drosophila* SL2, HEK-293 cells) containing overexpressed mouse Sp2 protein as well as by Electrophoretic Mobility Shift Assays (EMSAs) (see 3.1.5).

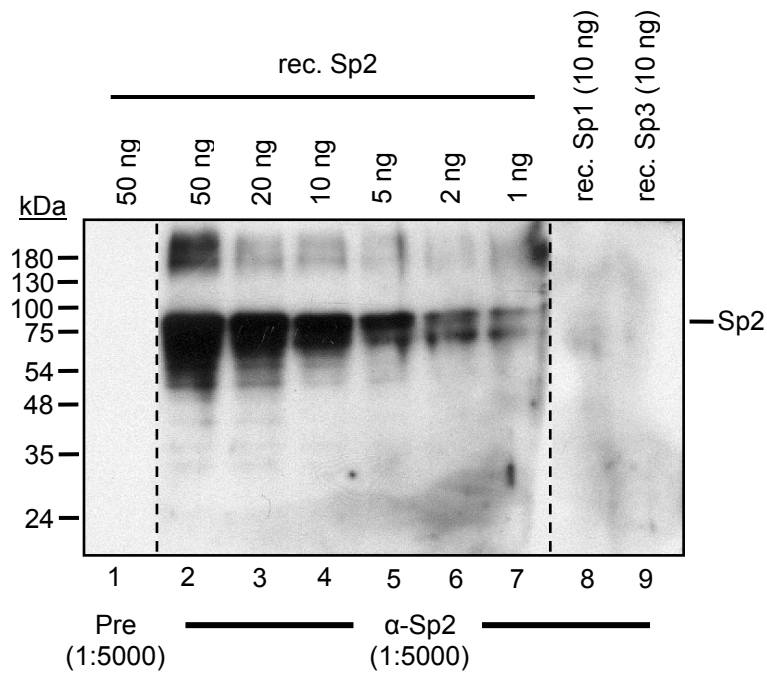


Fig. 3.1.1.4. Analysis of Sp2-specific rabbit antisera. Western Blot analysis. Decreasing amounts of recombinant mouse Sp2 protein (50-1 ng) were separated through 10% SDS-PAGE, blotted to a PVDF membrane and incubated with Sp2-specific antiserum (lane 2-7). Antiserum was obtained from immunized rabbit no. 193 (“Zwick”), 10 days after the first boost. As control, rabbit pre-immune serum was used (Pre; lane 1). To analyze antiserum cross reactivity towards other Sp factors, 10 ng of each recombinant Sp1 and Sp3 protein were incubated with the antiserum (lane 8-9). Pre-immune and antiserum were diluted 1:5000.

3.1.2 Endogenous Sp2 protein expression

When initiating this thesis, no data about endogenous Sp2 expression were available (see 1.4). To study Sp2 protein expression, Western Blot analyses were performed using various cell lines as well as adult mouse tissues.

3.1.2.1 Endogenous Sp2 protein expression in various cell lines

To study endogenous Sp2 protein expression in cells, nuclear extracts derived from two mouse (mouse embryonic fibroblasts (MEF); mouse embryonic stem cells (ES cells)) and two human (human embryonic kidney cells (HEK-293); Ishikawa cells) cell lines were investigated by Western Blot analysis. As depicted in Fig. 3.1.2.1, Sp2

protein is detectable in all four cell lines by two distinct bands of approximately 75 kDa. Whether these two bands reflect two translationally controlled Sp2 isoforms, alternative splice variants or different posttranslational modifications, has to be clarified. The Sp2-specific antiserum not only detects endogenous mouse Sp2 but also human Sp2. This can be explained by the high sequence homology between mouse and human Sp2 protein (see 3.2.1).

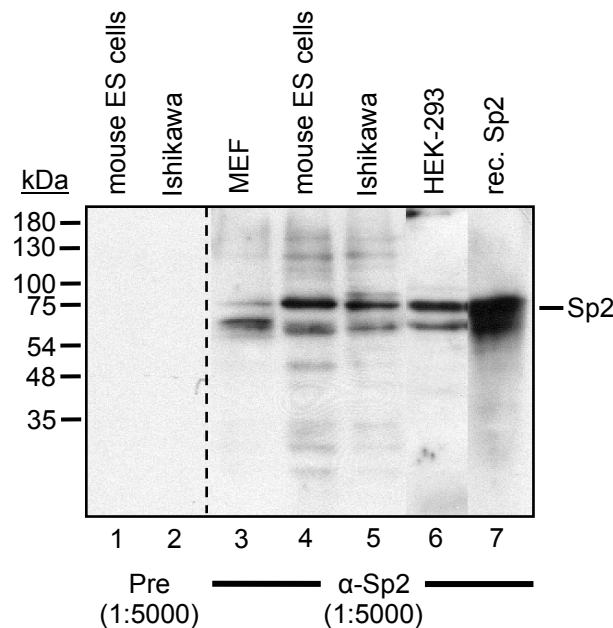


Fig. 3.1.2.1. Endogenous Sp2 protein expression in various cell lines. Western Blot analysis. 28 μ g nuclear protein extracts of MEF (lane 3), undifferentiated mouse ES (lane 4), Ishikawa (lane 5) and HEK-293 cells (lane 6) were separated through 10% SDS-PAGE, blotted to a PVDF membrane and incubated with Sp2-specific antiserum. 20 ng recombinant Sp2 protein were used as positive control (lane 7). To control specificity, 28 μ g nuclear protein extracts from undifferentiated mouse ES (lane 1) and Ishikawa cells (lane 2) were incubated with rabbit pre-immune serum (Pre). Pre-immune and Sp2 antiserum (rabbit no. 193; “Zwick“) were diluted 1:5000.

3.1.2.2 Endogenous Sp2 protein expression in adult mouse tissues

To determine Sp2 expression pattern in the adult mouse, total cell extracts from adult mouse organs of a Sp2 wildtype mouse were analyzed in Western Blot. The extracts were kindly provided by Grigore Rischitor and prepared as described in Rischitor (PhD thesis, 2005). Although Sp2 expression signals were weak, Sp2 protein is present in all tested organs, however with a certain variation of the expression level

(Fig. 3.1.2.2). Whereas high expression levels could be detected for liver, kidney and heart, the expression values of lung and muscles were comparably low. Nevertheless, also in the adult mouse, Sp2 is at least widely expressed and not restricted to a subset of organs or body regions.

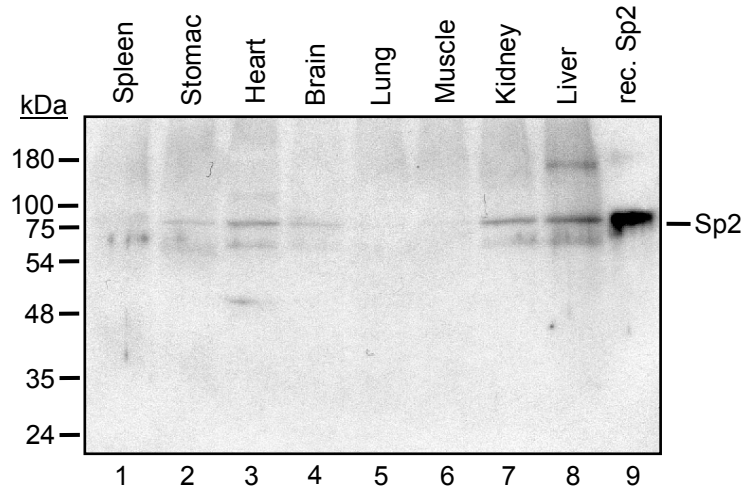


Fig. 3.1.2.2. Endogenous Sp2 protein expression in adult mouse tissues. Western Blot analysis. Each 30 µg protein derived from total cell extracts of different adult mouse organs were separated through 10% SDS-PAGE as indicated above, blotted to a PVDF membrane and incubated with Sp2-specific antiserum (lane 1-8). 20 ng recombinant Sp2 protein were used as positive control (lane 9). Sp2 antiserum (rabbit no. 193; "Zwick") was diluted 1:5000.

3.1.3 Subcellular localization of Sp2 protein

To obtain information concerning the subcellular localization of endogenous Sp2, immunofluorescence analysis of Sp2 protein in HEK-293 cells was performed. Endogenous Sp2 is expressed exclusively in the nucleus, thus displaying the same subcellular localization as Sp1 and Sp3 (Fig. 3.1.3).

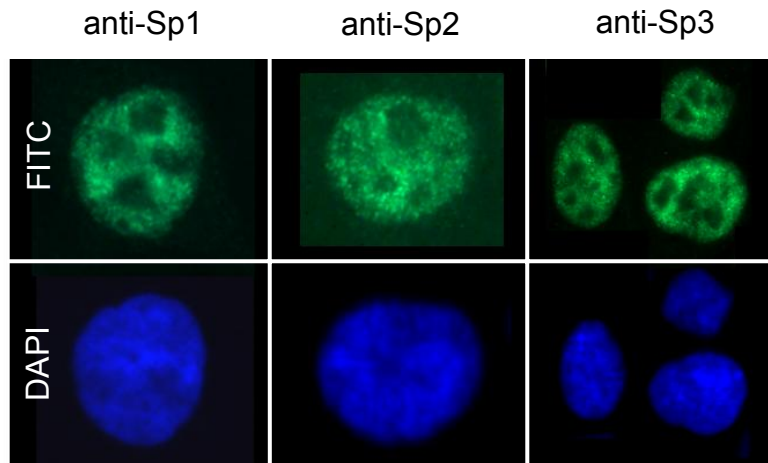


Fig. 3.1.3. Subcellular localization of Sp2 protein in HEK-293 cells. Detection of Sp1, Sp2 and Sp3 protein by α -Sp1 (1:3000), α -Sp2 rabbit no. 193 “Zwick” (1:5000) or α -Sp3 sc-644 (1:1000) and a FITC-conjugated secondary antibody (1:300) as indicated. Nuclei are DAPI-stained. Immunofluorescence microscope pictures.

3.1.4 Transactivation properties of Sp2 protein overexpressed in SL2 cells

An important aspect of transcription factor function is its ability to regulate gene expression. To investigate Sp2 transactivation properties, reporter assays were performed in *Drosophila* SL2 cells. As described in 2.2.4.3-4, 4 μ g of a luciferase (pGL3-Prom) or CAT reporter plasmid (pBCAT-2) were transiently co-transfected along with 2 μ g of a Sp-independent β -galactosidase expression plasmid (p97b) and 0.02-0.5 μ g of each Sp1, Sp2 or both, Sp1 + Sp2 expression constructs (Fig. 3.1.4). In the reporter plasmid pGL3-Prom a luciferase gene is driven by the Simian Virus 40 (SV40) promoter containing five Sp1 binding sites (GC boxes), whereas in the pBCAT-2 plasmid an artificial promoter harbouring two Sp1 binding sites of the HTLV promoter and an E1b-TATA box regulates CAT gene expression. The β -galactosidase expression plasmid p97b was used to compare transfection efficiencies.

As depicted in Fig. 3.1.4, Sp1 (0.5 μ g) strongly activates both promoters (SV40: 84 fold; BCAT-2: 21 fold activation). However, on both promoters, Sp2 activation values did not exceed the values derived from the empty plasmid (pPacHD). Moreover, Sp2

had no effect on Sp1 activity (and vice versa) under these conditions. Only for the BCAT-2 promoter, a slight decrease of Sp1 activity from 11 fold to 4 fold activity was detectable in the case of Sp1 and Sp2 co-transfection (Fig. 3.1.4.B, column 2 and 6). However, this can be a normal variation.

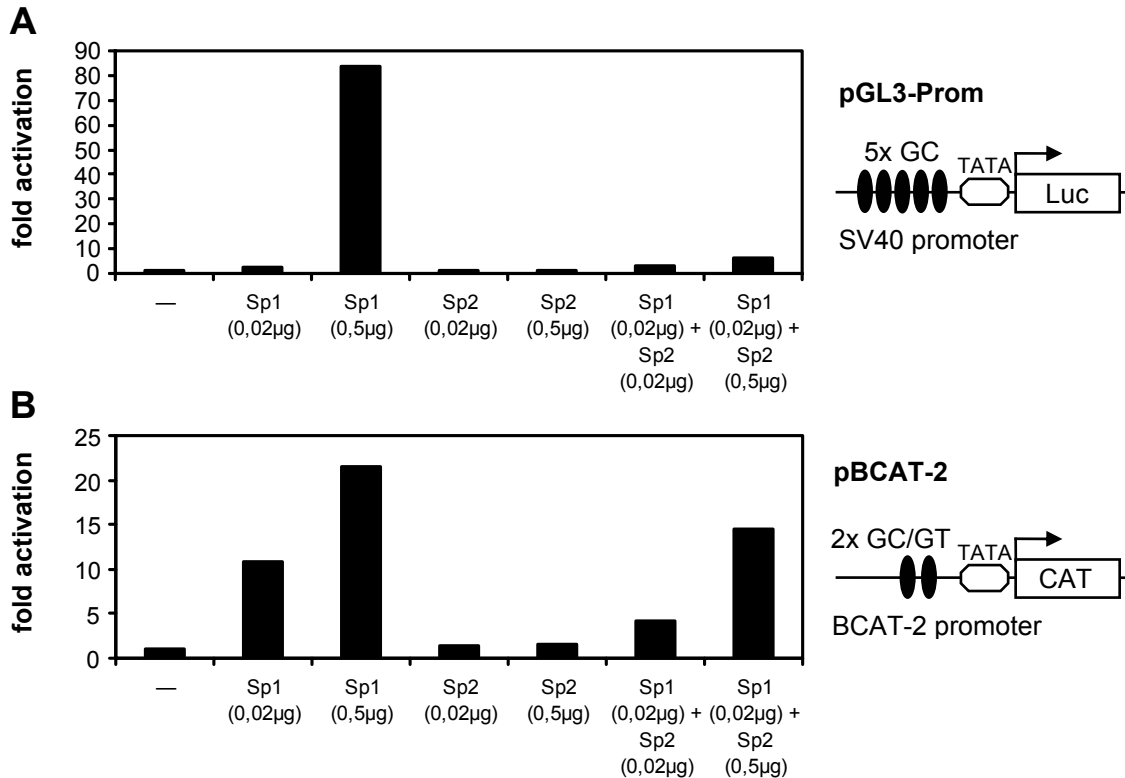


Fig. 3.1.4. Transactivation properties of full-length Sp2 protein overexpressed in SL2 cells. *Drosophila* SL2 cells were transiently co-transfected with 4 µg of pGL3-Prom (A) or pBCAT-2 reporter (B), 2 µg p97b β-galactosidase expression plasmid and 0.02 or 0.5 µg pPac-Sp1 and/or pPac-mSp2(1-606) as indicated. As negative control, the empty plasmid pPacHD was co-transfected with the reporter constructs (A, B; column 1). Transfection was performed by the Calcium-phosphate method. Cells were harvested 48 hours after transfection. **A.** Luciferase activity assay. **B.** Chloramphenicol acetyl transferase (CAT) activity assay.

3.1.5 DNA binding capacity of Sp2 protein

As transcription factor Sp2 did not activate the GC/GT-box-driven luciferase and CAT reporter genes, we next investigated whether Sp2 protein binds to these sequences. The interaction between transcription factors and DNA is a complex and often regulated process, crucial for the activation capability of the transcription factors (see 1.1-3). When initiating this thesis work, no detailed report concerning Sp2-DNA

interactions was available (apart from the few data of Kingsley and Winoto, 1992). However, based on the high sequence homology of the DNA-binding domain (zinc finger region) within the individual Sp family members (see 1.4), Sp2 binding to the classical GC boxes, was expected.

To study Sp2 DNA binding properties, Electrophoretic Mobility Shift Assays (EMSAs) were performed using overexpressed and endogenous Sp2 protein derived from insect or mammalian cell lines. The binding capacity towards various oligonucleotides harbouring different potential Sp2 binding sites like GC, GT or CT boxes as well GC box mutants (see 2.1.9.2) was investigated. In addition to full-length Sp2 protein (3.1.5.1-4), the DNA binding capacity of different Sp2 deletion mutants was analyzed (3.1.5.5-7).

3.1.5.1 GC box binding capacity of full-length Sp2 protein overexpressed in *Drosophila* SL2 cells

To investigate Sp2 GC-box-binding capacity, full-length Sp2 protein was overexpressed in SL2 cells (FuGENE 6 transfection) and nuclear extracts were incubated with a GC box oligonucleotide as described (Sp2-G; see 2.1.9.2) and analyzed by Electrophoretic Mobility Shift Assay (EMSA). Nuclear extracts derived from Sp1- and mock- transfected SL2 cells were used as positive and negative controls, respectively. As depicted in Fig. 3.1.5.1, no interaction between overexpressed Sp2 protein and the DNA is visible although Sp2 protein is detectable in Western Blot. Thus, Sp2 incapability to activate reporter gene expression in transactivation assays (see 3.1.4) can be explained by the lack of DNA binding.

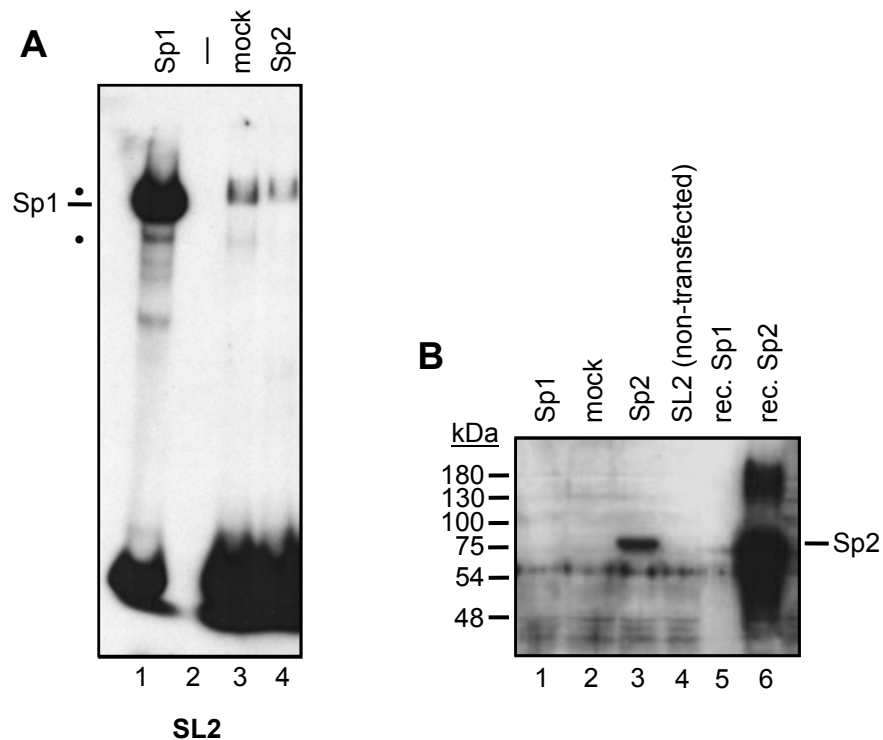


Fig. 3.1.5.1. GC box binding capacity of full-length Sp2 protein overexpressed in Drosophila SL2 cells. **A.** Electrophoretic Mobility Shift Assay (EMSA). 4% native PAA gel. 5 μ g nuclear extract of either Sp1- (pPac-Sp1; lane 1), mock- (pPacHD; lane 3) or Sp2-transfected (pPac-mSp2(1-106); lane 4) SL2 cells were incubated with the 32 P-labeled GC box oligonucleotide Sp2-G (GC box sequence: GGGGCGGGG; see also 2.1.9.2). (*), unspecific signal. FuGENE 6 transfection. **B.** Western Blot analysis to detect Sp2 protein expression in the transfected SL2 cells. 10 μ g of SL2 nuclear extracts were separated through 10% SDS-PAGE as indicated, blotted to a PVDF membrane and incubated with Sp2-specific antiserum (rabbit no. 193; "Zwick"); dilution 1:5000. 30 ng recombinant Sp2 protein were used as positive control (lane 6), 30 ng recombinant Sp1 protein as negative control (lane 5).

3.1.5.2 GC box binding capacity of endogenous Sp2 protein (MEF, HEK-293 and HeLa cells)

To exclude artifacts from Drosophila SL2 cell expression, nuclear extracts from MEF, HEK-293 and HeLa cells were used to investigate DNA binding properties of endogenous Sp2 protein. Extracts were incubated with the GC box oligonucleotide Sp2-G (GC box sequence: GGGGCGGGG; see also 2.1.9.2) and analyzed by Electrophoretic Mobility Shift Assay (EMSA). As in the case of SL2-overexpressed Sp2 protein, no DNA interaction could be detected in supershift assays, although Sp2 protein was detectable in Western Blot (Fig. 3.1.5.2).

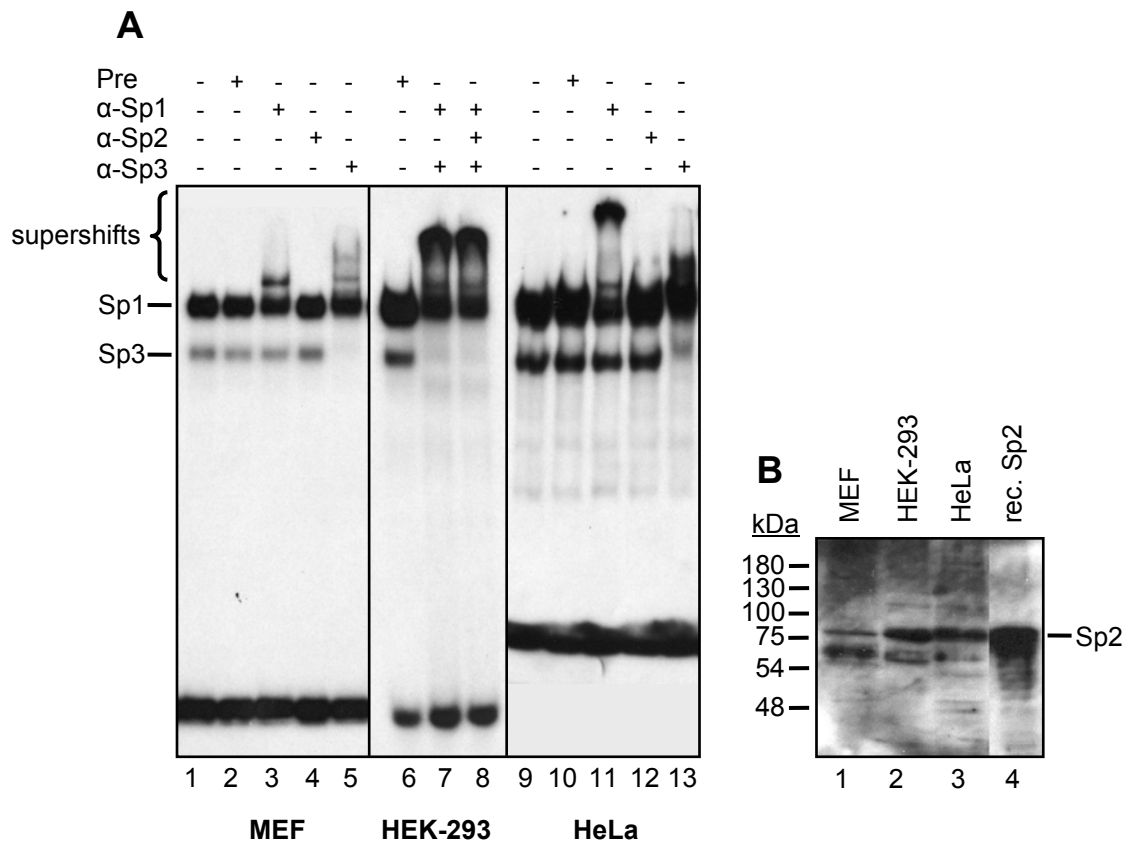


Fig. 3.1.5.2. Potential GC box binding activity of endogenous Sp2 in various cell lines. A. Electrophoretic Mobility Shift Assays (EMSAs). 4% native PAA gel. Nuclear extracts derived from MEF (each 2 µg; lane 1-5), HEK-293 (each 8 µg; lane 6-8) or HeLa cells (each 5 µg; lane 9-13) were incubated with the ³²P-labeled GC box oligonucleotide Sp2-G in presence (+) or absence (-) of antibodies/-sera against Sp1-3 or pre-immune serum (Pre) as indicated. To see that the Sp2-specific antiserum (rabbit no. 193; “Zwick”) is able to recognize Sp2 protein in EMSA, the reader is referred to Fig. 3.1.5.6.D. **B.** Western Blot analysis to monitor Sp2 expression in the respective cells. MEF (15 µg; lane 1), HEK-293 (20 µg; lane 2) and HeLa (20 µg; lane 3) nuclear extracts were separated through 10% SDS-PAGE as indicated, blotted to a PVDF membrane and incubated with Sp2-specific antiserum (rabbit no. 193; “Zwick”); dilution 1:5000. 10 ng recombinant Sp2 protein were used as positive control (lane 4).

As visible in Fig. 3.1.5.2.A, the presence of other Sp family members (like Sp1 and Sp3) binding with high affinity to the GC box oligonucleotide, extremely impede a potential Sp2 detection. To prevent from this, all further experiments were performed with Sp2 protein, mainly overexpressed in *Drosophila* SL2 cells, which are characterized by the lack of endogenous Sp factors (Suske, 2000).

3.1.5.3 Binding capacity of SL2-overexpressed full-length Sp2 protein towards GT and CT boxes

As described in chapter 1.3, the DNA-binding domain (zinc finger region) is highly conserved among the individual members of the Sp/XKLF super-family. However, Sp2 displays a slightly altered amino acid sequence in the first zinc finger (chapter 1.4). For this, Kingsley and Winoto (1992) claimed a higher binding affinity of Sp2 protein towards GT than GC boxes. To consider this, Sp2 DNA binding capacity was investigated using GT (GGGGTGTGG) as well as CT box (CGCCTCCCC or TCCCTCCCC) oligonucleotides (see also 2.1.9.2). However, SL2-overexpressed Sp2 protein did neither bind to the GT nor to the CT box oligonucleotides (Fig. 3.1.5.3.A).

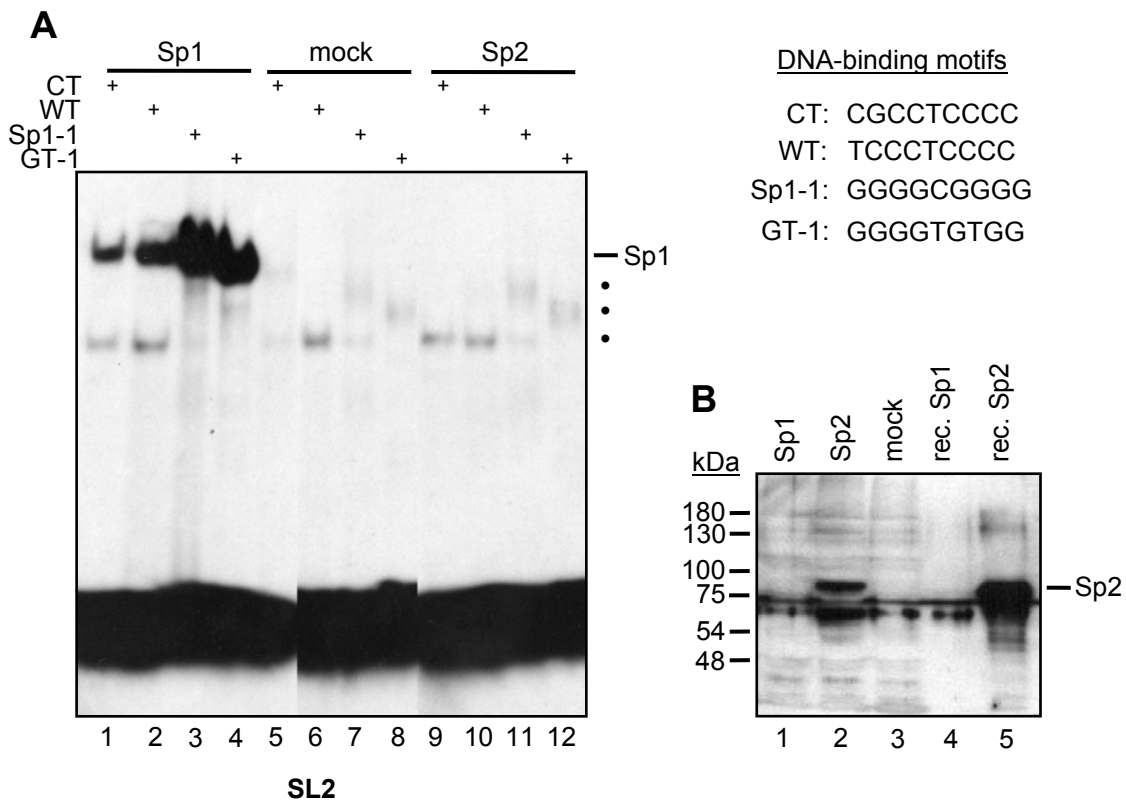


Fig. 3.1.5.3. Binding capacity of SL2-overexpressed, full-length Sp2 protein towards GT and CT boxes. **A.** *Left:* Electrophoretic Mobility Shift Assay (EMSA). 4% native PAA gel. 4 µg nuclear extract of either Sp1- (pPac-Sp1; lane 1-4), mock- (pPacHD; lane 6-8) or Sp2-transfected (pPac-mSp2(1-106); lane 9-12) SL2 cells were incubated with either a ³²P-labeled CT box (CT, WT-1), GT box (GT-1) or GC box (Sp1-1) oligonucleotide as indicated (for detailed oligonucleotide sequences, see 2.1.9.2). (*), unspecific signal. FuGENE 6 transfection. *Right:* Sequences of the DNA binding motifs inside the CT, WT, Sp1-1 and GT-1 oligonucleotides. **B.** Western Blot analysis to detect Sp2 protein expression in the transfected SL2 cells. 10 µg of SL2 nuclear extracts were separated through 10% SDS-PAGE as indicated, blotted to a PVDF membrane and incubated with Sp2-specific antiserum (rabbit no. 193;

“Zwick”); dilution 1:5000. 20 ng of recombinant Sp1 (lane 4) or Sp2 (lane 5) protein were used as negative and positive control, respectively.

3.1.5.4 Binding capacity of SL2-overexpressed full-length Sp2 protein towards GC box variants

To further consider the H→L substitution in the first zinc finger of Sp2, mutated GC box oligonucleotides were designed (2.1.9.2) following the so-called zinc finger code (see 1.3 and Fig. 3.1.5.4.A). To investigate Sp2 DNA binding capacity, nuclear extracts derived from SL2 cells were incubated with these mutated GC box oligonucleotides and analyzed in EMSA. As depicted in Fig 3.1.5.4.B, in contrast to overexpressed Sp1, Sp2 protein did not bind to any of the tested oligonucleotides.

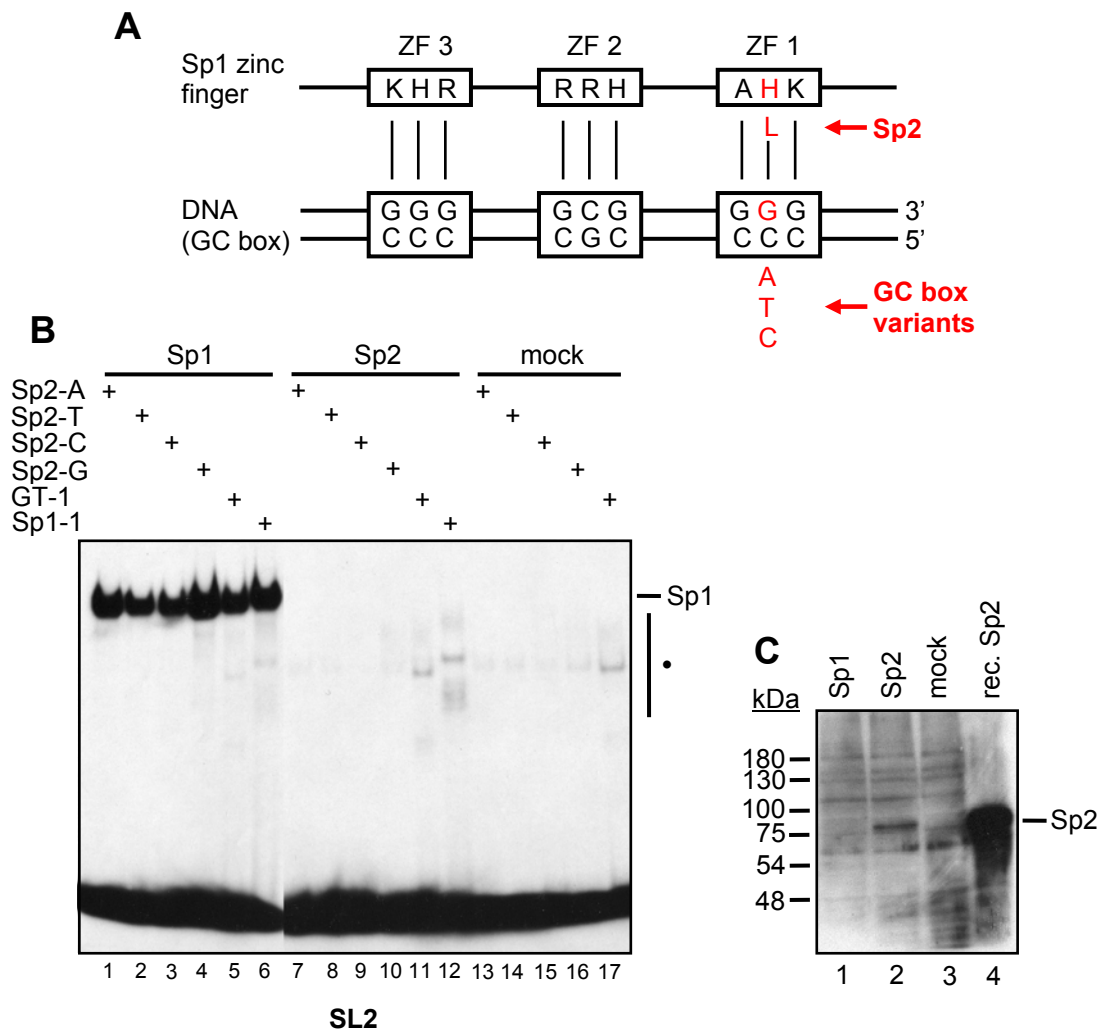


Fig. 3.1.5.4. Binding capacity of SL2-overexpressed full-length Sp2 protein towards GC box variants. **A.** Scheme of GC box oligonucleotide mutants. Following the so-called zinc finger code (see 1.3), three amino acids of each zinc finger enable the specific contact between a zinc finger and the

DNA by contacting each one nucleotide of the DNA sequence. Considering the H→L substitution in the first zinc finger of the Sp2 protein, GC box oligonucleotides were designed exhibiting a mutation (depicted in red) of the corresponding second G nucleotide. **B.** Electrophoretic Mobility Shift Assay (EMSA). 4% native PAA gel. Nuclear extract of either Sp1- (each 5 µg; pPac-Sp1; lane 1-6), Sp2- (each 10 µg; pPac-mSp2(1-106); lane 7-12) or mock-transfected (each 10 µg; pPacHD; lane 13-17) SL2 cells were incubated with either a ³²P-labeled GC box (Sp1-1 or Sp2-G), a GT box (GT-1) or one of three mutated GC box oligonucleotides (Sp2-A, -C, -T) as indicated (for oligonucleotide sequences, see 2.1.9.2 or Fig. 3.1.5.3.A). (•), unspecific signals. FuGENE 6 transfection. **C.** Western Blot analysis to detect Sp2 expression in SL2 cells. 10 µg of SL2 nuclear extracts were separated through 10% SDS-PAGE as indicated, blotted to a PVDF membrane and incubated with Sp2-specific antiserum (rabbit no. 193; “Zwick”); dilution 1:5000. 30 ng recombinant Sp2 protein were used as positive control (lane 4).

3.1.5.5 DNA binding capacity of the Sp2 DNA-binding domain overexpressed in SL2 cells

Electrophoretic Mobility Shift Assays (EMSAs) were performed using nuclear extracts from SL2 cells that were transiently transfected with the zinc finger domain of Sp2 only. As depicted in Fig. 3.1.5.5, the Sp2 zinc finger fragment Sp2 (aa478-606) is able to bind to the classical GC box as well as GC box variants with similar affinity. The binding affinity towards the GT box oligonucleotide, however, appears to be lower. This suggests that sequences inside the Sp2 protein carry an inhibitory function and somehow prevent Sp2 from binding to DNA. Thus, the H→L substitution in the first zinc finger of Sp2 seems to have no influence on Sp2 DNA binding capacity.

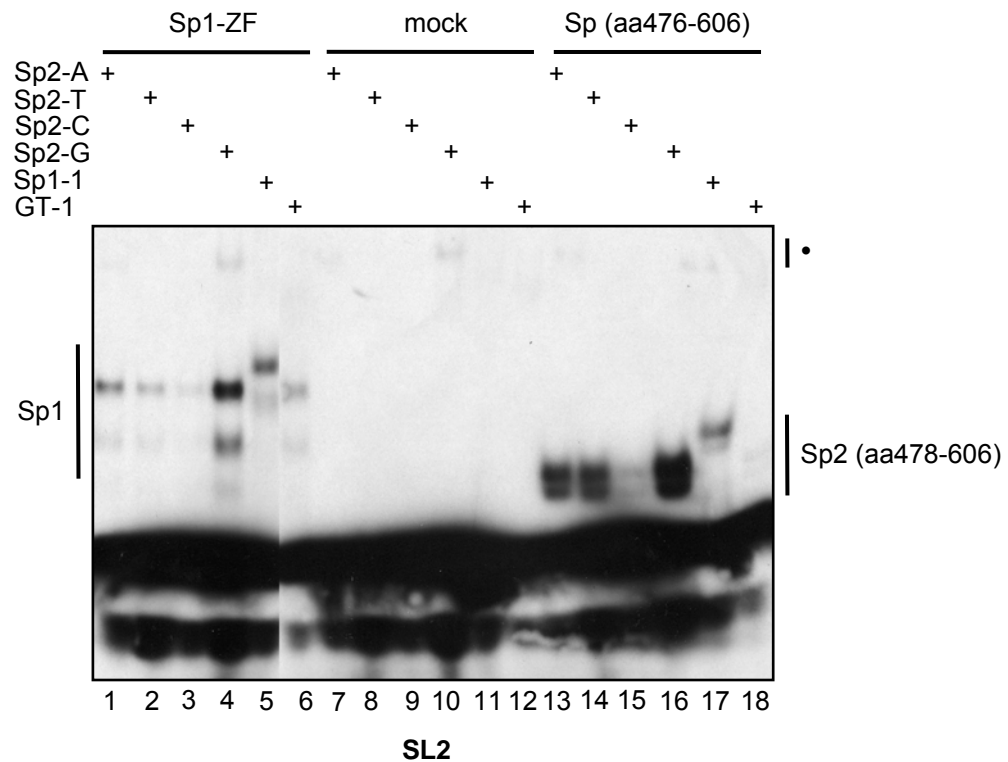


Fig. 3.1.5.5. DNA binding capacity of the Sp2 DNA-binding domain overexpressed in SL2 cells. Electrophoretic Mobility Shift Assay (EMSA). 4% native PAA gel. 5 µg nuclear extract derived from SL2 cells, which have been transfected (FuGENE 6) with an empty vector (pPacHD; lane 7-12) or an expression plasmid for the zinc finger region of either Sp1 (pPacUbx-Sp1ZF; lane 1-6) or Sp2 (pPacUbx-mSp2(478-606); lane 13-18), were incubated with different ³²P-labeled oligonucleotides as indicated (for oligonucleotide sequences, see 2.1.9.2 or Fig. 3.1.5.3-4.A). (•), unspecific signals.

3.1.5.6 DNA binding capacity of SL2-overexpressed N-terminal Sp2 protein deletion mutants

To identify putative sequences that prevent full-length Sp2 protein from binding to DNA, a series of N-terminal Sp2 deletion mutants were generated (Fig. 3.1.5.6.A; see also 2.1.10.2.2) and tested for GC box binding by Electrophoretic Mobility Shift Assay (EMSA). As depicted in Fig. 3.1.5.6.B, all Sp2 truncations lacking the first 179 N-terminal amino acids are able to bind to the GC box oligonucleotide. However, the intensity of the DNA binding signal decreased with the length of the Sp2 fragments. No or only a very weak signal was detected for the deletion fragment Sp2 (aa160-606), lacking the first 159 N-terminal amino acids. As all fragments, which the antibody is able to detect, display similar expression levels in Western Blot (Fig.

3.1.5.6.C), the weak or absent signal has to be interpreted as a reduction or total loss of the DNA binding capability.

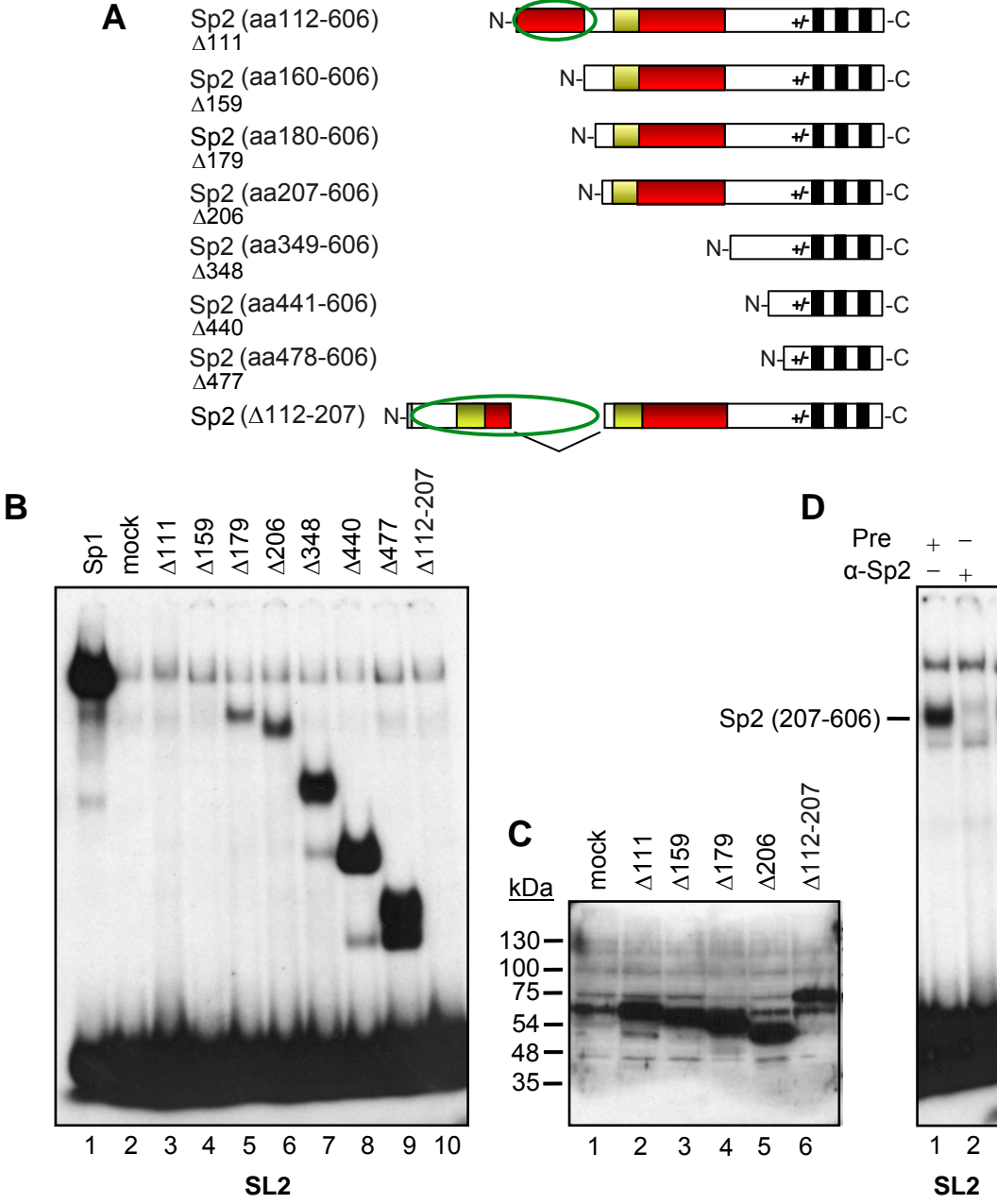


Fig. 3.1.5.6. DNA binding capacity of SL2-overexpressed N-terminal Sp2 deletion mutants. **A.** Schematic drawing of N-terminal Sp2 deletion mutants (see also 2.1.10.2.2). Numbers in brackets represent start and end amino acid of the protein fragments. Glutamine-rich putative transactivation domains are depicted in red, zinc fingers in black, serine/threonine-rich regions in yellow. (+/-) represents an area of charged amino acids. Sp2 (Δ112-207) is fused to a Flag/HA tag (depicted in grey) derived from the plasmid pPacHD-Flag; all other fragments are fused to an Ubx leader sequence derived from the plasmid pPacUbx. Green ellipse point out the N-terminal amino acids being present in the Sp2 deletions exhibiting no DNA binding capacity in EMSA experiments. **B.** Electrophoretic Mobility Shift Assay (EMSA). 4% native PAA gel. 5 μg nuclear extract derived from SL2 cells, which have been transfected with either Sp1 (pPac-Sp1; lane 1), mock (pPacHD; lane 2) or

one of the Sp2 constructs (lane 3-10), were incubated with the ³²P-labeled GC box oligonucleotide Sp2-G and loaded on the gel as indicated. FuGENE 6 transfection. **C.** Western Blot analysis to detect expression of the N-terminal Sp2 deletions in SL2 cells. 10 µg SL2 nuclear extract were separated through 8% SDS-PAGE as indicated, blotted to a PVDF membrane and incubated with Sp2-specific antiserum (rabbit no. 193; "Zwick"), dilution 1:5000. Sp2 deletion fragments starting with amino acid 349, 441 or 478 were not detected by the antiserum. **D.** Electrophoretic Mobility Shift Assay (EMSA) with 5 µg of SL2 nuclear extract expressing the N-terminal deletion mutant Sp2 (aa 207-606) to demonstrate functionality of the Sp2-specific antiserum (rabbit no. 193; "Zwick") in EMSA. Pre, pre-immune serum. 4% native PAA gel.

To investigate whether only a part of the N-terminal 179 amino acid is responsible for the loss of DNA binding capacity, amino acids 112 to 207 were deleted and the Sp2 deletion protein (Sp2 (Δ 112-207)) tested for DNA binding by Electrophoretic Mobility Shift Assay (EMSA). However, the deletion fragment also was unable to bind to the GC box oligonucleotide. (Fig. 3.1.5.6.B). This result indicates that the entire N terminus from amino acid 1-179 contains sequences that prevent SL2-overexpressed Sp2 from binding.

3.1.5.7 DNA binding capacity of Gal4-fused C-terminal Sp2 deletion mutants overexpressed in HEK-293 cells

Based on the result that the N-terminal amino acids 1-179 influence the DNA binding capacity of Sp2, we investigated whether the N terminus also inhibits Sp2-DNA interactions when being N-terminally fused to a heterologous Gal4 DNA-binding domain. For this, C-terminal truncated Sp2 protein fragments were generated lacking the zinc finger region and further C-terminal parts of the protein. In addition, one Sp2 fragment (= Gal4-Sp2 (aa207-471)) was generated lacking both, zinc finger domain and the N-terminal region around domain A (Fig. 3.1.5.7.A; see also 1.4 and 2.1.10.2.3). All four Sp2 fragments were fused at the N terminus to a Gal4 DNA-binding domain. The Gal4-Sp2 fusion constructs were overexpressed in HEK-293 cells and tested for DNA binding capacity in Electrophoretic Mobility Shift Assays (EMSAs). As depicted in Fig. 3.1.5.7.B, a diffuse signal in the range of the gel slots was observed for Gal4-Sp2 (aa6-471), Gal4-Sp2 (aa6-357) and Gal4-Sp2 (aa6-215). In contrast to that, no binding signal was obtained for the Gal4-Sp2 (aa207-471) fusion fragment, although it lacks the first 207 amino acids, which seem to influence

Sp2 DNA binding capacity in the N-terminal Sp2 deletion mutants (see 3.1.5.6). However, all fragments exhibit similar expression levels in Western Blot (Fig.3.1.5.7.C).

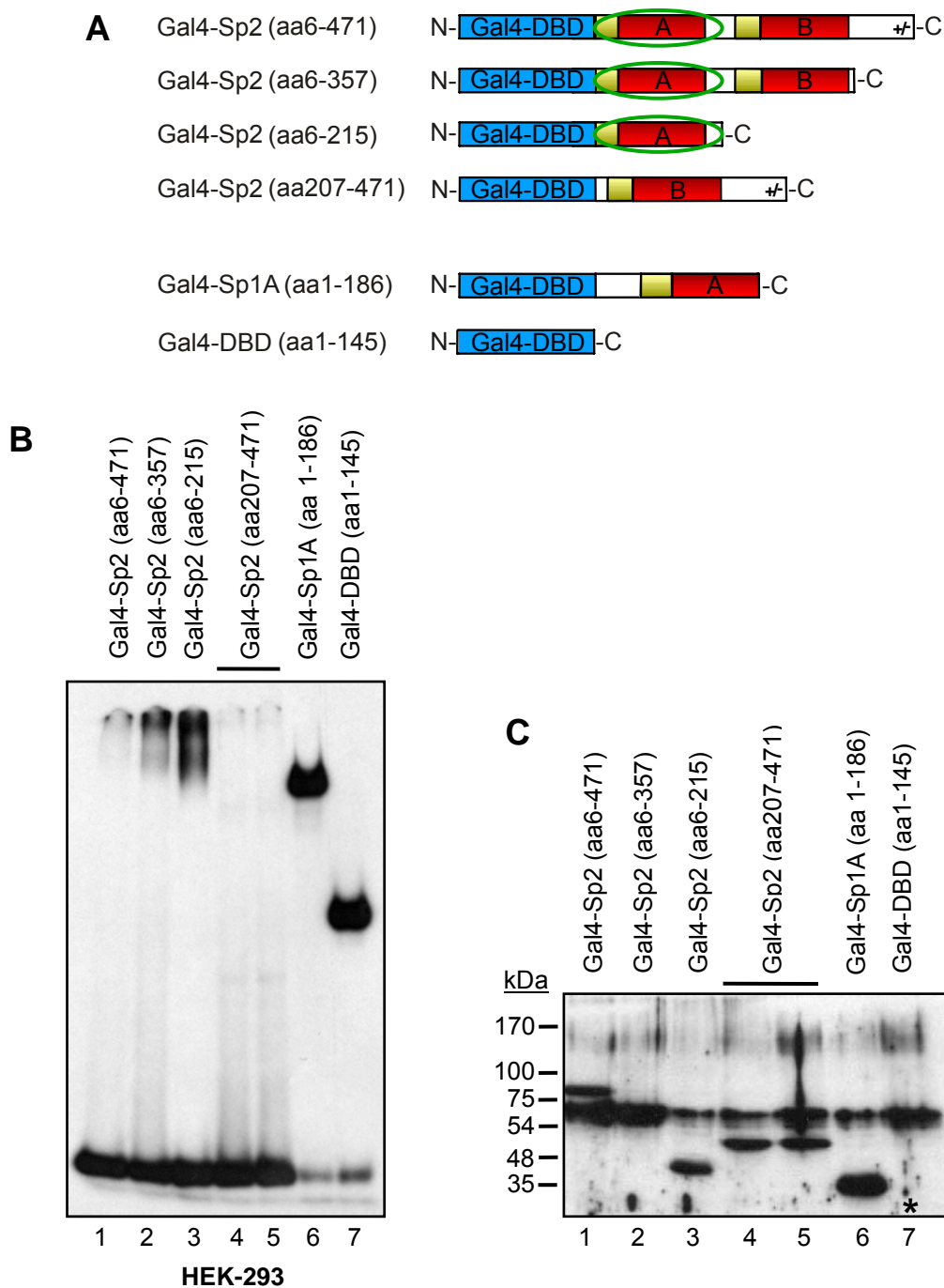


Fig. 3.1.5.7. DNA binding capacity of Gal4-Sp2 fusions overexpressed in HEK-293 cells. A. Schematic drawing of Gal4-fused C-terminal Sp2 deletion mutants (see also 2.1.10.2.3). Numbers in brackets represent start and end amino acid of the protein fragments. Glutamine-rich domains are depicted in red, serine/threonine-rich regions in yellow. (+/-) represents an area of charged amino acids. Instead of possessing the zinc finger region, the Sp2 truncations are fused at the N terminus to a Gal4 DNA-binding domain derived from pSG424Gal4, depicted in blue. Green ellipse point out the

N-terminal amino acids being present in the Sp2 deletion fragments described in 3.1.5.6, which exhibited no DNA binding in EMSAs. **B.** Electrophoretic Mobility Shift Assay (EMSA). 4% native PAA gel. Each 2 µg nuclear extract derived from HEK-293 cells, which have been transfected by FuGENE 6 with each one of the constructs described in (A), were incubated with a ³²P-labeled Gal4 binding site oligonucleotide and loaded on the gel as indicated. To investigate Gal4-Sp2 (aa207-471) DNA binding capacity, EMSA was performed in duplicates with two different nuclear extract preparations containing overexpressed Gal4-Sp2 (aa207-471). **C.** Western Blot analysis to detect expression of the Gal4-fused Sp2 deletions in HEK-293 cells. HEK-293 nuclear extracts (10 µg) were separated by 8% SDS-PAGE as indicated, blotted to a PVDF membrane and incubated with an anti-Gal4-DBD antibody (USA: St. Cruz; no. sc-577), dilution 1:1000. (*) The comparably short Gal4-DBD (aa1-145) fragment could not be detected as it ran out from gel.

These results suggest that the 179 N-terminal amino acids also impair “correct” interaction between the Gal4 DNA-binding domain and the corresponding DNA sequence. Thus, the impairment of DNA binding by the Sp2 N terminus appears to be independent of the position of the DNA-binding domain (zinc finger → C terminus; Gal4-DBD → N terminus). Moreover, the lack of any binding activity in the Gal4-Sp2 (aa207-471) mutant indicates that also sequences between amino acid 207-471 are involved in the DNA binding inhibition.

3.1.6 Transactivation properties of Sp2 deletion mutants displaying DNA binding capacity

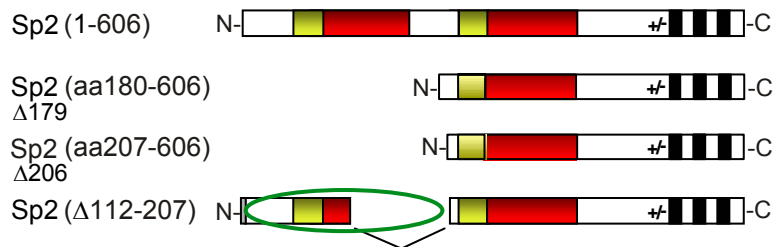
3.1.6.1 Transactivation properties of N-terminal Sp2 deletion mutants overexpressed in SL2 cells

Based on the findings that Sp2 deletion fragments lacking the first 179 N-terminal amino acids are able to bind DNA in Electrophoretic Mobility Shift Assays (EMSAs) (see 3.1.5.6), transactivation assays with these Sp2 deletion mutants were performed. Sp2 (aa180-606) and Sp2 (aa207-606), both binding to GC box oligonucleotides in EMSA, as well as Sp2 (Δ112-207) lacking amino acids 112-207 were tested for their capability to activate reporter gene expression. In addition, control experiments with full-length Sp1 and Sp2 or an empty vector (pPachD) were performed. As reporters, plasmid pGL3-Prom (containing a luciferase gene driven by the SV40 promoter) and p(GC)₂-CAT (containing a CAT gene driven by an artificial

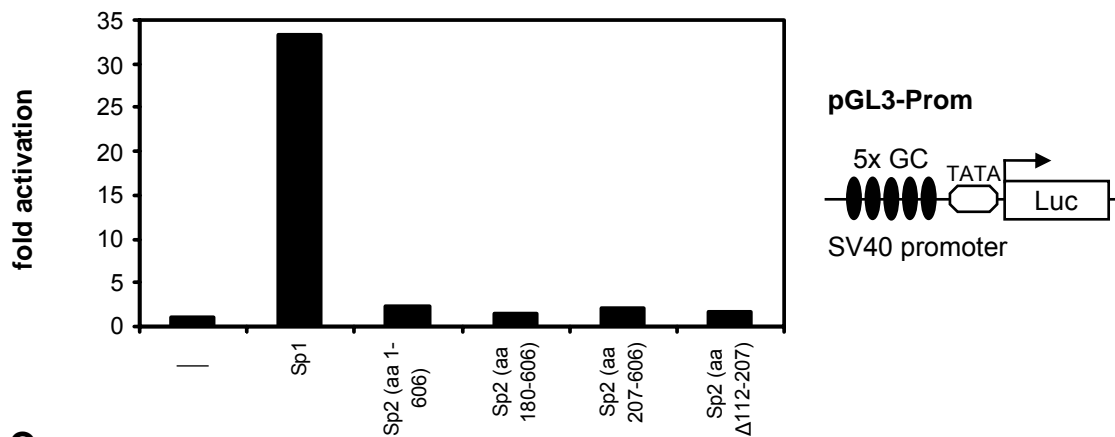
promoter consisting of two GC boxes and the E1b-TATA box) were used. The β -galactosidase expression plasmid p97b served to compare transfection efficiencies.

As depicted in Fig. 3.1.6.1.B-C, activation values of all tested Sp2 deletion fragments are comparable to the values obtained from mock transfections (pPacHD). Although binding to DNA, the Sp2 deletion fragments Sp2 (aa180-606) and Sp2 (aa207-606) lacking the glutamine-rich domain A did not activate reporter gene expression in these assays. In contrast to that, Sp1 strongly activated luciferase and CAT expression (SV40: 33 fold; (GC)₂-CAT: 308 fold activation). Possible reasons for inactivity will be discussed in chapter 4.2.

A



B



C

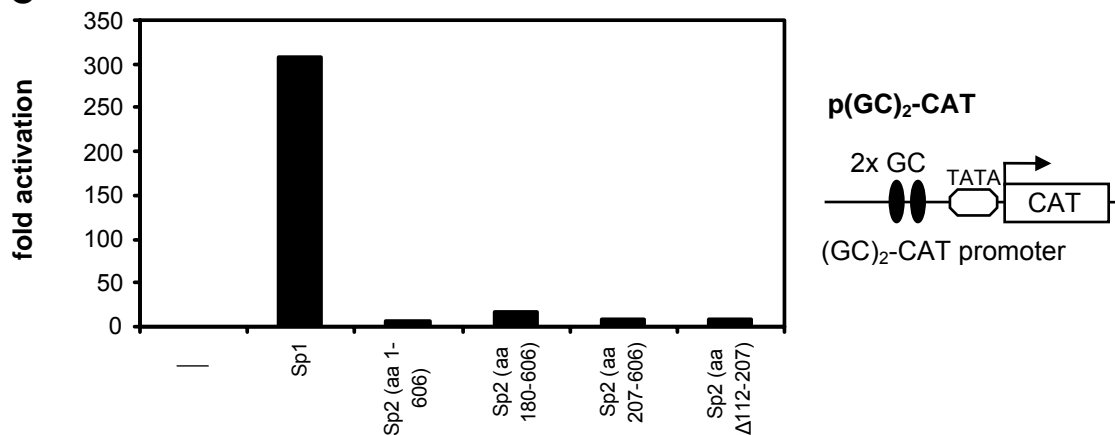


Fig. 3.1.6.1. Transactivation properties of N-terminal Sp2 deletion mutants in SL2 cells. A. Schematic drawing of N-terminal Sp2 deletion mutants used for transactivation analysis (see also 2.1.10.2.2 and Fig. 3.1.5.6.A). Numbers in brackets represent start and end amino acid of the protein fragments. **B-C.** Drosophila SL2 cells were transiently transfected with 4 µg of either pGL3-Prom (B) or p(GC)₂-CAT reporter (C), 2 µg p97b β-galactosidase expression plasmid and 0.5 µg of either pPacHD (column 1), pPac-Sp1 (column 2), pPac-mSp2(1-606) (column 3), pPac-mSp2(180-606) (column 4), pPac-mSp2(207-606) (column 5) or pPac-mSp2(Δ112-207) (column 6), as indicated. Transfection was performed by the Calcium-phosphate method. Cells were harvested for transactivation assays 48 hours after transfection. **B.** Luciferase activity assay. **C.** Chloramphenicol acetyl transferase (CAT) activity assay.

3.1.6.2 Transactivation properties of Gal4-Sp2 deletion mutants overexpressed in HEK-293 cells

Transactivation properties of the Gal4-fused C-terminal Sp2 deletion proteins, at least three of them exhibiting a diffuse DNA binding signal in Electrophoretic Mobility Shift Assays (EMSAs) (see 3.1.5.7), were investigated by luciferase reporter assays. In addition, control experiments with plasmids expressing either only the Gal4 DNA-binding domain (pSG424Gal4) or a Gal4-fused Sp1-A fragment (pSG424Gal4-Sp1A) were performed. As reporter plasmid, pGAWG5E1b, which contains a luciferase gene driven by an artificial promoter consisting of five Gal4 binding sites and the E1b-TATA box, was used. HEK-293 cells were either transiently or stably transfected with the reporter construct and co-transfected with the β-galactosidase expression plasmid pRSV-β-Gal (to standardize transfection efficiencies) and the Gal4-fused expression plasmids.

As depicted in Fig. 3.1.6.2, the three Gal4-Sp2 fragments containing the N-terminal amino acids 1-179 did not activate reporter gene expression. No differences were detected between cells being transiently or stably transfected with the reporter construct. Investigation of the fusion fragment Gal4-Sp2 (aa207-471) displaying no DNA interactions resulted in a 7.3 fold activation in the stably transfected cells. However, no activation was detected in transiently transfected cells, supposing that the 7.3 fold activation reflects the normal variation. In contrast to the Gal4-Sp2 fragments, Sp1 displays 21-53 fold luciferase activity in all experiments.

In summary, no Sp2-mediated activation of reporter gene expression could be detected in these experiments. Possible reasons for Sp2 inactivity will be discussed in chapter 4.2.

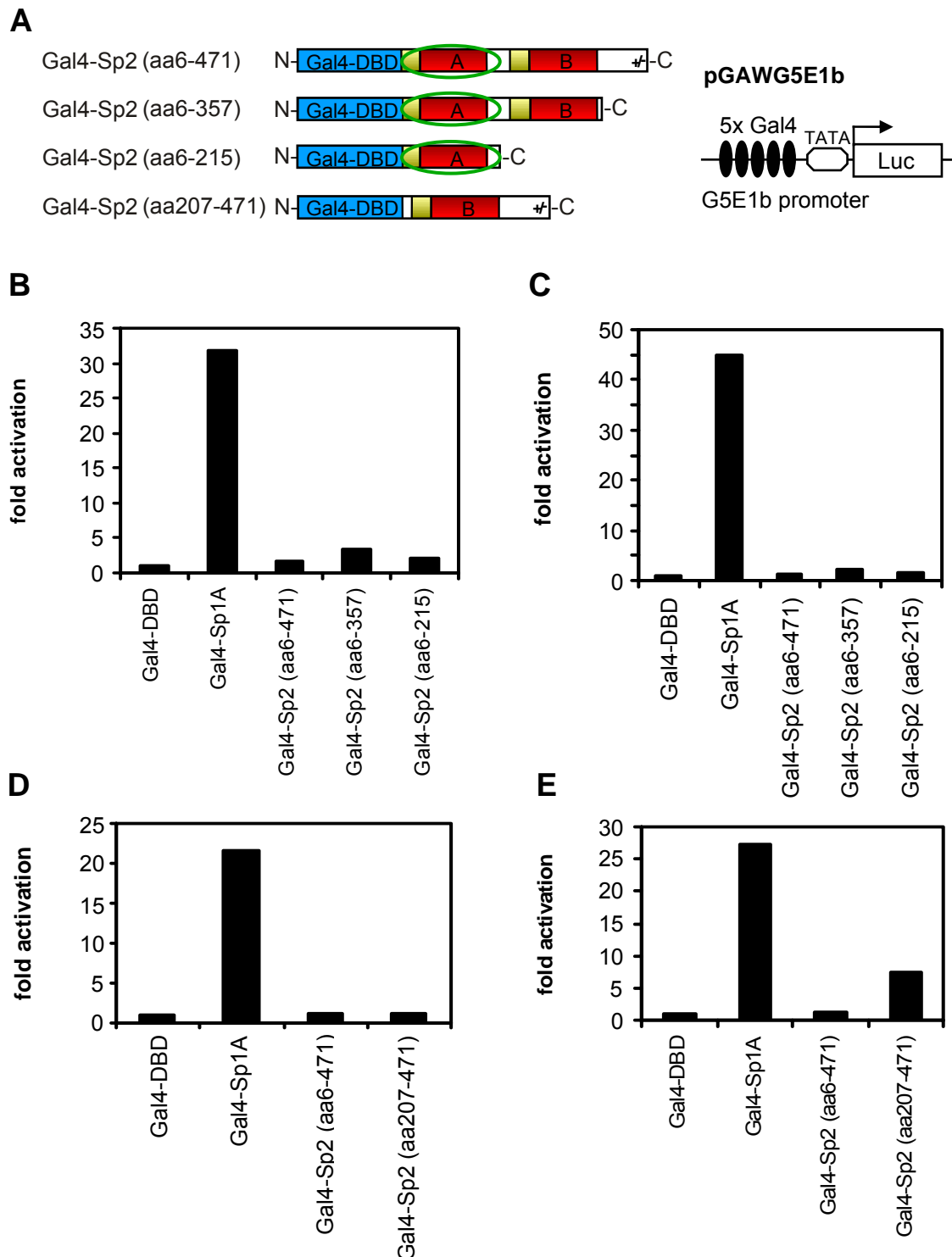


Fig. 3.1.6.2. Transactivation properties of Gal4-Sp2 deletion mutants in HEK-293 cells. *A. Left:* Schematic drawing of Gal4-fused C-terminal Sp2 deletion mutants (see also 2.1.10.2.3). Numbers in brackets represent start and end amino acid of the protein fragment. *Right:* Schematic drawing of the artificial G5E1b promoter and the luciferase reporter gene in the plasmid pGAWG5E1b. **B-E.**

Luciferase activity assays. HEK-293 cells were either transiently or stably (3 μ g) transfected with a luciferase reporter (pGAWG5E1b) and co-transfected with 0.5 μ g β -galactosidase expression plasmid (pRSV- β -Gal) and 1 μ g of each Gal4 expression construct, as indicated. Transient transfections were performed by the jetPEI™ method (Biomol, Hamburg). Cells were harvested for reporter assays 48 hours after transfection. **B+D.** HEK-293 cells, transiently transfected with the luciferase reporter plasmid pGAWG5E1b. **C+E.** HEK-293 cells, stably transfected with the luciferase reporter plasmid pGAWG5E1b (HEK-293-pGAWG5E1b cells).

3.2 Generation of Sp2 gene targeted mice

3.2.1 Sp2 gene structure

To design a knockout vector for Sp2 gene targeting in the mouse, the exon-intron structure of the Sp2 gene had to be identified. For this, genomic Sp2 DNA sequence derived from the Celera data base (Celera access. no.: mCG 13240) was aligned with different Sp2 cDNA sequences present in the data base of the National Center for Biotechnology Information (NCBI). Exon-intron transitions were identified by the splice donor and acceptor site nucleotides GT/(GC) and AG. For this alignment, mouse (NCBI access. no.: BC021759 and NM_030220) as well as human cDNA sequences (NCBI access. no.: BC016680 and D28588) were used.

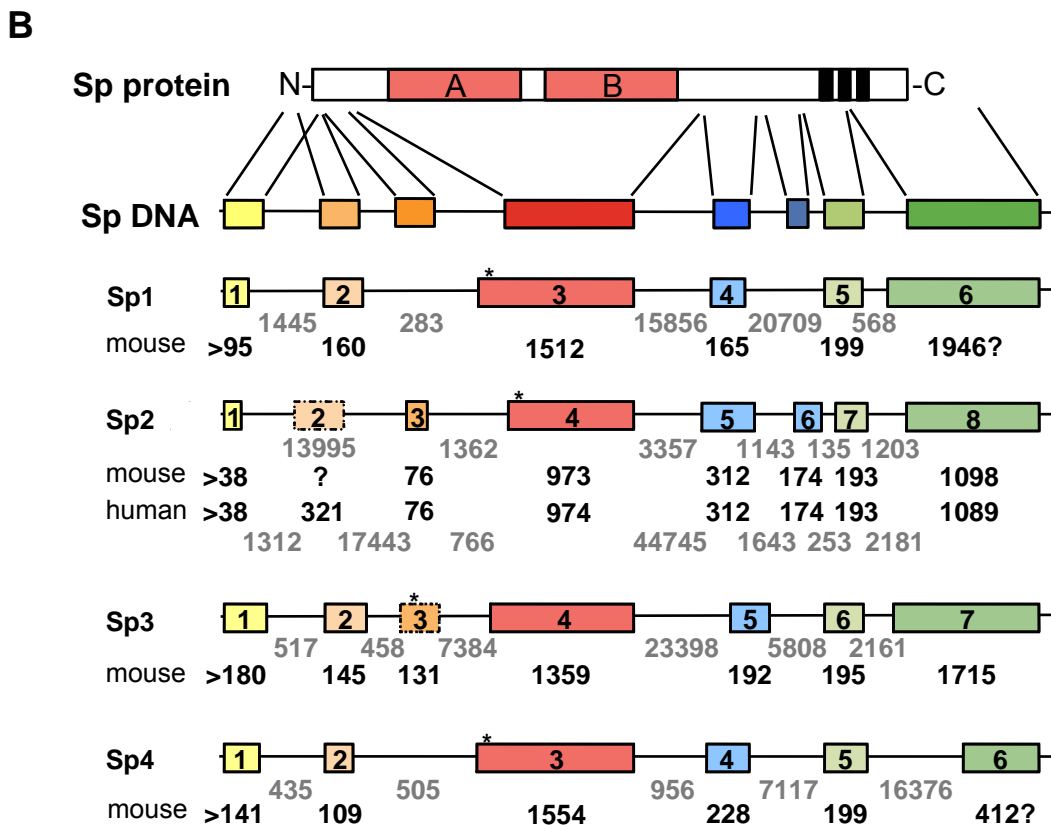
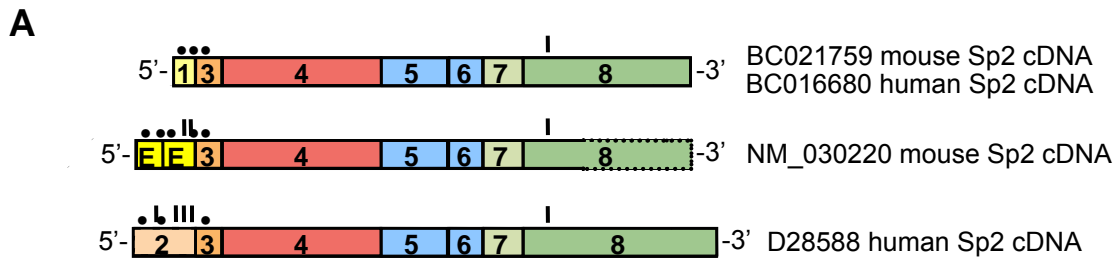


Fig. 3.2.1. Sp2 gene structure. A. Sp2 exon identification. Mouse genomic Sp2 DNA sequences (Celera data base; access. no.: mCG 13240) were aligned with different Sp2 cDNA sequences (NCBI data base) as indicated. Exons are numbered and coloured as described in B. Dots represent AUGs being in frame in the exons 1, 2, 3 and the two upstream exons (E); vertical lines represent stop codons. (Sequence of exon 8 is incomplete in the cDNA sequence NM_030220 (dotted line)). **B.** Exon-intron structure of Sp1-4 (modified after Krüger (diploma work, 2002)). Size of murine and human Sp2 gene exons (black numbers) and introns (grey numbers) are indicated. As the transcription start site in the first exon was not determined experimentally, only a minimal size could be calculated. The size of the last exon was determined until the start of the poly-adenylation sequence. As exon 2 is absent in the mouse cDNA sequences, size values could not be determined (= ?). Exon colour code: Red exon encodes the two glutamine-rich putative transactivation domains (A+B), yellow and orange exons code for the N-terminal protein region, blue exons for the region between glutamine-rich domains and zinc finger region, and green exons encode the C-terminal protein region including the zinc finger domain. The existence of the dashed exon 2 in mouse has to be proven. (*) represents the codons for the Sp box.

As depicted in Fig. 3.2.1.A, the murine and human Sp2 gene consists of at least 7 exons (exon 1 and 3-8). Exons 3-8 are present in all analyzed cDNAs, whereas exon 1 is only present in the Sp2 cDNAs no. BC021759 (mouse) and BC016680 (human) but not in NM_030220 (mouse) and D28588 (human). By sequence comparisons of Sp2 exon 1 with the first exon of Sp1, Sp3 and Sp4, a consensus sequence of ATG-AG/CC-G at the 3'-end of the exon could be identified (Krüger, diploma work, 2002). As the transcription start site in exon 1 is not known, only a minimal size could be calculated (see Fig. 3.2.1.B). Nevertheless, Sp2 mRNA and thus Sp2 protein is longer than originally described (Kingsley and Winoto, 1992).

Alignment of human Sp2 cDNA no. D28588 (but not BC016680) with mouse Sp2 genomic DNA (Celera access. no.: mCG 13240) supports the existence of a potential second exon around nt 16520-16660 (genomic Sp2 DNA) which is absent in the published mouse cDNAs. Probably this exon is a target of alternative splicing in mouse, which has to be proven experimentally. In addition, two potential upstream exons with unknown function (genomic Sp2 DNA: nt 10000-10106; nt 11368-11542) could be identified by sequence alignment of mouse Sp2 genomic DNA with cDNA no. NM_030220 (mouse).

Regarding translation, mouse (BC021759) and human (BC016680) Sp2 mRNA exhibit a relatively large number of putative translation start sites (in-frame AUGs) at

the 5'-region. The first start site, which not results in an uncompleted translation product, is located at the 3'-end of exon 1 (cDNA no. BC021759: nt 32 and BC016680, nt 31), according to the situation in Sp1, Sp3 and Sp4 (see above and Krüger; diploma work, 2002). This AUG is directly flanked downstream by two AUGs in exon 3 (e.g., cDNA no. BC021759: nt 47 and 53). The two putative upstream exons also contain three in-frame AUGs, however followed by two stop codons after 43 amino acids regarding the first AUG (see Fig. 3.2.1.A).

As depicted in Fig. 3.2.1.B, Sp2 displays a similar exon-intron structure as Sp1, Sp3 and Sp4: Exon 4 encodes the glutamine-rich putative transactivation domains and the characteristic Sp box (see 1.4), exons upstream of exon 4 for the N-terminal part of the protein, exon 7 for the first two and exon 8 for the third zinc finger as well as for the short region C-terminal to the third zinc finger. In contrast to Sp1, Sp3 and Sp4, the region between putative transactivation and DNA-binding domain is encoded by two exons, exon 5 and 6, the latter encoding the buttonhead box (see 1.4). It has to be emphasized that Sp2 (like Sp1, Sp3 and Sp4) exhibits not one but two glutamine-rich putative transactivation domains. This inconsistency towards previous results (e.g., Kingsley and Winoto, 1992) correlates with the finding that the Sp2 N-terminus is longer than previously described.

As depicted in Tab. 3.2.1, Sp2 displays high sequence identity on the protein as well as on the RNA level between mouse and human. Moreover, the overall identity between Sp2 and Sp1, Sp3 and Sp4 protein sequences is at least 24% and is mainly based on the highly conserved zinc finger domain.

<u>cDNA alignment</u>	<u>Identity [%]</u>
BC021759 (mouse) x BC016680 (human)	ca. 82%
<u>protein alignment</u>	
BC021759 protein (mouse) x BC016680 protein (human)	ca. 94%
BC021759 protein (mouse) x Sp1 protein sequence (mouse)	ca. 24% *
BC021759 protein (mouse) x Sp3 protein sequence (human)	ca. 25% *
BC021759 protein (mouse) x Sp4 protein sequence (mouse)	ca. 28% *

Tab. 3.2.1. Sequence identity [%] between murine and human Sp2 cDNA and protein as well as between Sp1-4 protein. DNA-Strider analysis. (*) Identity mainly based on the highly conserved zinc finger sequences.

Based on the high conservation of exon 3-8 in all analyzed cDNAs and on exon 1 homology between Sp 1-4, all experiments in this work requiring cDNA were performed by using mouse cDNA no. BC021759. The cDNA sequence together with the mouse Sp2 genomic DNA sequence is attached in 6.2. For all other cDNA sequences, the reader is referred to the NCBI data base.

3.2.2 Strategy to target the Sp2 gene in the mouse

Based on the analysis of the Sp2 gene structure (see 3.2.1), a knockout vector was designed to target the Sp2 gene in mouse for the investigation of Sp2 function *in vivo*. This vector exhibits the option for the generation of both, a constitutive and a conditional Sp2 knockout (Fig. 3.2.2.A).

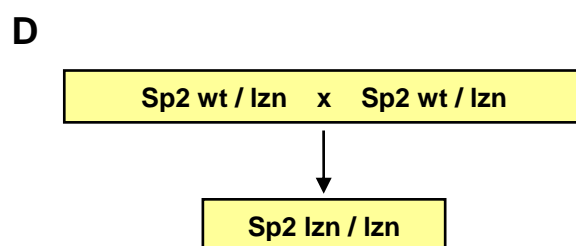
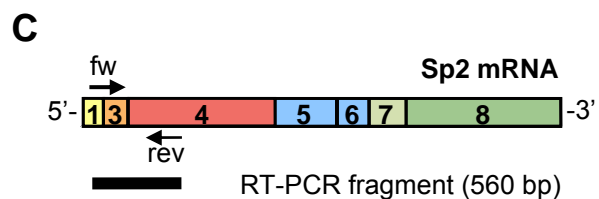
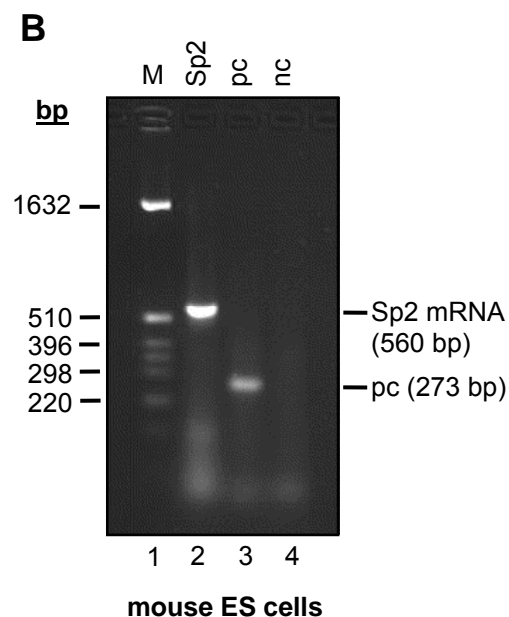
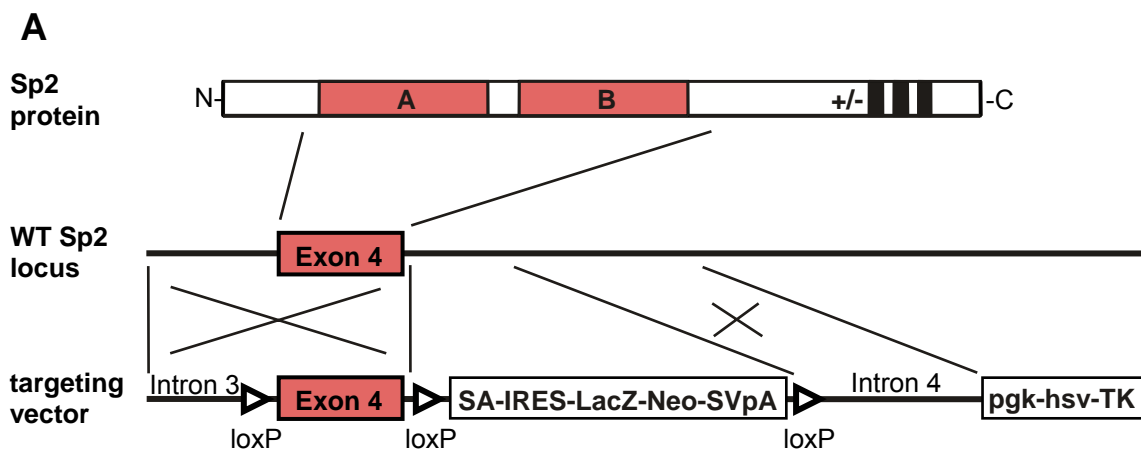


Fig. 3.2.2. Strategy to target the Sp2 gene in the mouse. **A.** Scheme of the Sp2 protein, wildtype locus and targeting vector. Zinc fingers of the DNA binding domain are depicted as black beams, glutamine-rich domains A and B as well as exon 4 encoding these domains are depicted in red. The floxed exon 4 as well as the floxed SA-IRES-LacZ-Neo-SVpA sequence replace exon 4 after homologous recombination. The LacZ-Neo gene is driven by the endogenous Sp2 promoter. Genomic fragments (intron 3 and 4), necessary for the homologous recombination, are indicated. The negative hsv-TK selection gene driven by a pgk promoter is located downstream of the homologous sequences. Triangles represent loxP sites enabling Cre-driven recombination. SA, splice acceptor site; IRES, internal ribosomal entry site; LacZ, gene encoding β -galactosidase; Neo, neomycin resistance gene encoding aminoglycoside phosphotransferase; SVpA, poly-adenylation site derived from Simian Virus; pgk, phosphoglycerate kinase promoter; hsv-TK, gene encoding the herpes simplex virus Thymidine Kinase. For further details concerning the targeting vector, the reader is also referred to chapter 3.2.4-5. **B.** RT-PCR with RNA derived from undifferentiated mouse ES cells to prove Sp2 gene expression. 5 μ l of each RT-PCR reaction were loaded on a 1.8% agarose gel. Primers MSp2-ES-RT-fw and -rev were used to detect Sp2 mRNA (560 bp fragment; lane 2). Primers against the myc gene were used as positive control (pc; 273 bp fragment; lane 3); the negative control reaction was performed without RNA (nc; lane 4). M, pBR-322 marker after *HinfI* digestion (lane 1). **C.** Scheme of RT-PCR strategy. Sp2 mRNA; exons 1-8 are indicated following the colour code of 3.2.1. Arrows represent primer MSp2-ES-RT-fw (= fw) and MSp2-ES-RT-rev (= rev) used to amplify Sp2 cDNA by PCR resulting in a 560 bp RT-PCR fragment (black beam). **D.** Mouse crossing scheme. Mice being heterozygous for the entire targeting vector (Izn/wt) were received from ES cell transfection, blastocyst injection and crossing of the resulting chimeric mice (see 1.5). The heterozygous mice then were crossed to homozygosity (Izn/Izn), exhibiting a functional knockout (see 3.2.8-9).

The knockout construct consists of a floxed exon 4, which encodes the two glutamine-rich domains A and B, flanked by two genomic regions (in the following named intron 3 and 4), necessary for the homologous recombination. Exon 4 has already been successfully used to generate the Sp3 knockout (Bouwman et al., 2000) and, for this, seems to be appropriate to target also Sp2.

Inside the homologous region, downstream of exon 4, a floxed selection cassette was inserted. The cassette consists of a promoter-less LacZ-Neo fusion gene (LacZ encodes β -galactosidase, Neo encodes the aminoglycoside phosphotransferase leading to neomycin resistance, thus enabling positive selection of the targeted embryonic stem cells), an internal ribosomal entry site (IRES) as well as a splice acceptor site (SA) upstream of the LacZ-Neo fusion gene, and a poly-adenylation sequence derived from the Simian Virus (SVpA) downstream of the LacZ-Neo fusion gene.

As the Sp2 gene is expressed in mouse embryonic stem (ES) cells (see Fig. 3.2.2.B), the LacZ-Neo gene could be set under the control of the endogenous Sp2 promoter. This has the advantage of an increased selection efficiency. Moreover, it opens the possibility of Sp2 expression studies by LacZ stainings (X-Gal, BluoGal) in the targeted animals (see 2.2.4.7).

The SA, SVpA and IRES elements inside the cassette are required for the transcriptional and translational control of the Sp2, LacZ and Neo gene. In the knockout mice, Sp2 exons 1-4 will be fused to the LacZ-Neo gene by the splice acceptor site (SA) during transcription. The poly-adenylation signal (SVpA) downstream of the Neo gene will stop transcription of this fusion product. Thus, Sp2 exons 5-8 should not be transcribed in the targeted animals. The IRES sequence enables an independent translation of the LacZ-Neo transcript, which allows to perform LacZ-based expression studies as well as positive selection by neomycin resistance.

Unlike the SA-IRES-LacZ-Neo cassette, the hsv-TK (herpes simplex virus Thymidine Kinase) gene, regulated by the phosphoglycerate kinase promoter (pgk), is located downstream of the homologous sequences, thus serving as negative selection marker (see 1.5).

As described in chapter 1.5, mouse embryonic stem (ES) cells were transfected with the targeting vector and selected by the two selection markers (Neo, hsv-TK). Positively selected ES cell were injected into blastocysts and chimeric mice, followed by mice being heterozygous (lzn/wt) for the entire construct, were generated. These mice were crossed to homozygosity (lzn/lzn) leading to a functional knockout (Fig. 3.2.2.D; see also 3.2.8-9).

Due to the floxed exon 4, the targeting vector also enables the generation of constitutive or conditional Sp2 knockout mice based on Cre-driven exon 4 deletion (see 1.5).

3.2.3 Screening for Sp2 genomic DNA

To receive genomic fragments of the murine Sp2 gene, a 129/ola mouse cosmid library at the RZPD Deutsches Ressourcenzentrum für Genomforschung GmbH (Heidelberg) was screened with a DNA probe against exon 4. The probe was amplified by PCR with primer MSp2-Ex4-RZPD-fw and MSp2-Ex4-RZPD-rev (see 2.1.9.3). In total, 15 clones were identified and four of them scrutinized by Southern Blot analysis. Clone number MPMGc121L17390Q2 positively tested by Southern Blotting was utilized for PCR amplification of the genomic fragments exon 4, intron 3 and intron 4 (targeting construct).

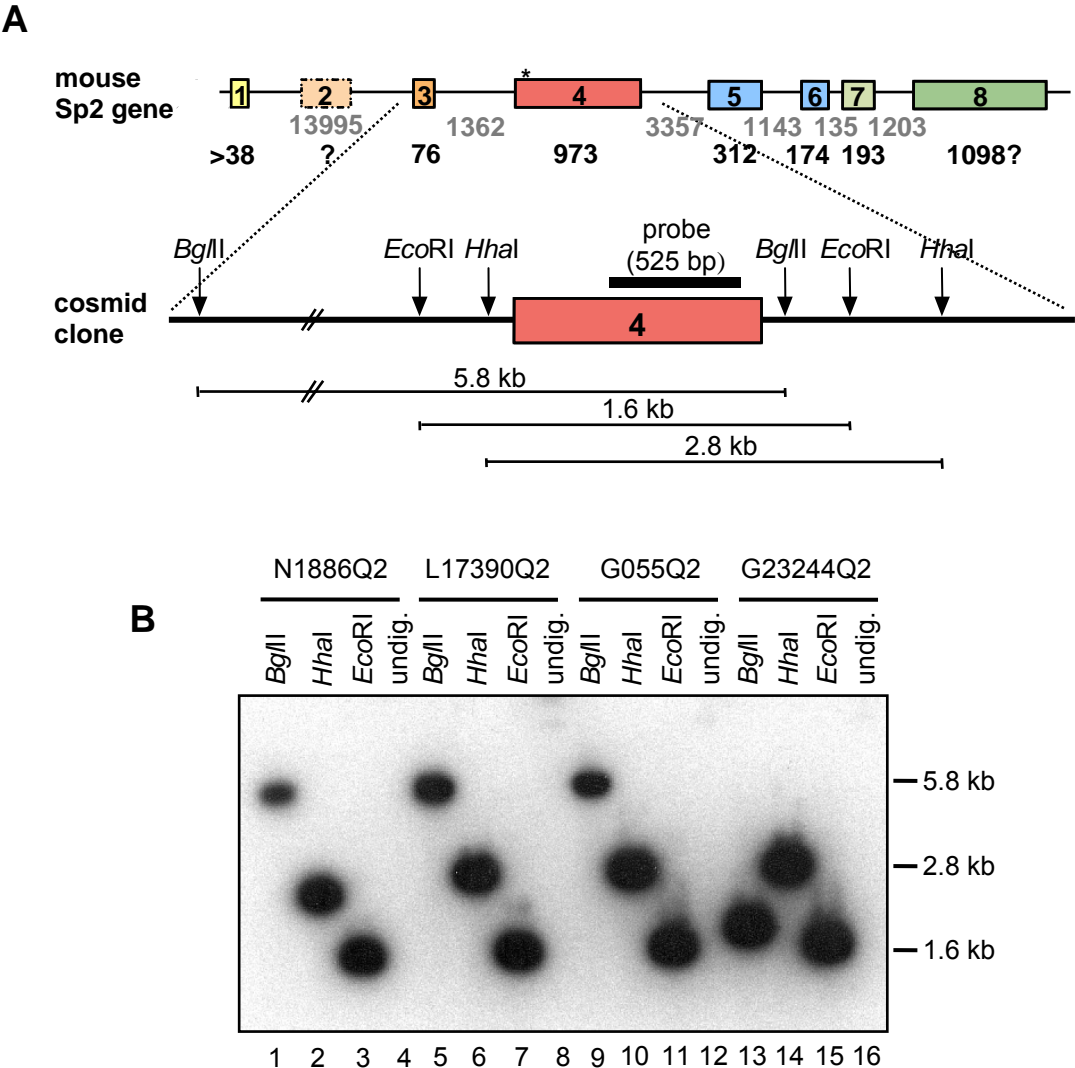


Fig. 3.2.3. Screening for Sp2 genomic DNA. A. Scheme of the Southern Blot strategy. A 129/ola mouse cosmid library was screened with a 525 bp DNA probe against exon 4 at the RZPD Deutsches Ressourcenzentrum für Genomforschung GmbH (Heidelberg). The identified cosmid clones were digested with either *BglII*, *HhaI* or *EcoRI* as indicated and hybridized with the labeled probe (Southern Blot). Clones containing exon 4 exhibit a 5.8 kb fragment for *BglII*, a 2.8 kb fragment for *HhaI* and a

1.6 kb fragment for *EcoRI* digestion. **B.** Southern Blot analysis of cosmid clones no. MPMGc121N1886Q2 (lane 1), MPMGc121L17390Q2 (lane 2), MPMGc121G055Q2 (lane 3), MPMGc121G23244Q2 (lane 4). The probe (Ex4-RZPD) was amplified by PCR with primer Sp2-Ex4-RZPD-fw and Sp2-Ex4-RZPD-rev (see 2.2.2.2.3).

3.2.4 Generation of the Sp2 knockout construct

To target the Sp2 gene in mouse, a knockout construct was designed displaying the option for the generation of both, a constitutive and a conditional Sp2 knockout (see 3.2.2). The starting vector for the Sp2 knockout was the pPNT plasmid (Fig. 3.2.4.A), containing a pgk-driven neomycin resistance (Neo) and herpes simplex Thymidine Kinase gene (hsv-TK) (Tybulewicz et al., 1991; see also Mangold, diploma work, 1995). The PCR-amplified and sequenced Sp2 genomic fragments intron 3, intron 4 and exon 4 (see 2.2.2.2.3 and Fig 3.2.4.B) as well as the three hybridized loxP sites (see 2.2.2.2.4) were subcloned into pPNT as described above (2.1.10.2.4) and depicted in Fig. 3.2.4.C. As a final cloning step, the pgk-driven neomycin resistance gene (Neo) of the pPNT plasmid was replaced by the En2-SA-IRES-LacZ-Neo-SVpA fragment (see 3.2.2) derived from plasmid pGT1,8-IRES- β Geo (Mountford and Skarnes, unpublished; see also Mangold, diploma work, 1995) (Fig. 3.2.4.A). To verify correct integration of the subcloned fragments, characteristic restriction digestions were performed and the constructs sequenced (data not shown). A scheme of the final Sp2 knockout construct pPNT-cSp2ko as well as all pre-constructs including fragment lengths are depicted in Fig. 3.2.4.C.

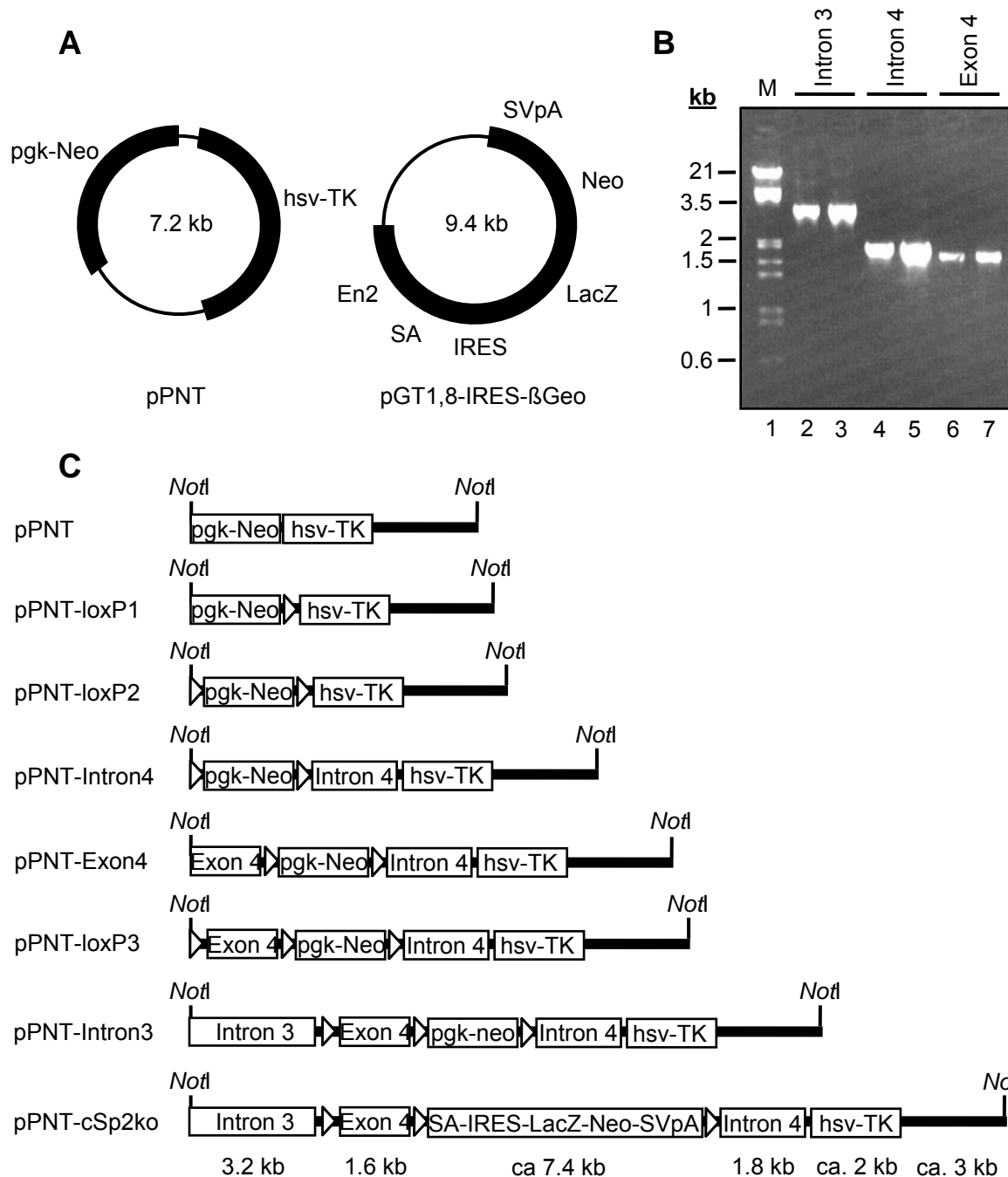


Fig. 3.2.4. Generation of the Sp2 knockout construct. **A.** Scheme of the starting plasmid pPNT (Tybulewicz et al., 1991) and the plasmid pGT1,8-IRES-βGeo harbouring the En2-SA-IRES-LacZ-Neo-SVpA fragment (Mountford and Skarnes, unpublished). **B.** PCR-amplified genomic fragments intron 3 (lane 2+3), intron 4 (lane 4+5) and exon 4 (lane 6+7). M, Lambda 1 marker after *EcoRI* and *HindIII* digestion (lane 1). 1% agarose gel. **C.** Scheme of the final Sp2 knockout construct (pPNT-cSp2ko) and all pre-constructs after *NotI* linearization. Triangles represent loxP sites. For details, the reader is referred to 2.1.10.2.4.

3.2.5 Functional analysis of Cre-driven loxP site recombinase capacity

To investigate whether the loxP sites present in the Sp2 knockout construct are functional, i.e. whether Cre recombinase is able to delete DNA sequences flanked by the loxP sites, a functionality assay in *E. coli* 294-Cre was performed as described in 2.2.2.2. For this purpose, the final Sp2 knockout construct pPNT-cSp2ko and the pre-construct pPNT-loxP2 were used. Sp2 knockout construct pPNT-cSp2ko contains three loxP sites flanking exon 4 and the En2-SA-IRES-LacZ-Neo-SVpA cassette (in total 9 kb). Pre-construct pPNT-loxP2 harbours two loxP sites flanking the pgk-driven neomycin resistance gene (1.8 kb). Both fragments should be deleted due to Cre recombinase expression after transformation into *E. coli* 294-Cre bacteria. As depicted in Fig. 3.2.5, loxP-flanked DNA sequences of both constructs are successfully deleted by Cre recombinase, i.e. all three loxP sites are functional.

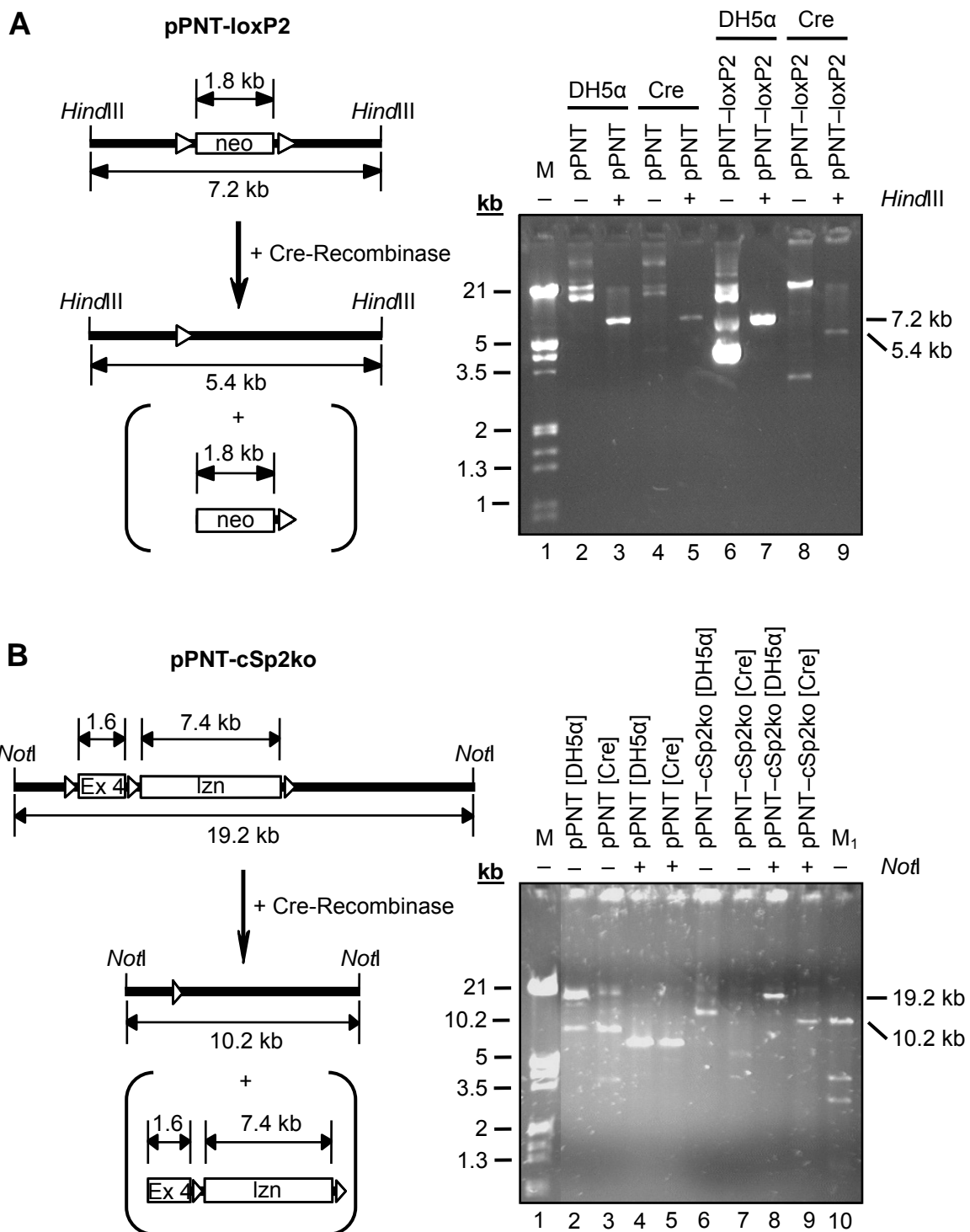


Fig. 3.2.5. Functional analysis of Cre-driven loxP site recombinase capacity. **A.** *E. coli* DH5 α and *E. coli* 294-Cre bacteria were transformed with either pPNT-loxP2 or the control plasmid pPNT. Plasmids were *Hind*III-digested and separated through a 0.6% agarose gel as indicated. As controls, undigested plasmids were loaded. Scheme represents situation before (above) and after Cre recombination (below) of pPNT-loxP2 in *E. coli* 294-Cre. Cre recombination leads to the deletion of the floxed Neo cassette (not visible in gel; depicted in brackets) resulting in a remaining 5.4 kb plasmid, visible in gel (lane 9). However, when transforming *E. coli* DH5 α with pPNT-loxP2, a ca. 7.2 kb fragment is expected after linearization (lane 7). In contrast to this, no Cre-driven recombination is observed for the loxP-less control plasmid pPNT (lane 5+3). M, Lambda 1 marker after *Eco*RI and

*Hind*III digestion (lane 1). **B.** The same experiment, described in (A), was performed with the final Sp2 knockout construct pPNT-cSp2ko and the loxP-less control plasmid pPNT. Instead of *Hind*III digestion plasmids were linearized by *Not*I. Expected pPNT-cSp2ko fragment sizes visible in gel are 10.2 kb for successful Cre recombination (lane 9) and ca. 19.2 kb without Cre recombination (lane 8). In contrast to this, no Cre-driven recombination is observed for the loxP-less control plasmid pPNT (lane 5+4). Samples were loaded as indicated. M, Lambda 1 marker after *Eco*RI and *Hind*III digestion (lane 1). M₁ pPac-cSp2ko after *Bam*HI digestion (lane 10). 0.4% agarose gel.

3.2.6 Sp2 gene targeting in mouse embryonic stem cells

Mouse embryonic stem (ES) cells were transfected with the conditional Sp2 knockout construct and selected for homologous recombination as described (2.2.4.5). Positive selected clones were tested by Southern Blotting (2.2.2.2.6) for homologous integration of the knockout construct (Izn/wt; see also 3.2.2 and 1.5) and used for blastocyst injections. Fig. 3.2.6 depicts strategy and results of ES cell genotyping by Southern Blot analysis.

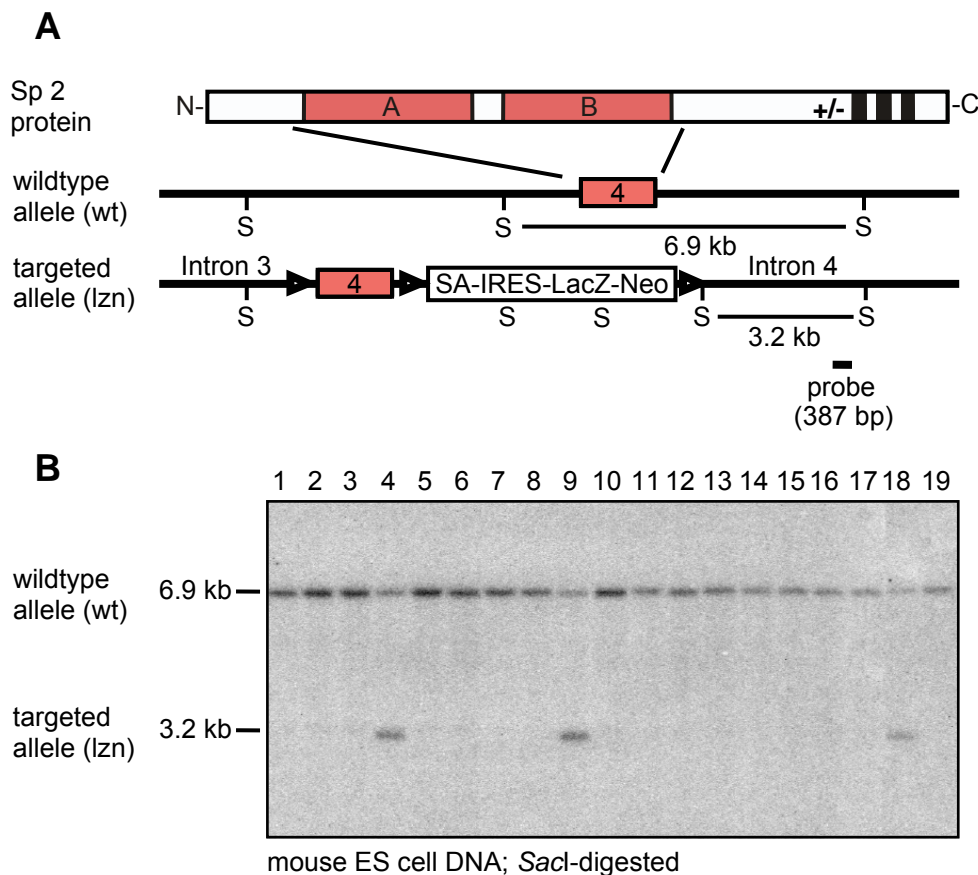


Fig. 3.2.6. Sp2 gene targeting in mouse embryonic stem cells. **A.** Southern Blot strategy for ES cell genotyping. Wildtype (wt) and targeted (Izn) Sp2 alleles are depicted. A 387 bp probe hybridizes with an intronic region downstream of exon 4 (= intron 4). *Sac*I (S) digestion of Sp2 heterozygous ES

cell DNA (Izn/wt) should result in a 6.9 kb fragment representing the wildtype allele (wt) and a 3.2 kb fragment representing the targeted Sp2 allele exhibiting homologous integration of the conditional Sp2 knockout construct (Izn). **B.** Southern Blot analysis of 19 targeted ES cell clones after *SacI* (S) digestion. The 387 bp probe (ES-a-SB) was obtained by PCR amplification using primer Sp2-ES-a-SB-fw and Sp2-ES-a-SB-rev (2.2.2.2.3).

3.2.7 Genotyping of constitutive Sp2 Izn/Izn knockout mice

Targeted embryonic stem (ES) cells were injected into blastocysts and implanted into pseudo-pregnant foster mice as described above (1.5 and 2.2.4.5) resulting in chimeric mice. Chimeric mice were crossed to heterozygosity (see 1.5 and 3.2.2). Sp2 heterozygous mice display one wildtype allele (wt) and one targeted allele (Izn) having replaced exon 4 by the conditional Sp2 knockout construct (floxed exon 4 + floxed En2-SA-IRES-LacZ-Neo-SVpA cassette). These mice were crossed to homozygosity (Izn/Izn) and genotyped by PCR (see 2.2.2.2.3). Fig. 3.2.7 depicts an example of genotyped mouse embryos.

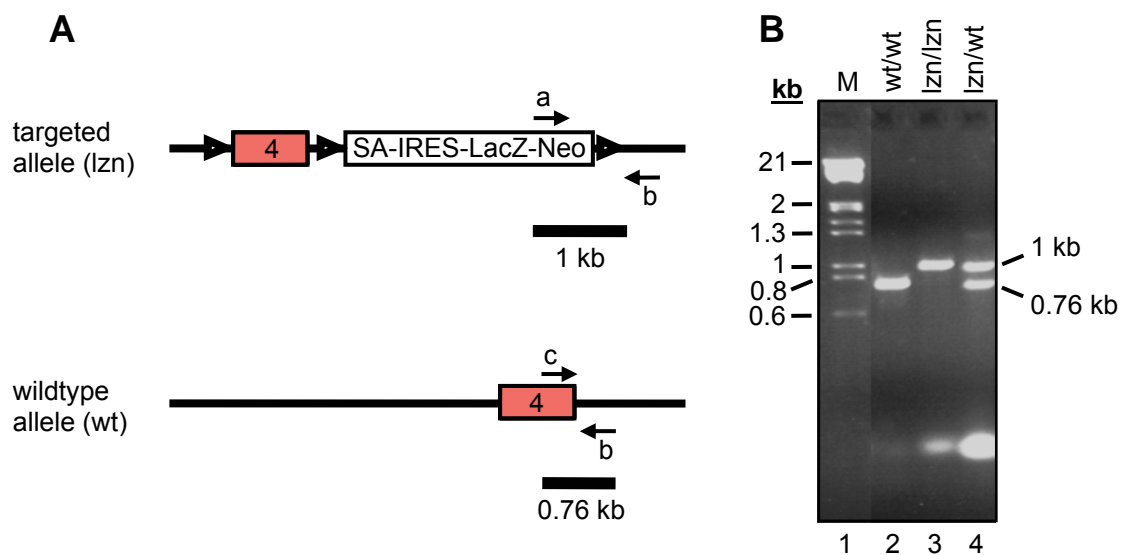


Fig. 3.2.7. Genotyping of targeted hetero- and homozygous Sp2 mice. **A.** Scheme of the PCR strategy. Annealing region of the three used primers and resulting PCR fragments are depicted for wildtype (wt) and targeted Sp2 allele (Izn). Primers MSp2Ex/L3Seq3 (c) and MSp2In2Seq4 (b) allow the detection of the Sp2 wildtype allele (0.76 kb fragment), primers Neo (a) and MSp2In2Seq4 (b) the detection of the targeted Sp2 locus (1 kb fragment). **B.** PCR with DNA templates derived from mouse embryo tail tips (see 2.2.2.2.1) and primers indicated above demonstrating wt/wt (lane 2), Izn/Izn (lane 3) and Izn/wt (lane 4) situation. Detailed PCR conditions are described in 2.2.2.2.3. M, Lambda 1 marker after *EcoRI* and *HindIII* digestion (lane 1). 1% agarose gel.

3.2.8 Sp2 expression in targeted Sp2 lzn/lzn knockout mice

PCR genotyping (3.2.7) demonstrated the correct targeting of the Sp2 gene in the mouse. To verify that the homozygous integration of the construct leads to a functional constitutive Sp2 knockout, Sp2 expression in the lzn/lzn mice was examined by Northern Blot analysis (3.2.8.1) and RT-PCR (3.2.8.2). Moreover, to study Sp2 protein expression during embryonic development, β -galactosidase activity measurements were performed by LacZ stainings (X-Gal, BluoGal) of day E12.5 mouse embryos (3.2.8.3).

3.2.8.1 Detection of a Sp2-lzn fusion mRNA in the targeted mice

To investigate whether the Sp2 knockout strategy leads to the loss of a functional Sp2 transcript, Northern Blot analyses were performed with total RNA from homozygous (lzn/lzn), heterozygous (lzn/wt) as well as from wildtype (wt/wt) mouse embryos (day E18.5). As depicted in Fig. 3.2.8.1, the expected Sp2 transcript of ca. 2.5-3 kb was detected in wildtype (wt/wt) and a ca. 6-8 kb transcript representing the fusion mRNA of the Sp2 exons 1-4 with the IRES-LacZ-Neo-SVpA cassette in homozygous (lzn/lzn) embryos. According to this, both transcripts are detectable in heterozygous embryos (lzn/wt).

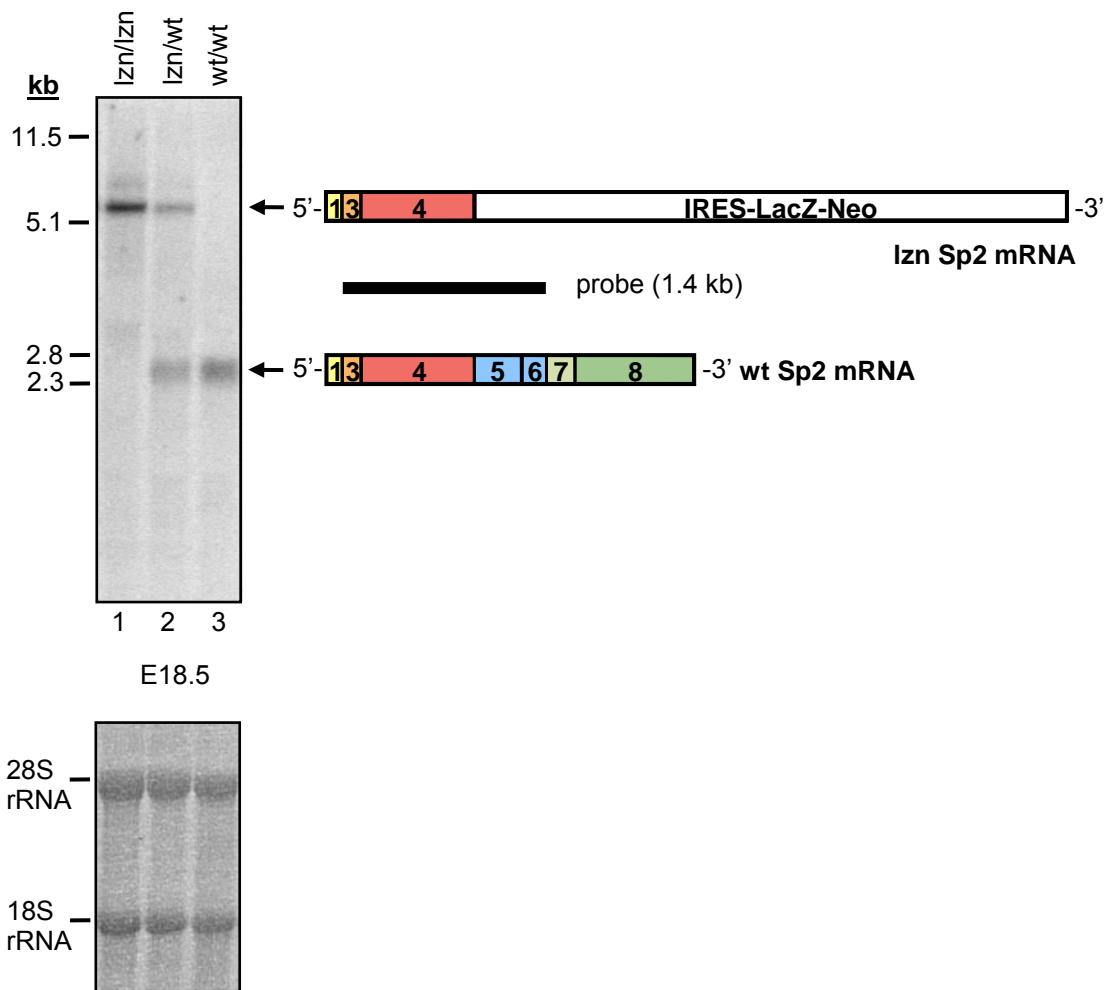


Fig. 3.2.8.1. Detection of a Sp2-Izn fusion mRNA in Sp2-targeted mice. Northern Blot analysis. 20 μ g of total RNA from day E18.5 embryos isolated by the LiCl/urea method (see 2.2.2.1.1) were separated through a 1% denaturing formaldehyde agarose gel, transferred to a nylon membrane and hybridized with a radioactive labeled probe against exon 3-6 (Ex 3-6 NB; see 2.2.2.1.4). 28S and 18S rRNA signals on the gel demonstrate comparable RNA loading amounts in all investigated samples. Wt, wildtype Sp2 allele; Izn, targeted Sp2 allele.

3.2.8.2 Absence of exon 5-8 in the targeted mice

The existence of a Sp2-LacZ-Neo fusion transcript requires to prove the absence of exons downstream of the LacZ-Neo cassette (exon 5-8) in this transcript. For this, an RT-PCR experiment was performed with RNA derived from fetal liver and brain (day E18.5) using primers against exon 5 and 6 (see 2.2.2.1.3). In wildtype but not Izn/Izn embryos these primers should anneal within exon 5 and 6, thus resulting in a 228 bp PCR fragment. As expected, Fig. 3.2.8.2 demonstrates the presence of exon 5 and 6 sequences in the wildtype (wt/wt) but not in the targeted (Izn/Izn) fetal transcript.

Consistently, a weak 228 bp signal appears with RNA from heterozygous embryos (lzn/wt).

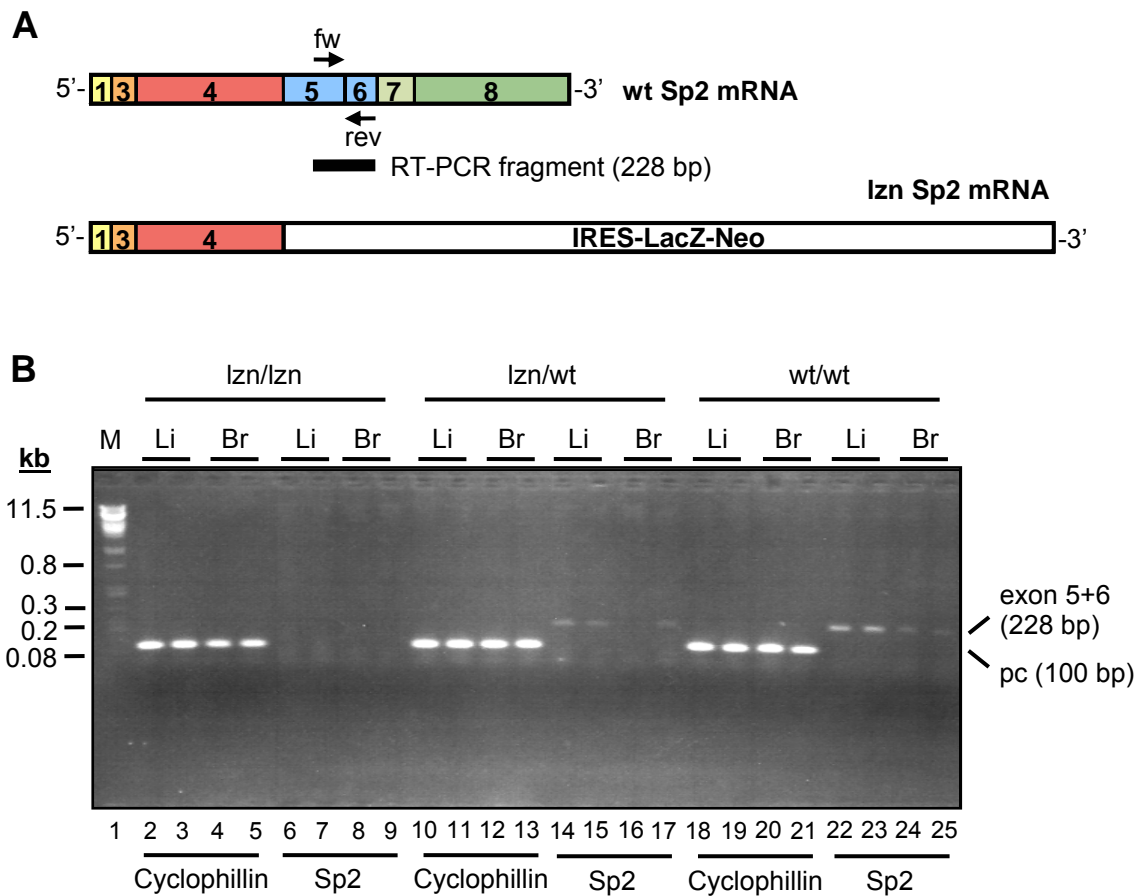


Fig. 3.2.8.2. Absence of exon 5-8 in the targeted lzn/lzn mice. A. Scheme of RT-PCR strategy. Wildtype Sp2 mRNA (wt) and targeted Sp2-lzn fusion mRNA (lzn) are depicted. Exons 1-8 are indicated following the colour code of 3.2.1. Arrows represent primer Sp2-Ex5-RT-fw (= fw) and Sp2-Ex6-RT-rev (= rev) that should amplify exon 5 and 6 in the Sp2 wildtype (wt) but not in the targeted (lzn) RNA, resulting in a 228 bp RT-PCR fragment (black beam). **B.** RT-PCR with RNA obtained from liver (Li) and brain (Br) of lzn/lzn (lanes 2-9), lzn/wt (lanes 10-17) and wt/wt (lanes 18-25) mouse embryos (day E18.5) using primer Sp2-Ex5-RT-fw and Sp2-Ex6-RT-rev (for detailed information, see 2.2.2.1.3). As internal reaction control (pc), primers Cyclophilin-fw and -rev directed against sequences of the cyclophilin gene were used resulting in a ca. 100 bp fragment. Each reaction was performed in duplicates. M, Lambda 1 marker after *Pst*I digestion (lane 1). 1.6% agarose gel.

3.2.8.3 Endogenous Sp2 expression in mouse embryos

To determine Sp2 expression pattern during embryogenesis, β -galactosidase activity in day E12.5 mouse embryos was investigated by X-Gal and BlueGal staining of lzn/lzn, lzn/wt and wt/wt embryos (see 2.2.4.7). Lzn/lzn embryos are characterized by

integration of the conditional knockout construct at the Sp2 locus of both alleles (see 1.5 and 3.2.2), thus exhibiting the β -galactosidase gene (LacZ) under the control of the endogenous Sp2 promoter. If the Sp2 promoter is active in these embryos, the LacZ gene is expressed and β -galactosidase activity can be detected by X-Gal or BluoGal stainings of the embryos. Thus, all blue-stained embryonic tissues or body regions display Sp2 expression. On the other hand, lzn/wt embryos exhibit homologous integration of the conditional Sp2 knockout construct only on one allele. For this, a weaker β -galactosidase activity should be observed. In contrast to that, no activity should be detected in homozygous wildtype (wt/wt) embryos. Fig. 3.2.8.3 demonstrates that Sp2 is widely expressed in day E12.5 embryos. Moreover, the intensity of X-Gal and BluoGal staining is lower in the heterozygous lzn/wt and absent in the wt/wt embryos, as expected.

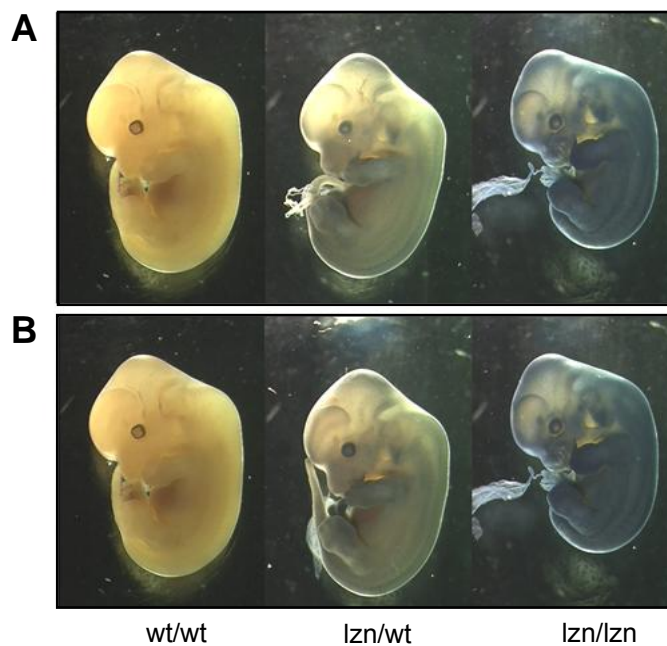


Fig. 3.2.8.3. LacZ expression in day E12.5 mouse embryos. Lateral view of day E12.5 mouse embryos. Blue colour represents β -galactosidase activity, i.e. Sp2 expression in the corresponding tissue or body region. Wt, wild type Sp2 locus; lzn, targeted Sp2 allele. **A.** X-Gal staining. **B.** BluoGal staining.

3.2.9 Preliminary characterization of targeted Sp2 lzn/lzn mice

Two aspects were investigated to characterize Sp2 lzn/lzn knockout mice being available at the end of the thesis period: viability (3.2.9.1) and the occurrence of growth abnormalities during embryonic development (3.2.9.2).

3.2.9.1 Post- and pre-natal viability of targeted Sp2 lzn/lzn mice

Heterozygous Sp2 (lzn/wt) knockout mice are viable, reproduce normal and display no obvious phenotype after birth. Also no obvious abnormalities during embryonic development were detectable. In contrast to this, homozygous Sp2 lzn/lzn knockout mice exhibit a strongly reduced viability: only 1 out of 73 born mice are lzn/lzn (Tab. 3.2.9.1). This embryo died during the first weeks after birth (Nynke Gillemans, personal communication). To find out whether homozygous Sp2 lzn/lzn knockout mice already die during embryonic development, embryos of three different developmental time points (E18.5, E12.5 and E9.5) were analyzed. As depicted in Tab. 3.2.9.1, Sp2 lzn/lzn embryos exhibit a normal Mendelian distribution at all three time points. Thus far, no dramatic effect of Sp2 protein deficiency on viability is visible in these animals. However, Sp2 lzn/lzn embryos at day E18.5 slightly deviate from the Mendelian distribution but probably due to the low number of analyzed samples.

	<u>Sp2 lzn/lzn</u>	<u>Sp2 lzn/wt</u>	<u>Sp2 wt/wt</u>	<u>total number</u>
born	1 (1.4%)	52 (71.2%)	20 (27.4%)	73
day E18.5	10 (18.5%)	31 (57.4%)	13 (24.1%)	54
day E12.5	14 (24.6%)	31 (54.4%)	12 (21.1%)	57
day E9.5	7 (24.1%)	15 (51.7%)	7 (24.1%)	29
expected distribution	25%	50%	25%	

Tab. 3.2.9.1. Post- and pre-natal viability of targeted Sp2 lzn/lzn mice. Genotype distribution in [n] and [%] of either new-born Sp2 mice or day E18.5, E12.5 and E9.5 embryos. Mouse genotyping was performed as described in 3.2.7 and 2.2.2.2.3. Wt, wildtype Sp2 allele; lzn, targeted Sp2 allele.

3.2.9.2 Reduced growth in day E18.5 Sp2 *lzn/lzn* embryos

Consistent with the findings, that Sp2 deficiency influences viability of new-born and probably also day E18.5 embryos, a significant reduction in body size and weight of day E18.5 Sp2 knockout (*lzn/lzn*) embryos compared to wildtype (*wt/wt*) and heterozygous (*lzn/wt*) embryos is detectable, however with a strong variation (Fig. 3.2.9.2.A-B). This reduction is associated with a developmental retardation, at least in some of the *lzn/lzn* embryos (see Fig. 3.2.9.2.B). In addition, the body size at earlier developmental time points (E12.5 and E9.5) was determined. However, no statistically significant differences between knockout and wildtype or heterozygous embryos were observed (Fig. 3.2.9.2.A+C). This favours the conclusion that Sp2 deficiency effects normal mouse development in later embryonic stages (at least E18.5). However, the mechanisms leading to this Sp2-based abnormal development have to be explored.

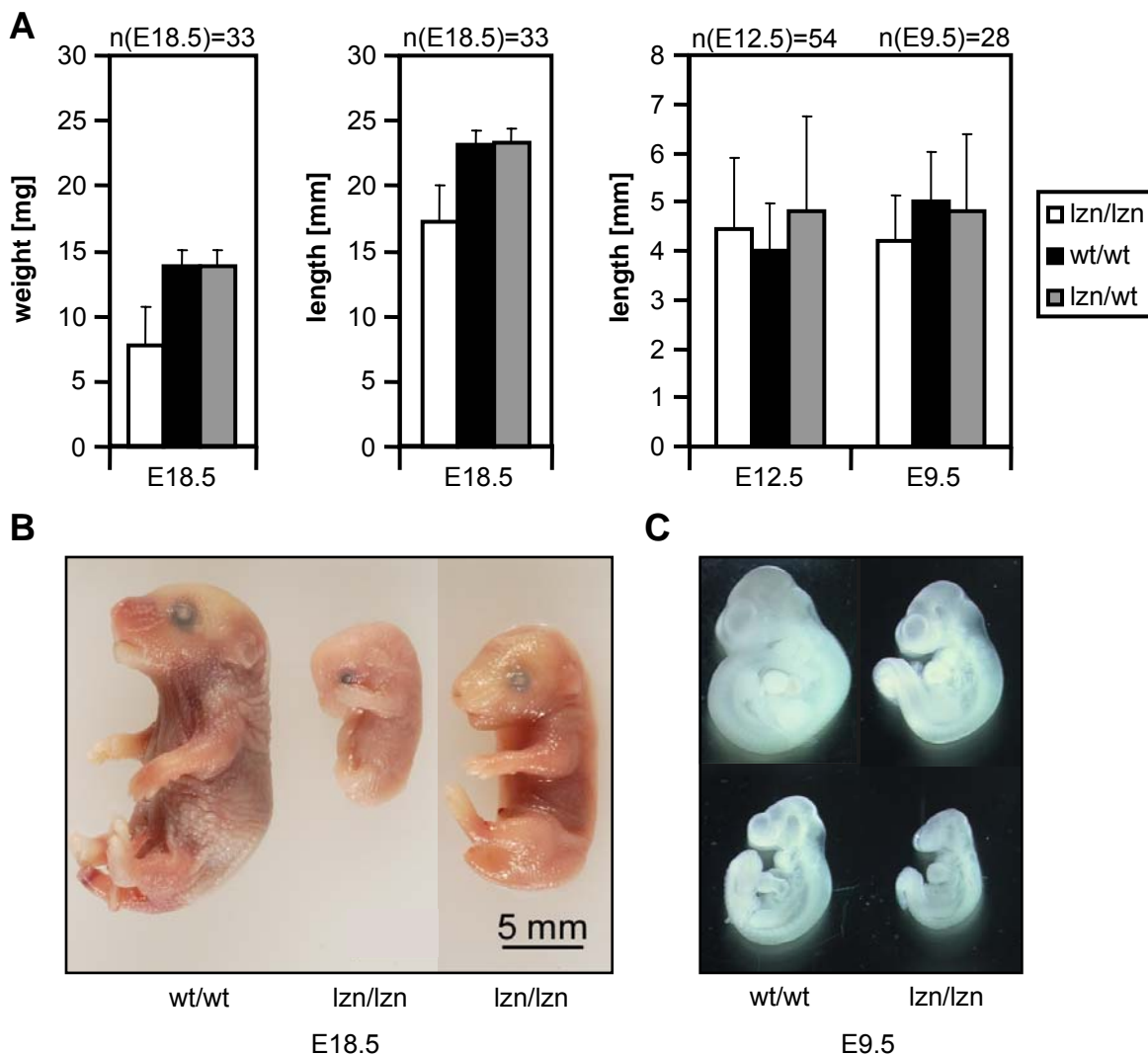


Fig. 3.2.9.2. Growth reduction in day E18.5 Sp2 lzn/lzn embryos. **A.** Weight [mg] and size distribution [mm] in day E18.5 mouse embryos as well as size distribution in E12.5 and E9.5 mouse embryos. Wt, wildtype Sp2 allele; lzn, targeted Sp2 allele. N, number of analyzed embryos per each developmental time point. **B.** Example of wildtype (wt/wt) and Sp2-targeted (lzn/lzn) mouse embryos of embryonic day E18.5. Lateral view. **C.** Examples of wildtype (wt/wt) and Sp2-targeted (lzn/lzn) mouse embryos of embryonic day E9.5. Lateral view.

4. Discussion

4.1 Structure comparisons between Sp2 and the glutamine-rich family members Sp1, Sp3 and Sp4

Many data exist supporting the fundamental importance of the glutamine-rich Sp family members Sp1, Sp3 and Sp4 in a diverse set of biological processes (Philipsen and Suske, 1999; Bouwman and Philipsen, 2002). However, until now, such information lacks for the transcription factor Sp2, which represents the less conserved member of the glutamine-rich Sp subgroup. The studies reported herein were conducted to initiate the characterization of structural and functional properties of the transcription factor Sp2 on the molecular as well as on the physiological level.

Concerning the gene and protein structure, our analyses demonstrate high similarities between transcription factor Sp2 and the glutamine-rich family members Sp1, Sp3 and Sp4. The Sp2 gene is characterized by a similar exon-intron structure as it is described for these factors. The two most downstream exons of the Sp2 gene (exon 7 and 8; see Fig. 3.2.1.B) encode the characteristic DNA-binding domain, a large central exon (exon 4) codes for the two glutamine-rich and serine/threonine-rich domains, and the exons upstream of exon 4 encode the N terminal part of the protein.

Specific for the Sp2 gene is the existence of two exons (exon 5 and 6; see Fig. 3.2.1.B) encoding the protein region between domain B and the DNA-binding domain. In Sp1, Sp3 and Sp4 this region is encoded by only one exon. Whereas exon 6 encodes the so-called Btd box (see 1.4 and 3.2.1) and displays homology towards the corresponding exon in the remaining glutamine-rich Sp members (Sp1, Sp4: exon 4; Sp3: exon 5; see Fig. 3.2.1.B), the evolution and function of exon 5 has to be clarified. However, it is conspicuous that the length of exon 5 plus exon 4 together (the latter encoding domain A and B and being smaller in size as the corresponding gene in Sp1, Sp3 and Sp4) is comparable with the length of the exon encoding domain A and B in Sp1, Sp3 and Sp4. Supposing Sp2 as the evolutionary most basal factor among the glutamine-rich Sp factors (Kolell and Crawford, 2002; Suske, 1999), it might be that the large exon coding for the transactivation domains

in Sp1, Sp3 and Sp4 evolved from two individual exons. Whether Sp2 exon 5 displays specialized function, has to be explored.

Also specific for the Sp2 gene is one exon, which is present in the human cDNA upstream of exon 3 but not in the published mouse cDNA sequences. Whether this exon exists in mouse and whether it is target of alternative splicing has to be determined. The occurrence of alternative splicing of the third exon in Sp3 (Krüger, diploma work, 2002) could be a hint that such an event is also expectable for Sp2.

In addition, two potential exons upstream to exons 1 (the latter harbouring the translation start codon; see Fig. 3.2.1.A) were identified in Sp2 by alignment with the mouse cDNA NM_030220. Although these exons also contain several AUG codons, the reading frames are closed by stop codons. Moreover, they are not present in the other Sp2 cDNAs. Thus, it remains to be clarified whether these are “real” exons.

Apart from these structural differences on the gene level, the similarities in the exon-intron structure as well as the corresponding functional protein domains support the relationship between Sp2 and the glutamine-rich family members Sp1, Sp3 and Sp4. On protein level, Sp1-4 share several functional domains like the aforementioned highly conserved DNA binding domain, a stretch of 81 amino acids forming three C2H2 zinc fingers, and the glutamine- and serine/threonine-rich regions. Contrary to previous reports (e.g. Kingsley and Winoto, 1992; Bouwman and Philipsen, 1992), Sp2 protein also possesses two glutamine-rich domains (domain A and B) and two serine/threonine-rich regions.

Different to the glutamine-rich Sp factors is the lack of the D domain in Sp2, a region located C-terminally to the zinc fingers. In Sp1, domain D together with the two transactivation domains A and B are required for a synergistic transcription activation by binding to multiple sites (Pascal and Tjian, 1991). In theory, this function should be absent in Sp2. However, its ability to form multimers via the glutamine-rich B domain like in Sp1 (Pascal and Tjian, 1991) should be present. As a result of the absence of the D domain, the Sp2 protein is shorter in size as Sp1, Sp3 and Sp4. This again supports its evolutionary most distant position within the glutamine-rich Sp subgroup.

To summarize the analysis of the mouse Sp2 gene and protein structure, it is to retain that Sp2 protein shares many of the characteristic structural features with the related glutamine-rich Sp family members Sp1, Sp3 and Sp4 like DNA binding and glutamine-rich domains. However, as described above, also differences are observed when comparing protein and gene structure. Nevertheless, the assumption that Sp2 should not be recognized as a member of the Sp family, which is favoured by Kolell and Crawford (2002) based on the H→L substitution in the first zinc finger and the assumed absence of domain A, can be excluded. Whether and how the structural distinctions influence the functional specificity of Sp2 has to be unravelled. The results of the functional *in vitro* and *in vivo* studies presented in this thesis work suggest that these structural differences indeed affect Sp2 function (as it will be discussed in the following chapters).

4.2 Regulation of Sp2 DNA binding capacity and transactivation properties

When initiating this thesis work, no functional data about the transcription factor Sp2 were available. To analyze Sp2 function, the ability of Sp2 to activate reporter gene expression was investigated. Based on the structural similarities towards the other glutamine-rich family members (zinc finger domain, glutamine-rich regions, etc.) and the results from expression and subcellular localization studies, a similar binding affinity towards the classical Sp1 binding site (GC box) was expected. Therefore, reporter constructs were used, in which the expression of the reporter gene was controlled by a promoter containing Sp1 binding sites. However, no Sp2-driven activation was detectable.

Investigation of the Sp2 capacity to bind to GC-box-containing DNA sequences resulted in the finding that full-length Sp2 protein –although highly expressed– is unable to bind to the classical Sp1 binding site. Also when using DNA sequences harbouring other binding motifs, full-length Sp2 failed to bind. However, when analyzing N-terminally truncated Sp2 protein fragments, it turned out that those fragments lacking the first 179 N-terminal amino acids are able to bind DNA (GC boxes). This led to the assumption that the N terminus somehow is involved in the regulation of Sp2 DNA binding capacity and that the exchange of a critical amino acid in the first zinc finger does not influence sequence specificity as it was previously

postulated (e.g. Kingsley and Winoto,1992; Moorefield, 2004). Yet, although binding DNA, the N-terminal Sp2 deletions did not activate reporter gene expression.

Unexpected results were obtained, when C-terminally truncated, zinc-finger-less Sp2 deletion mutants were fused at the N terminus to a Gal4 DNA-binding domain. Those Sp2 fragments containing the first 179 N-terminal amino acids were able to bind the Gal4 DNA-binding site but exhibited an extremely retarded migration in gel electrophoresis (see Fig. 3.1.5.7.B). However, when analyzing a Gal4-Sp2 fusion protein fragment lacking these amino acids, no binding signal for this fragment was detectable. In addition, compared to a Gal4-Sp1 control, the Gal4-Sp2 deletions did not activate reporter gene expression.

These results suggest that a putative regulatory function is not only restricted to the N-terminal part of the Sp2 protein but also involves a region located around the B domain. Assuming this, three questions arise that should be discussed: (1) What is the general nature of the regulating mechanism; (2) why are the Gal4-Sp2 fusions containing the N-terminus as well as the region around the B domain (e.g. fragment Gal4-Sp2 (aa6-471)) still able to bind DNA; and (3) why does a N-terminal deletion of the first 206 amino acids (fragment Sp2 (aa207-606)) rescue from binding inhibition but not when the same Sp2 fragment is fused at the N terminus to a Gal4 DNA-binding domain (fragment Gal4-Sp2 (aa207-471))?

Regarding the general mechanisms regulating Sp2 DNA binding activity, several putative alternatives are imaginable, including regulation by posttranslational modification and/or protein-protein interactions, either inhibiting or activating the DNA binding capacity. As data bank research for known protein modification or interaction motifs in the Sp2 sequence did not reveal any helpful information, the further discussion is rather speculative.

Regulation of protein activity by posttranslational modifications is a well-known mechanism and examples can be found within the Sp family members. Transcription factor Sp1 activity, e.g., is regulated by phosphorylation and glycosylation. Depending on cell type and stimuli, phosphorylation of Sp1 protein can either increase (e.g. Merchant et al., 1999; Rohlff et al., 1997), decrease (e.g. Armstrong et al., 1997; Borellini et al., 1990; Leggett et al., 1995), or not affect (Jackson et al.,

1990) Sp1 DNA binding activity. Reduced O-GlcNAcylation of Sp1 protein is reported to result in a decreased Sp1 binding activity (Han and Kudlow, 1997). However, other reports correlate increased O-GlcNAcylation of Sp1 protein with a positive (Han and Kudlow, 1997; Du et al., 2000) or negative (Yang et al., 2001) regulation of Sp1 activity.

Another possibility of regulating Sp factor activity by posttranslational modification is the inhibition of Sp3 transactivity by SUMO. SUMO regulates Sp3 transactivation capacity by attaching to a critical lysine residue within an inhibitory domain (Dennig et al., 1996; Sapetschnig et al., 2002). Based on its functional variety (reviewed in Seeler and Dejean, 2003), it is conceivable that SUMO also regulates DNA binding specificity of transcription factors. However, no classical SUMO site is present in the Sp2 protein.

Whether and how Sp2 DNA binding capacity is regulated by posttranslational modifications has to be explored. Western Blot analyses of endogenous Sp2 expression in various cell lines resulted in two differently migrating protein species detected by the Sp2-specific antiserum (see Fig. 3.1.2.1-2). Apart from reflecting the existence of alternative splice variants or translationally controlled protein isoforms, this might also be a hint for posttranslational modifications. It is conceivable, that a modification influences protein conformation and by this enables the interaction of Sp2 protein with the DNA. On the other hand, it is possible that a modification like e.g. glycosylation somehow masks the zinc finger domain thus preventing Sp2 from binding. However, it is difficult to explain, why the deletion fragment Sp2 (aa207-606) is able to bind DNA whereas the same construct N-terminally fused to a Gal4 DNA-binding domain is unable to interact.

Another putative mechanism regulating Sp2 DNA binding capacity would be based on protein-protein interactions. An interacting protein could either function as activator or inhibitor of Sp2-DNA interaction. Our results obtained from DNA binding studies with N-terminal truncated Sp2 protein fragments would suggest the existence of an inhibitory protein binding within the first 179 N-terminal amino acids of Sp2 protein. This inhibitor could prevent Sp2 from binding to DNA by conformational changes and/or by masking the zinc finger domain. As described above, our results obtained from the Gal4-Sp2 studies, however, indicate that also a region around the

B domain is involved in the regulation process. Moreover, the fact that Sp2 DNA binding inhibition occurs as well in insect (*Drosophila* SL2) and mammalian (HEK-293) cells favours the existence of a protein that is highly abundant.

During the period of this thesis work, Horowitz and co-workers published results on transactivation and DNA binding properties of Sp2 (Moorefield et al., 2004). Consistent with our findings, they observed that SL2-overexpressed Sp2 protein does not or at the most very weakly activate the Hamster DHFR promoter. According to this, Sp2 protein that has been overexpressed by recombinant baculoviruses in Sf9 insect cells did only bind to these sequences when being used in a 20-50 fold excess compared to Sp1- or Sp3-containing extracts. Also when using several GC-rich oligonucleotide variants, Sp2-DNA interactions remained comparably low. In addition, Sp2 protein was observed to be expressed in many human and mouse cell lines. However, no or only low binding affinity was apparent in extracts prepared from these cells. When incubating recombinant Sp2 protein derived from baculovirus-infected Sf9 cells with increasing amounts of the mammalian cell extracts before analyzing its DNA binding activity, the weak interactions between DNA and the recombinant Sp2 protein could be totally blocked. Western Blot analyses of the extract mixes demonstrated that the loss of DNA binding capacity was not due to Sp2 protein degradation. From this, Horowitz and co-workers concluded that Sp2 DNA binding activity is inhibited by one or more proteins in mammalian cells, thus supporting our results. By protein-protein binding assays, two interacting proteins were identified, a 84 kDa protein (p84) specifically binding to the glutamine-rich domains of Sp2 and a 74 kDa protein (p74) also binding to the transactivation domains of Sp1 and Sp3. Whereas p84 function is unclear, p74 is reported to decrease the transactivation capacity of at least Sp1 (Murata et al., 1994). Whether p74 also negatively influence the Sp2 DNA binding capacity and whether it functionally interacts with p84, has to be explored. Moreover, Horowitz and co-workers could not prove whether p84 inhibits Sp2-DNA interactions.

Besides the regulation by one or two inhibitory proteins binding to two distinct sites within the Sp2 protein, it is also conceivable that one protein contacts Sp2 by two or more binding sites, e.g. within the first 179 N-terminal amino acids as well as the N-terminal part of the B domain. As depicted in Fig. 4.2, the DNA binding incapability of Sp2 fragments containing the first 179 N-terminal amino acids on the one hand and

the ability for DNA interactions of those fragments lacking these amino acids on the other hand thus would be explainable. Assuming one contact sequence being located around the B domain, it could be suggested that a small percentage of the fragments Sp2 (aa180-606) (not depicted here) and Sp2 (aa207-606), which contain parts of the sequence, still interact with the inhibitor. This would explain the gradual effect observed in gelelectrophoresis for these fragments in comparison to the fragments completely lacking the N-terminal part of domain B (see Fig. 3.1.5.6).

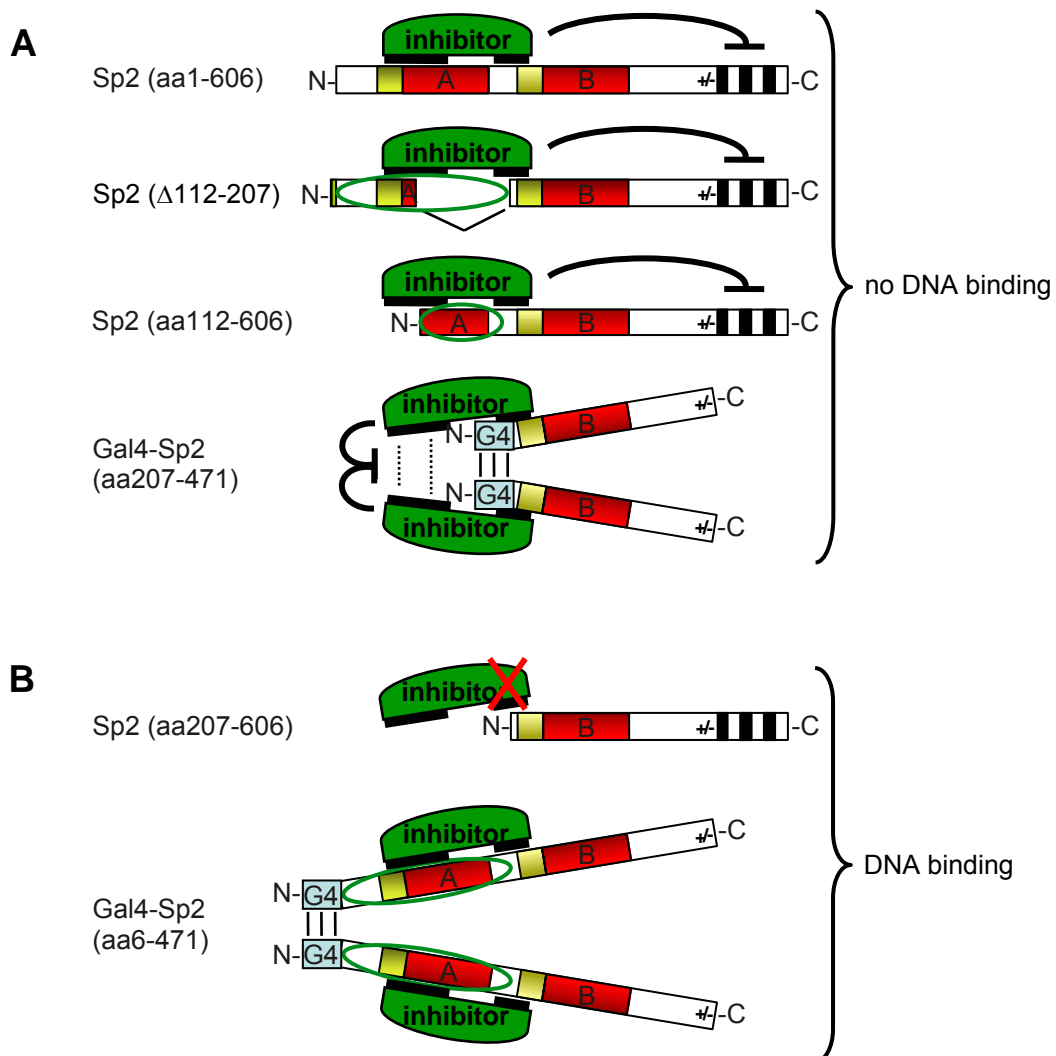


Fig. 4.2. Hypothetical regulation model of Sp2-DNA interactions. A. This model suggests the existence of an inhibitory protein (depicted in green) regulating Sp2-DNA interactions. The inhibitor contacts Sp2 through two putative binding sites located at the N-terminal parts of domain A and B. Probably due to conformational changes and/or masking of the zinc finger domain, this prevents Sp2 from contacting the DNA. In the case of Gal4-Sp2 (aa207-471), the inhibitor protein binds to the N-terminal region of the B domain of each individual Gal4-Sp2 molecule dimerized through the Gal4 DNA-binding domain. These interactions mask the Gal4-DNA binding sites thus leading to DNA binding incapability. Dotted lines represent putative interactions between two inhibitor molecules. **B.** In

the fragment Sp2 (aa207-606), the inhibitor is unable to bind (indicated as red cross) due to the lack of the second binding site. This can not be compensated by either a second inhibitor molecule or the Gal4 DNA-binding domain like in the corresponding Gal4 fusion fragment. The interaction with the dimerized Gal4-Sp2 (aa6-471) fragments should not influence their DNA binding capacity but results in a reduced mobility of the fragments during gelelectrophoresis due to the interaction with the two inhibitor proteins. **A-B.** Green ellipse point out the N-terminal amino acids being present in the Sp2 deletion mutants exhibiting no DNA binding capacity in EMSA experiments (see 3.1.5.6).

Whereas GC boxes can be bound by transcription factor monomers, the Gal4 binding sites require dimerization, which occurs through the Gal4 DNA-binding domain. Assuming an inhibitory protein, which also contacts sequences of the B domain, it can be postulated for the Gal4-Sp2 (aa207-471) fragment that the inhibitor binds to the N-terminal region of the B domains of each fusion fragment. This could either impede the dimerization process due to steric reasons or could mask the dimerized Gal4 DNA-binding domains and prevent them from interacting with the DNA. As the unfused fragment Sp2 (aa207-606) appears not to interact with the inhibitory protein, it could be assumed that a Gal4-Sp2 (aa207-606) monomer would also be unable for this interaction. Only when being dimerized this interaction should be possible. Thus, the inhibitor should not prevent from dimerization but impede Gal4-DNA interactions by masking the Gal4 DNA-binding domain.

Regarding the situation for those Gal4-Sp2 fusions including the first 179 N-terminal amino acids (e.g. Gal4-Sp2 (aa6-471)): Assuming an inhibitor, which binds within the first 179 N-terminal amino acids as well as the N-terminal part of the B domain, such an effect as described for Gal4-Sp2 (aa207-471) should not be expected. Here, the Gal4 DNA-binding domain should remain unmasked because both transactivation domains of Sp2 (A and B) are accessible for the inhibitor. In addition, the ability of these Sp2 fragments to bind to DNA although interacting with the inhibitor could also explain the reduced mobility observed in gelelectrophoresis (see Fig. 3.1.5.7).

In addition to the Sp2-DNA interaction studies with unfused or Gal4-fused truncated Sp2 protein fragments, transactivation assays were performed using those fragments being able to bind to the GC boxes. However, although binding to DNA, these fragments were unable to activate reporter gene expression on different GC-box-containing promoters.

This could be explained by the lack of one transactivation domain in the analyzed protein fragments Sp2 (aa180-606), Sp2 (aa207-606), Gal4-Sp2 (aa6-215), Gal4-Sp2 (aa207-606) or the occupancy of the transactivation domains in the Gal4-fused Sp2 fragments including the first 179 N-terminal amino acids (Gal4-Sp2 (aa6-471), Gal4-Sp2 (aa6-357)). However, it is reported for Sp1 that one activation domain is sufficient for transactivation (Courey and Tjian, 1988). This would favour a repressing function of transcription factor Sp2 in the regulation of gene expression.

During the period of this thesis work, Phan et al. (2004) postulated a repressive function of transcription factor Sp2 on the expression of the carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) in prostate cancer cells. A repressive function of Sp2 would support the results presented here in this thesis. Their statement was based on the detection of an increased Sp2 expression level along with a decreased CEACAM1 expression in prostate cancer cells. Sp2 expression was detected by using the Sp2-specific Santa Cruz antibody no. sc-643 (Santa Cruz Biotechnology; USA: Santa Cruz). When trying to detect recombinant Sp2 protein by the related Sp2-specific Santa Cruz antibody no. sc-17814 during this thesis work, we observed that the antibody does not detect the recombinant protein (data not shown). In addition, when analyzing Sp2 protein expression in mouse tissues, the antibody detected a protein, which is smaller in size than expected for Sp2. This could also be observed for the expression data presented by Phan et al. (2004). From this, it can be concluded that the Santa Cruz antibody no. sc-643 might also be unable to detect the Sp2 protein. Therefore, the results of Phan et al. have to be interpreted cautiously.

In summary, our data suggest a regulation of the Sp2-DNA interaction capacity and by that its function as regulator of gene expression. Although exhibiting structural similarities like the highly conserved DNA-binding domain, the individual glutamine-rich Sp family members display a high functional specificity as demonstrated by the different knockout phenotypes. The functional specificity is achieved e.g. by differences in expression pattern or posttranslational modifications (Bouwman and Philipsen, 2002). The regulation of Sp2-DNA interaction capacity probably represents a further mechanism leading to this functional specificity.

Further steps in analyzing the mechanisms of Sp-DNA binding and transactivation regulation could be the attempt to narrow down the putative inhibitor binding sites by generating further deletion mutants. In addition, a yeast-two-hybrid screen for proteins interacting with the putative binding regions could be a valuable approach. On the other hand, the investigation of Sp2 posttranslational modifications should be performed. Moreover, a focus should also be directed on the analysis of Sp2-DNA interactions *in vivo* by CHIP assays.

4.3 Sp2 is essential for normal mouse development

To unravel the biological function of Sp2, Sp2-targeted mice were generated. These mice are characterized by the insertion of a LacZ-Neo cassette downstream of exon 4, which encodes the two glutamine-rich domains A and B. The insertion results in a fusion transcript consisting of the Sp2 exons 1-4 and the LacZ-Neo mRNA. RT-PCR experiments demonstrated that the exons 5-8 are absent in the Sp2-LacZ-Neo (Sp2-lzn) fusion transcript. It is expected that translation of the Sp2-lzn transcript leads to a non-functional Sp2 protein fragment. However, Western Blot analyses of Sp2-targeted and wildtype mouse tissue extracts (day E18.5) did not allow to distinguish between the targeted and the wildtype Sp2 protein (data not shown). From this, it has to be concluded that the translated Sp2-lzn transcript includes IRES and LacZ sequences at the C terminus, thus being similar in size as wildtype Sp2 protein. To confirm that the LacZ-Neo insertion reflects Sp2 deficiency and to determine putative differences between the Sp2 lzn/lzn and a Sp2 *-/-* phenotype, the generation of Sp2 knockout mice based on exon 4 deletion by crossing Sp2 lzn/lzn with Cre recombinase-expressing mice is in progress.

Regarding the Sp2 lzn/lzn phenotype: Whereas heterozygous Sp2-targeted mice (Sp2 lzn/wt) do not differ from wildtype animals, homozygous targeted mice (Sp2 lzn/lzn) display a clear phenotype: only one out of 73 born mice was lzn/lzn. This animal died within the first weeks after birth. To find out whether Sp2 lzn/lzn mice already die during embryonic development, mouse embryos of day E18.5, E12.5 and E9.5 were investigated. Whereas embryos appear normal until day E12.5, Sp2 lzn/lzn embryos at day E18.5 display a significant reduction in body size and weight, however with a strong variation. This favours the conclusion that the mutation of the

Sp2 gene effects normal mouse development at least at later embryonic stages. As Sp2-targeted mice were only available at the end of this thesis work, a detailed analysis of the mechanisms leading to the abnormal development could not be performed. For this, further studies are required.

As discussed above, Sp2 is closely related to the glutamine-rich Sp family members Sp1, Sp3 and Sp4, all of them recognizing the same DNA elements. However, the essential biological functions appear to be different. Whereas Sp1 mutant embryos are already severely retarded in the early development and die around day E10.5 of gestation (Marin et al., 1997), Sp3-deficient mice develop normal until birth but die a few minutes post-natum due to respiratory failure (Bouwman et al, 2000). In addition, they are characterized by abnormalities in tooth and bone development (Bouwman et al., 2000; Göllner et al, 2001b) and an impaired hematopoiesis (Van Loo et al., 2003). Disruption of the Sp4 gene revealed that Sp4 also is important for early post-natal survival since approximately two-thirds of the Sp4-deficient mice die within the first weeks after birth for unknown reasons (Göllner et al., 2001a). Surviving mice are growth retarded and either exhibit a pronounced delay in sexual maturation (females) or an abnormal reproduction behaviour (males).

Although displaying different phenotypes, overlapping functions of the individual Sp family members might exist *in vivo*. This is supported e.g. by the observation that Sp3 mRNA is up-regulated in Sp4-deficient mice suggesting a functional compensation of Sp4 deficiency by increased Sp3 levels (Göllner et al., 2001a). The existence of a certain functional redundancy within the glutamine-rich Sp family members could explain why loss of Sp2 does not visibly affect the early mouse development (at least until day E12.5).

Another possible explanation for this is based on the assumption that the interaction between transcription factor Sp2 and its corresponding DNA binding site in a promoter is a regulated process (see 4.2). Although widely expressed in mammals, Sp2 thus might control the expression of only a well defined set of genes at certain developmental time points or under distinct cellular or environmental conditions depending on the nature of the regulating mechanism. This functional specificity could be a reason why the effect of Sp2-deficiency is not visible at the early stages of development and not in all of the day E18.5 embryos.

To elucidate the reasons for the abnormal development of the targeted animals, a detailed histological analysis of the E18.5 Sp2 *lzn/lzn* embryos is required. As well, day E16.5 and E14.5 embryos should be examined to determine whether an influence of Sp2 deficiency is also observed at this stage of development. Additionally, as mentioned before, the generation of Sp2 knockout mice based on the complete deletion of exon 4 and the comparison with the *lzn/lzn* phenotype should be performed. Moreover, as the deletion of transcription factor Sp2 leads to lethality, the generation of tissue- or time-point-specific conditional knockout mice is necessary to unravel precisely its role in adult mice. Thus far, our results demonstrate that the loss of transcription factor Sp2 has an tremendous effect on the normal development and viability of the mouse.

In addition to the experiments proposed above, two other approaches should be performed in future: On the one hand, the occurrence or influence of redundancy in the Sp2 mouse phenotype due to the further individual Sp family members should be investigated by the generation of double or compound Sp knockouts. On the other hand, a second aim should be the identification of Sp2 target genes by the microarray technique.

5. References

- Anderson, K.P., Kern, C.B., Crable, S.C. and J.B. Lingrel.** 1995. Isolation of a gene encoding a functional zinc finger protein homologous to erythroid Krüppel-like factor: identification of a new multigene family. *Mol. Cell. Biol.* **15**:5957-5965.
- Armstrong, S.A., Barry, D.A., Leggett, R.W. and C.R. Mueller.** 1997. Casein kinase II-mediated phosphorylation of the C terminus of Sp1 decreases its DNA binding activity. *J. Biol. Chem.* **272**:13489-13495.
- Athanikar, J.N., Sanchez, H.B. and T.F. Osborne.** 1997. Promoter selective transcriptional synergy mediated by sterol regulatory element binding protein and Sp1: a critical role for the Btd domain of Sp1. *Mol. Cell. Biol.* **17**:5193-5200.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidma, J.G., Smith, J.A. and K. Stuhl.** 1987. *Current protocols in molecular biology*, Greene Publishing Associates and Wiley-Intersciences, New York.
- Bakovic, M., Waite, K.A. and D.E. Vance.** 2000. Functional significance of Sp1, Sp2, and Sp3 transcription factors in regulation of the murine CTP:phosphocholine cytidyltransferase alpha promoter. *J. Lipid Res.* **41**:583-594.
- Bieker, J. and C.M. Southwood.** 1995. The erythroid Krüppel-like factor transactivation domain is a critical component for cell-specific inducibility of a β -globin promoter. *Mol. Cell. Biol.* **15**:852-860.
- Birnbaum, M.J., van Wijnen, A.J., Odgren, P., Last, T.J., Suske, G., Stein, G.S. and J.L. Stein.** 1995. Sp1 trans-activation of cell cycle regulated promoters is selectively repressed by Sp3. *Biochemistry* **34**:16503-16508.
- Bishop, C.E., Boursot, P., Baron, B., Bonhomme, F. and D. Hatat.** 1985. Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. *Nature* **315**:70-72.
- Bond, B.J. and N. Davidson.** 1986. The *Drosophila melanogaster* actin 5C gene uses two transcription initiation sites and three polyadenylation sites to express multiple mRNA species. *Mol. Cell. Biol.* **6**:2080-2088.
- Borellini, F., Aquino, A., Josephs, S.F. and R.I. Glazer.** 1990. Increased expression and DNA-binding activity of transcription factor Sp1 in doxorubicin-resistant HL-60 leukemia cells. *Mol. Cell. Biol.* **10**:5541-5547.
- Bouwman, P., Göllner, H., Elsasser, H.P., Eckhoff, G., Karis, A., Grosveld, F., Philipsen, S. and G. Suske.** 2000. Transcription factor Sp3 is essential for post-natal survival and late tooth development. *Embo J.* **19**:655-661.
- Bouwman, P. and S. Philipsen.** 2002. Regulation of the activity of Sp1-related transcription factors. *Mol. Cell. Endocrinol.* **195**:27-38.

- Bradford, M.M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **72**:248-254.
- Bradley, A., Evans, M., Kaufman, M.H. and E. Robertson.** 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**:255-256.
- Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R. and K. Nasmyth.** 1985. Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* **41**:41-48.
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. and H. Cedar.** 1994. Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**:435-438.
- Breathnach, R. and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**:349-383.
- Bucher, P.** 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**:563-578.
- Buchholz, F., Angrand, P.-O. and A.F. Stewart.** 1996. A simple assay to determine the functionality of Cre or FLP recombination targets in genomic manipulation constructs. *Nucleic Acids Res.* **24**:3118-3119.
- Cook, T., Gebelein, B., Mesa, K. Mladek, A. and R. Urrutia.** 1998. Molecular cloning and characterization of TIEG2 reveals a new subfamily of transforming growth factor-beta-inducible Sp1-like zinc finger-encoding genes involved in the regulation of cell growth. *J. Biol. Chem.* **273**:25929-25936.
- Courey, A.J., Holtzman, D.A., Jackson, S.P. and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59**:827-836.
- Courey, A.J. and R. Tjian.** 1988. Analysis of Sp1 *in vivo* reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**:887-898.
- Dennig, J., Beato, M. and G. Suske.** 1996. An inhibitor domain in Sp3 regulates its glutamine-rich activation domains. *EMBO J.* **15**:5659-5667.
- Dennig, J., Hagen, G., Beato, M. and G. Suske.** 1995. Members of the transcription factor family control transcription from the uteroglobin promoter. *J. Biol. Chem.* **270**:12737-12744.
- Doetschman, T., Gregg, R.R., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and O. Smithies.** 1987. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**:576-578.

- Du, X.L., Edelstein, D., Rossetti, L., Fantus, I.G., Goldberg, H., Ziyadeh, F., Wu, J. and M. Brownlee.** 2000. Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing Sp1 glycosylation. *Proc. Natl. Acad. Sci. USA* **97**:12222-12226.
- Dynan, W.S.** 1989. Modularity in promoters and enhancers. *Cell* **58**:1-4.
- Dynan, W.S. and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**:79-87.
- Di Nocera, P.P. and I.B. Dawid.** 1983. Transient expression of genes introduced into cultured cells of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **80**:7095-7098.
- Emili, A., Greenblatt, J. and C.J. Ingles.** 1994. Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. *Mol. Cell. Biol.* **14**:1582-1593.
- Evans, M.J. and M.H. Kaufman.** 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**:154-156.
- Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T. and D. Rhodes.** 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**:483-487.
- Faisst, S. and S. Meyer.** 1992. Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.* **20**:3-26.
- Fried, M. and D.M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**:6505-6525.
- Garner, M.M. and A. Revzin.** 1986. The use of gel electrophoresis to detect and study nucleic acid-protein interactions. *TIBS* **11**:395-396.
- Gidoni, D., Dynan, W.S. and R. Tjian.** 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature* **312**:409-413.
- Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P. and R. Tjian.** 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. *Science* **230**:511-517.
- Gigliani, B., Comi, P., Ronchi, A., Mantovani, R. and S. Ottolenghi.** 1989. The same nuclear proteins bind the proximal CACCC box of the human beta-globin promoter and a similar sequence in the enhancer. *Biochem. Biophys. Res. Commun.* **164**:149-155.
- Gill, G.** 2003. Post-translational modification by the small ubiquitin-related modifier SUMO has big effects on transcription factor activity. *Curr. Opin. Genet. Dev.* **13**:108-113.
- Gill, G., Pascal, E., Tseng, Z.H. and R. Tjian.** 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the *Drosophila*

TFIID complex and mediates transcriptional activation. Proc. Natl. Acad. Sci. USA **91**:192-196.

Gillemans, N., Tewari, R., Lindeboom, F., Rottier, R., de Wit, T., Wijgerde, M., Grosveld, F. and S. Philippsen. 1998. Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the beta-globin locus control region in vivo. Genes Dev **12**:2863-2873.

Göllner, H., Dani, C., Phillips, B., Philippsen, S. and G. Suske. 2001. Impaired ossification in mice lacking the transcription factor Sp3. Mech. Dev. **106**:77-83.

Göllner, H., Bouwman, P., Mangold, M., Karis, A., Braun, H., Rohner, I., Del Rey, A., Besedovsky, H.O., Meinhardt, A., van den Broek, M., Cutforth, T., Grosveld, F., Philippsen, S. and G. Suske. 2001. Complex phenotype of mice homozygous for a null mutation in the Sp4 transcription factor gene. Genes Cells **6**:689-697.

Goodrich, J.A. and R. Tjian. 1994. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell **77**:145-156.

Hagen, G., Dennig, J., Preiß, A., Beato, M. and G. Suske. 1995. Functional analyses of the transcription factor Sp4 reveal properties distinct from Sp1 and Sp3. J. Biol. Chem. **270**:24989-24994.

Hagen, G., Müller, S., Beato, M. and G. Suske. 1992. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. Nucleic Acids Res. **20**:5519-5525.

Hagen, G., Müller, S., Beato, M. and G. Suske. 1994. Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J. **13**:3843-3851.

Hall, C.V., Jacob, P.E., Ringold, G.M. and F. Lee. 1983. Expression and regulation of *Escherichia coli* lacZ gene fusions in mammalian cells. J. Mol. Appl. Genet. **2**:101-109.

Han, I. and J.E. Kudlow. 1997. Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility. Mol. Cell. Biol. **17**:2550-2558.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557-580.

Hansen, T.O., Bundgaard, J.R., Nielsen, F.C. and J.F. Rehfeld. 1999. Composite action of three GC/GT boxes in the proximal promoter region is important for gastrin gene transcription. Mol. Cell. Endocrinol. **155**:1-8.

Harrison, S.M., Houzelstein, D., Dunwoodie, S.L. and R.S. Beddington. 2000. Sp5, a new member of the Sp1 family, is dynamically expressed during development and genetically interacts with Brachyury. Dev Biol **227**:358-372.

Hoess, R.H., Ziese, M. and N. Sternberg. 1982. P1 site-specific recombination: nucleotide sequence of the recombining sites. Proc. Natl. Acad. Sci. USA **79**:3398-3402.

- Hoey, T., Weinzierl, R.O.J., Gill, G., Chen, J.-L., Dynlacht, B.D. and R. Tjian.** 1993. Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell* **72**:247-260.
- Ihn, H. and M. Trojanowska.** 1997. Sp3 is a transcriptional activator of the human alpha2(I) collagen gene. *Nucleic Acids Res.* **25**:3712-3717.
- Imataka, H., Sogawa, K., Yasumoto, K.-I., Kikuchi, Y., Sasano, K., Kobayashi, A., Hayami, M. and Y. Fujii-Kuriyama.** 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, S.P., MacDonald, J.J., Lees-Miller, S. and R. Tjian.** 1990. GC box binding induces phosphorylation of SP1 by a DNA-dependent protein kinase. *Cell* **63**:155-165.
- Jackson, S.P. and R. Tjian.** 1988. O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55**:125-133.
- Jamieson, A.C., Miller, J.C. and C.O. Pabo.** 2003. Drug discovery with engineered zinc-finger proteins. *Nature Reviews* **2**:361-368.
- Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and S.T. Smale.** 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* **14**:116-127.
- Kadonaga, J.T.** 2004. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* **116**:247-257.
- Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and R. Tjian.** 1987. Isolation of cDNA encoding transcription factor SP1 and functional analysis of the DNA binding domain. *Cell* **51**:1079-1090.
- Kadonaga, J.T., Courey, A.J., Ladika, J. and R. Tjian.** 1988. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* **242**:1566-1570.
- Kalff-Suske, M., Kunz, J., Grzeschik, K.-H. and G. Suske.** 1996. Human Sp3 Transcriptional Regulator Gene (SP3) maps to chromosome 2q31. *Genomics* **37**:410-412.
- Kalff-Suske, M., Kunz, J., Grzeschik, K.-H. and G. Suske.** 1995. Human Sp4 transcription factor gene (SP4) maps to chromosome 7p15. *Genomics* **26**:631-633.
- Karlseeder, J., Rotheneder, H. and E. Wintersberger.** 1996. Interaction of Sp1 with the Groth- and cell cycle-regulated transcription factor E2F. *Mol. Cell Biol.* **16**:1659-1667.
- Kennett, S.B., Udvardia, A.J. and J.M. Horowitz.** 1997. Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. *Nucleic Acids Res.* **25**:3110-3117.

- Kingsley, C. and A. Winoto.** 1992. Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. *Mol Cell Biol* **12**:4251-4261.
- Kolell, K.J. and D.L. Crawford.** 2002. Evolution of Sp transcription factors. *Mol. Biol. Evol.* **19**:261-222.
- Kornberg, R.D.** 1996. RNA polymerase II transcription control. *Trends Biochem Sci* **21**:325-326.
- Krajewska, W.M.** 1992. Regulation of transcription in eukaryotes by DNA-binding proteins. *Int. J. Biochem.* **24**:1885-1898.
- Kriwacki, R.W., Schultz, S.C., Steitz, T.A. and J.P. Caradonna.** 1992. Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1. *Proc. Natl. Acad. Sci. USA* **89**:9759-9763.
- Kutach, A.K. and J.T. Kadonaga.** 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol. Cell. Biol.* **20**:4754-4764.
- Kuwahara, J., Yonezawa, A., Futamura, M. and Y. Sugiura.** 1993. Binding of transcription factor Sp1 to GC box DNA revealed by footprinting analysis: different contact of three zinc fingers and sequence recognition mode. *Biochemistry* **32**:5994-6001.
- Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685.
- Landschulz, W.H., Johnson, P.F. and S.L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**:1759-1764.
- Lee, J. S., K. M. Galvin and Y. Shi.** 1993. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci. USA* **90**:6145-6149.
- Leggett, R.W., Armstrong, S.A., Barry, D. and C.R. Mueller.** 1995. Sp1 is phosphorylated and its DNA binding activity down-regulated upon terminal differentiation of the liver. *J. Biol. Chem.* **270**:25879-25884.
- Lemon, B. and R. Tjian.** 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* **14**:2551-2569.
- Letovsky, J. and W.S. Dynan.** 1989. Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. *Nucleic Acid Res.* **17**:2639-2653.
- Levine, M. and J.L. Manley.** 1989. Transcriptional repression of eukaryotic promoters. *Cell* **59**:405-408.

- Lewin, B.** 1990. Commitment and activation at Pol II promoters: A tail of protein-protein interactions . *Cell* **61**:1161-1164.
- Liang, Y., Robinson, D.F., Dennig, J., Suske, G. and W. E. Fahl.** 1996. Transcriptional regulation of the SIS/PDGF-B gene in human osteosarcoma cells by the Sp family of transcription factors. *J. Biol. Chem.* **271**:11792-11797.
- Liu, J.P., Baker, J., Perkins, A.S., Robertson, E.J. and A. Efstratiadis.** 1993. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell* **75**:59-72.
- Lowry, O.H., Rosebrough, N.J., Farr, A.L. and R.J. Randall.** 1951. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**: 265-275.
- Ma, J. and M. Ptashne.** 1987. Deletion analysis of Gal4 defines two transcriptional activating segments. *Cell* **48**:847-853.
- Macleod, D., Charlton, J., Mullins, J. and A.P. Bird.** 1994. Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev* **8**:2282-2292.
- Majello, B., de Luca, P., Hagen, G., Suske, G. and L. Lania.** 1994. Different members of the Sp1 multigene family exert opposite transcriptional regulation of the long terminal repeat of HIV-1. *Nucleic Acids Res.* **22**:4914-4921.
- Mansour, S.L., Thomas, K.R. and M.R. Capecchi.** 1988. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**:348-352.
- Marin, M., Karis, A., Visser, P., Grosveld, F. and S. Philipsen.** 1997. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* **89**:619-628.
- Martin, R.G.** 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA* **78**:7634-7638.
- Matera, A.G. and D.C. Ward.** 1993. Localization of the human Sp1 transcription factor gene to 12q13 by fluorescence in situ hybridization. *Genomics* **17**:793-794.
- Merchant, J.L., Du, M. and A. Todisco.** 1999. Sp1 phosphorylation by Erk 2 stimulates DNA binding. *Biochem. Biophys. Res. Commun.* **254**:454-461.
- Mermod, N., O'Neill, E.A., Kelly, T.J. and R. Tjian.** 1989. The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58**:741-753.
- Miller, I.J. and J.J. Bieker.** 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol. Cell. Biol.* **13**:2776-2786.

- Mitchell, P.J. and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**:371-378.
- Moorefield, K.S., Fry, S.J. and J.M. Horowitz.** 2004. Sp2 DNA binding activity and trans-activation are negatively regulated in mammalian cells. *J. Biol. Chem.* **279**:13911-13924.
- Murata, Y., Kim, H.G., Rogers, K.T., Udvadia, A.J. and J.M. Horowitz.** 1994. Negative regulation of Sp1 trans-activation is correlated with the binding of cellular proteins to the amino terminus of the Sp1 trans-activation domain. *J. Biol. Chem.* **269**:20674-20681.
- Nagai, K., Perutz, M.F. and C. Poyart.** 1985. Oxygen binding properties of human mutant hemoglobins synthesized in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **82**:7252-7255.
- Narayan, V.A., Kriwacki, R.W. and J.P. Caradonna.** 1997. Structures of zinc finger domains from transcription factor Sp1. Insights into sequence-specific protein-DNA recognition. *J. Biol. Chem.* **272**:7801-7809.
- Nguyen-Tran, V.T., Kubalak, S.W., Minamisawa, S., Fiset, C., Wollert, K.C., Brown, A.B., Ruiz-Lozano, P., Barrere-Lemaire, S., Kondo, R., Norman, L.W., Gourdie, R.G., Rahme, M.M., Feld, G.K., Clark, R.B., Giles, W.R. and K.R. Chien.** 2000. A novel genetic pathway for sudden cardiac death via defects in the transition between ventricular and conduction system cell lineages. *Cell* **102**:671-682.
- Nishida, M., Kasahara, K., Kaneko, M., Iwasaki, H. and K. Hayashi.** 1985. [Establishment of a new human endometrial adenocarcinoma cell line, Ishikawa cells, containing estrogen and progesterone receptors]. *Nippon Sanka Fujinka Gakkai Zasshi* **37**:1103-1111.
- Orphanides, G., Lagrange, T. and D. Reinberg.** 1996. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**:2657-2683.
- Pabo, C.O. and R.T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**:1053-1095.
- Pascal, E. and R.Tjian.** 1991. Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes Dev.* **5**:1646-1656.
- Pavletich, N.P., and C.O. Pabo.** 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**:809-817.
- Phan, D., Cheng, C.-J., Galfione, M., Vakar-Lopez, F., Tunstead, J., Thompson, N.E., Burgess, R.R., Najjar, S.M., Yu-Lee, L.-Y. and S.-H. Lin.** 2004. Identification of Sp2 as a transcriptional repressor of carcinoembryonic antigen-related cell adhesion molecule 1 in tumorigenesis. *Cancer Res.* **64**:3072-3078.
- Philipsen, S. and G. Suske.** 1999. A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res* **27**:2991-3000.

- Ptashne, M.** 1988. How eukaryotic transcriptional activators work. *Nature* **335**:683-689.
- Robertson, E., Bradley, A., Kuehn, M. and M. Evans.** 1986. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323**:445-448.
- Roeder, R.G.** 1991. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem Sci* **16**:402-408.
- Roeder, R.G.** 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**:327-335.
- Rohlf, C., Ahmad, S., Borellini, F., Lei, J. and R.I. Glazer.** 1997. Modulation of transcription factor Sp1 by cAMP-dependent protein kinase. *J. Biol. Chem.* **272**:21137-21141.
- Ryu, S., Zhou, S., Ladurner, A.G. and R. Tjian.** 1999. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* **397**:446-450.
- Roeder, R.G. and W.J. Rutter.** 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**:234-237.
- Roos, M.D., Su, K., Baker, J.R. and J.E. Kudlow.** 1997. O-glycosylation of an Sp1-derived peptide blocks known Sp1 protein interactions. *Mol. Cell. Biol.* **17**:6472-6480.
- Ryu, S., Zhou, S., Ladurner, A.G. and R. Tjian.** 1999. The transcriptional cofactor complex CRSP is required for activity of the enhancerbinding protein Sp1. *Nature* **397**:446-450.
- Sambrook, J., Fritsch, E.F. and Z. Maniatis.** 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, New York.
- Sapetschnig, A., Koch, F., Rischitor, G., Mennenga, T., and G. Suske.** 2004. Complexity of translationally controlled transcription factor Sp3 isoform expression. *J. Biol. Chem.* **279**:42095-42105.
- Sapetschnig, A., Rischitor, G., Braun, H., Doll, A., Schergaut, M., Melchior, F. and G. Suske.** 2002. Transcription factor Sp3 is silenced through SUMO modification by PIAS1. *Embo J.* **21**:5206-5215.
- Schneider, I.** 1972. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* **27**:353-365.
- Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schroder, C., Kemler, R., et al.,** 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Krüppel, a *Drosophila* segmentation gene. *Cell* **47**:1025-1032.
- Schoy, S., van Vooren, P., Szpirer, C. and J. Szpirer.** 1998. Assignment of Sp genes to rat chromosome bands 7q36 (Sp1), 10q31→q32.1 (Sp2), 3q24→q31 (Sp3)

and 6q33 (Sp4) and of the SP2 gene to human chromosome bands 17q21.3→q22 by in situ hybridization. *Cytogenet. Cell. Genet.* **81**:273-274.

Seeler, J.S. and A. Dejean. 2003. Nuclear and unclear functions of SUMO. *Nat. Rev. Mol. Cell. Biol.* **4**:690-699.

Sjottem, E., Anderssen, S. and T. Johansen. 1996. The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3' to the TATA box. *J. Virol.* **70**:188-198.

Smale, S. T. and D. Baltimore. 1989. The "initiator" as a transcription control element. *Cell* **57**:103-113.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A. and R.S. Kucherlapati. 1985. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**:230-234.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.

Southern, P.J. and P. Berg. 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* **1**:327-341.

Stark, W.M., Boocock, M.R. and D.J. Sherratt. 1992. Catalysis by site-specific recombinases. *Trends Genet.* **8**:432-439.

Struhl, K. 1988. The Jun oncoprotein, a vertebrate transcription factor, activates transcription in yeast. *Nature* **332**:649-650.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and J.W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**:60-89.

Su, K., Roos, M.D., Yang, X., Han, I., Paterson, A.J. and J.E. Kudlow. 1999. An N-terminal region of Sp1 targets its proteasomedependent degradation in vitro. *J. Biol. Chem.* **274**:15194-15202.

Subramaniam, M., Harris, S.A., Oursler, M.J., Rasmussen, K., Riggs, B.L. and T.C. Spelsberg. 1995. Identification of a novel TGF-beta-regulated gene encoding a putative zinc finger protein in human osteoblasts. *Nucleic Acids Res.* **23**:4907-4912.

Supp, D.M., Witte, D.P., Branford, W.W., Smith, E.P. and S.S. Potter. 1996. Sp4, a member of the Sp1-family of zinc finger transcription factors. is required for normal murine growth, viability, and male fertility. *Dev. Biol.* **176**:284-299.

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

Suske, G. 2000. Transient transfection of Schneider cells in the study of transcription factors. *Methods Mol. Biol.* **130**:175-187.

- Suske, G., Bruford, E. and S. Philipson.** 2005. Mammalian SP/KLF transcription factors: Bring in the family. *Genomics* **85**:551-556.
- Tanese, N., Saluja, D., Vassallo, M.F., Chen, J.L. and A. Admon.** 1996. Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100. *Proc. Natl. Acad. Sci. USA* **93**:13611-13616.
- Tybulewicz, V.L., Crawford, C.E., Jackson, P.K., Bronson, R.T. and R.C. Mulligan.** 1991. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell* **65**:1153-1163.
- Udvardia, A.J., Templeton, D.J. and J.M. Horowitz.** 1995. Functional interactions between the retinoblastoma (Rb) protein and Sp-family members: Superactivation by Rb requires amino acids necessary for growth suppression. *Proc. Natl. Acad. Sci. USA* **92**:3953-3957.
- Vallian, S., Chin, K.V. and K.S. Chang.** 1998. The promyelocytic leukemia protein interacts with Sp1 and inhibits its transactivation of the epidermal growth factor receptor promoter. *Mol. Cell. Biol.* **18**:7147-7156.
- Van Loo, P.F., Bouwman, P., Ling, K.W., Middendorp, S., Suske, G., Grosveld, F., Dzierzak, E., Philipson, S. and R.W. Hendriks.** 2003. Impaired hematopoiesis in mice lacking the transcription factor Sp3. *Blood* **102**:858- 866.
- Voss, D. and O. Pongs.** 1986. Upstream sequences modulate in vitro transcription from *Drosophila* yolk protein genes I and II. *Eur. J. Biochem.* **158**:25-32.
- Weinzierl, R.O.J., Dynlacht, B.D. and R. Tjian.** 1993. Largest subunit of *Drosophila* transcription factor-IID directs assembly of a complex containing TBP and a coactivator. *Nature* **362**:511-517.
- Wimmer, E.A., Frommer, G., Purnell, B.A. and H. Jackle.** 1996. buttonhead and D-Sp1: a novel *Drosophila* gene pair. *Mech. Dev.* **59**:53-62.
- Wimmer, E.A., Jäckle, H., Pfeifle, C. and S.M. Cohen.** 1993. A *Drosophila* homologue of human Sp1 is a head-specific segmentation gene. *Nature* **366**:690-694.
- Wingender, E.** 1988. Compilation of transcription regulating proteins. *Nucleic Acids Res.* **16**:1879-1902.
- Yang, X., Su, K., Roos, M.D., Chang, Q., Paterson, A.J. and J.E. Kudlow.** 2001. O-linkage of N-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability. *Proc. Natl. Acad. Sci. USA* **98**:6611-6616.
- Zhao, L. and L.S. Chang.** 1997. The human POLD1 gene. Identification of an upstream activator sequence, activation by Sp1 and Sp3, and cell cycle regulation. *J. Biol. Chem.* **272**:4869-4882.

Diploma and doctoral theses:

Braun, Harald: Regulation der lungenspezifischen Expression des Uteroglobin/CC10 Gens durch die Transkriptionsfaktoren HNF3, TTF-1, Sp1 und Sp3. 2001. Universität Gesamthochschule Kassel.

Doll, Andreas: Klonierung von Interaktionspartnern des Transkriptionsfaktors Sp3 durch ein genetisches Interaktions-Screening in *Saccharomyces cerevisiae*. 1998. Philipps-Universität Marburg.

Göllner, Heike: Identifizierung potentieller Zielgene der Transkriptionsfaktoren Sp1 und Sp3 in embryonalen Stammzelllinien und Etablierung Sp3-defizienter Fibroblastenzelllinien. 1998. Philipps-Universität Marburg.

Göllner, Heike: Die biologische Bedeutung der Transkriptionsfaktoren Sp3 und Sp4: Charakterisierung von Sp3- und Sp4-Knockout Mäusen. 2002. Philipps-Universität Marburg.

Krüger, Imme: Generierung von Knock-In-Konstrukten zur *in-vivo*-Mutagenese des IKEE-Motivs des Transkriptionsfaktors Sp3. 2002. Philipps-Universität Marburg.

Mangold, Monika: Charakterisierung des Maus Sp4-Gens und Konstruktion eines Vektors zur Herstellung Sp4 defizienter Mäuse. 1995. Philipps-Universität Marburg.

Rischitor, Grigore: Transcription factor Sp3 as target for SUMOylation *in vivo*. 2005. Philipps-Universität Marburg.

Stielow, Bastian: Analyse der E3-Ligase-Aktivität der PIAS-Proteine bei der SUMOylierung des Transkriptionsfaktors Sp3. 2003. Philipps-Universität Marburg.

6. Appendix

6.1 Abbreviation index

Additionally to the usual SI and IUPAC units, the following shortcuts were used:

A	Adenosin
APS	Ammonium peroxodisulfate
ATP	Adenosin triphosphate
bp	base pairs
BSA	Bovine serum albumine
C	Cytidin
°C	degree Celsius
CAT	Chloramphenicol acetyl transferase
cDNA	complementary DNA
cpm	counts per minute
dATP	desoxy Adenosin triphosphate
dCTP	desoxy Cytidin triphosphate
Del	deletion
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethylsulfoxide
DNA	Desoxy ribonucleic acid
dNTP	desoxy nucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediamine-tetraacetic acid
EGTA	Ethyleneglycol-tetraacetic acid
ES cells	embryonic stem cells
FBS	Fetal bovine serum
Fig.	Figure
fw	forward
G	Guanosin
G418	Neomycin analogon G418
HA	Haemaglutinine
HEPES	4-(2-Hydroxyethylene)-1-Piperazineethansulfonate

hsv-TK	herpes simplex virus Thymidine Kinase
HTLV	Human T-cell Lymphotropic Virus
IPTG	Isopropyl- β -D-galactoside
kb	kilo base pairs
kDa	kilo Dalton
Luc	Luciferase
lzn	LacZ-Neo (targeted Sp2 allele)
mRNA	messenger RNA
Neo	Neomycin resistance gene
nt	nucleotide
OD	optical density
PAA	Polyacrylamide
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pgk	phosphoglycerate kinase promoter
PMSF	Phenylmethylsulfonylfluoride
PVDF	Polyvinylidendifluoride
RNA	Ribonucleic acid
rev	reverse
RNase	ribonuclease
rpm	revolutions per minute
SDS	Sodium dodecyl sulfate
SSC	SDS sodium citrate buffer
SV40	Simian Virus 40
T	Thymidine
TBE	Tris borate EDTA buffer
TE	Tris EDTA buffer
TEMED	Tetramethylethylenediamin
Tris	Tris-hydroxy-methyl-amino-methane
U	Unit (enzyme activity)
V	Volt
v/v	volume per volume
wt	wildtype
w/v	weight per volume

6.2 Mouse Sp2 genomic DNA and cDNA sequences

Mouse Sp2 genomic DNA sequence; Celera access. no. mCG13240

Exon identification via alignment of the genomic Celera mouse Sp2 DNA sequence with two mouse (NCBI access. no. BC021759 and NM_030220) and two human (NCBI access. no. BC016680 and D28588) Sp2 cDNA sequences.

9001	ATGAAGACCT	GAGTTCATAT	CTCCAGTACC	CATGTCAAAA	ACAGTGTGCA	CATTCATAAC	
9061	CCAGGTTCCG	GGAAGGTGGA	GACAGAAGGG	TTTACTGGCC	CCAGTCTAGC	TGAATCTGCA	
9121	GGTTCTGGTT	CAATGAGAGA	CTGAGGACAA	CTATCGCCTT	CTTATGACCA	TATGCACACA	
9181	CATGTGCTTG	TACCAGAACT	CACANNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	
9241	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	
9301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	
9361	NNNNNNNNNN	NNNNNNNNNA	TGTCCTCGGC	GCTCAGCGCT	TACACTGTCC	CCTGGCTAAG	
9421	CCTTCCTTAA	CTAAGGCCCT	CTGTCTCAG	AACTGCAGGT	GGTTTGACC	AAGTCTGAG	
9481	TCTTCCCTGC	ATTGAGGTGA	TGGCTGCCAG	GCAGAGCGGG	ATGGGCAGTA	AGTATGTGGG	
9541	TCTGGGTACA	GAGGAAAGGA	CAGGAGTGCC	GTCCTACAAA	TTATGCTATC	ATCAATTAC	
9601	TGTGACCTCA	GCCAACTCCA	GCACCTTCTC	GAGCCTTGGT	TTCCCTTGGT	CATAAAATAT	
9661	GGGGTTAATT	GTCCCTCCTA	GGGAAGCTGG	GAGGGTTAAT	TGGGCAATGG	CTCTGAAGAG	
9721	AAGGAGATCT	ATAAATACAA	GGGATCATT	GAGGTATTT	CACAGCCGTT	TTGTGGCCCC	
9781	GATGGTATTA	TTAATATGAT	TATTGTTATT	GTTATCGCCA	CCATCACCCA	CTTCCCGAGT	
9841	CTCTGCCAGC	TGGGTCCTGC	TTCCAGGTGA	AGGAGCTCTG	CCAACAGCCT	CTGTGACCCT	
9901	CCACCCTCAG	GAGTGTCCGG	GACCCCTCAG	GAGGTATTT	CATTGTGGCA	TTACAGAGGG	
9961	GGTCAAGAGC	TTTAAGGGGA	GGGGGTGAGG	GGACGTGCCA	TGCTTGCTTA	GTGAGGGCTA	
10021	AAGGCTTTTA	GTCCAGGGAT	GACCCGTGTC	TGCTGTGGGA	AGCCTGGACT	CGAGAAGCCT	Exon
10081	GATGGCTCCA	TGGAAATGTG	CTCATG	GTGA	AGCAGGAGTT	TGAGGGGATG	(5' end
10141	GTGTTTTCTG	GGACGCTGAG	GGACGGGCAG	ATGGGGATGT	GCTCTCTTGA	GGAGGCAAAAT	unclear)
10201	GGCTCTTCTT	AAAGGAATCT	CTTTGTAGGG	GTCCTAGGAT	TCTTAAGGCC	TGGGGATGAA	
10261	GGACGGCCTA	GGCTGGCCAA	TAGTAGGCC	AGAATATCCA	CCAGGTTATC	TGTGTTTCTT	
10321	CCTTTGGGCA	TGTAAAACCA	CGGTCCAGGGG	CTTCTGAAAT	GCCTTGGGTG	CCCTCTGGCG	
10381	GCAACACGTG	AGTGCTACAG	AGGACCAACA	CCTGCCCAGA	GAGCACCTCC	TCCTGCCCTC	
10441	GGCAGAAAGG	AAGAACCCTAC	CAGAGGGATG	GCTGGAATGT	CCTTGCTGGG	AGATTCTGTG	
10501	GCAGTTGGCC	AGCAGCCATC	CACCTGGCCA	GGAAAGGCCA	GGTAGAGGG	AGCTGTGGGG	
10561	TCACAGCAAG	GGACGGAAGT	GGTATCCCTG	GGTTTGCAAC	ACGGGCAGTT	GGTGGCTGTA	
10621	TGTGTGCACC	AGGATGCACA	GTGTGGGGAG	CTAGAACATA	TGAGGGAATC	CACCTTTGGC	
10681	GTTCCCTGAT	GGTTCACCCA	AACCAAGGGT	TTTCCCATCA	AGCCACTGTC	GGGGTTTATG	
10741	CATCAGGGCA	TCAGGGATGT	ATGGGAGGGT	CTTCTCCATA	GGTGGAGAGA	GAATGAAGAA	
10801	AGAAAACCAG	AGGAGGAAAA	GGAAAGAGGCA	GAGGCCATGG	GGAGATCCTA	AGGCAAAAAGT	
10861	GAGCATCCTC	AGAAGGCTGA	GGGGGCTGTA	ATGGTTTGAA	GAATTAGCCT	GTCAAACCTG	
10921	GGCCATAGGA	GGGAGTGTGA	CCCCGTGTGG	AGGCCAAGGC	TGCCCTGCCCT	TAGACAAAAG	
10981	CAGTCAGGAC	ATGCTTCTAG	ACCAAGAGCA	ATGTTGAGCC	TTGTGGATCT	TTCTATATCT	
11041	TCCTGGGGAC	AGAGACCATG	ACTCCGTGTG	TCTCCTTCAG	TGTGCTCAGC	ACCCTTGGGG	
11101	ACTCTGGATA	GACAAGGGCA	GCCACATGCA	GAAGAGAACA	GAGAGGGCTG	GGAAAGCAGG	
11161	AGAGCCAGAG	AGGGCTTCAC	AGGAAGCTGA	AGCCAGGGAA	GCAATAGAGC	GGACATGGAG	
11221	CAAGCTGAGC	CCGAGTGAGA	AGTGCTCAAA	GGACCGCCAA	CAGCTGTGCG	CCTGCCAGCG	
11281	GCACGTGGCC	GGTAGTGGAG	GGCGCGCGGT	GGAGGTCAAG	GTAACATATG	ACCATACGTG	
11341	GTGGAGAACG	AGTAACCTTG	CACACTAGAA	ATGGGGACAG	AGTGGAAACG	TGGACTAGGA	
11401	GGAGGCTCAA	AGCAGCTGGA	GGCAGCACCG	ACAGAACTTG	CTTCCAGACC	CACCAGGGAC	Exon
11461	AAGATGGCGC	TGTGACGTGG	GTGATGTGGG	CAGACACAGC	TGGAACCGCC	TAGCCACCTT	
11521	TGCTCTTCAA	ACACCCTGTC	AGGTAGGGTG	GAACAAGGAG	CTCTTCAGGG	ATGGAGAGGC	
11581	AGGAGTGGGT	TGGCTCAGGG	CCTTTGGTGA	AGGAATGCTT	TGTTTACATC	TGCCCTTGCTT	
11641	ACACATGGAT	TACCTGAGGG	TCTTCTTCAT	ATAGGTGCTC	GGCTCCATCC	TAGGGACTGA	
11701	CATCCTGGGA	TGAGACAGGA	GCCCAACTTC	AGAACTGGTG	AAAGGCCGTC	TGCTGAGTCT	
11761	GACGTATAGT	CAGGACTCTG	ATGTGGTTAG	AGCACTTGCT	TTTTGTGCAC	GAGGCTCTGG	
11821	GTTCAATCCC	CAGTATAGGA	ATCAGGCATA	GTGGAACTCG	CTGCCTAGTC	TCTTGAAAT	
11881	GGAGACCTGT	GAGAAAAAAA	CAGTTGTGCG	GTGGTCTGAG	GGGGCTTATG	AATGAACTTG	
11941	AAAGACTCAT	CGGTAGGGTA	GGGAATGTGG	TTCAGTGGAC	CCTGGTTTCC	CAAGCATGAG	
12001	GCCCCAGATT	CAATGCCCCAG	TACAAGGGGC	AAACAAGCAA	ACACAAAAGG	TCCCTGTFTT	
12061	TGTCTGTGCC	TTTAATAAACC	TAAAAACAAA	ACAAACGTAC	ACCCCTTTAA	TCCCAGCACT	
12121	CAGAAGGCAG	AGGCAGGTGG	ATTTCTGTGA	CTATGAGACC	AGCTTGGTGT	ACAGAGTGAG	
12181	TCTACAGGGC	TACACAGAGA	AACCTGCCT	GGAGAAAACGA	AAAAAAAAC	CAGATAGCCA	
12241	AAACCCAAAA	ACAAACAAAC	AAACTGACAG	GCTTTAAAAG	GGCTGAACCA	TTCAATTTTAA	
12301	AACTATTGGT	ACCAGTTGGG	TCTGGTGGCC	CATGCCTGAG	AGCCTAGCAC	TTGGGAGGTA	
12361	GAGGCAGGAA	GATCATGAGT	TCTGGGCAAG	CCTGGGCTAT	TTGAGACCCT	GTCCATAAAA	
12421	CAACTAACCA	CCCAATCAGT	ACTACTGGTG	AATCCTAGGT	TTTCTCACTG	TCCCCCACCC	
12481	ACACTTTTCA	TGTGTCTTTC	ATGAGAGGGC	CTCTGGGAAG	GACAGAGTTA	GCTACCTCCA	
12541	GCTTTTCTCT	AGAAGAGGTT	TCCACCCAGC	AGGATAGTCA	GCCCCACCTC	CCCACCAAAA	
12601	CCCCTGCACC	CTAATGTACC	TGAGTCACTG	AATCCGGGGC	GAGCCCTTGG	CTACACAGCA	
12661	AAGCAAGCCA	AACTGAAAC	TGGCCCTGGG	AACAGAATGT	ATGGGCTAAA	CAAGTCTTAT	
12721	GCAGATGTAG	GAGTATGCAA	AGTCATTCTT	GAACCTCCGA	TCCATTTTCT	GATCATCTCG	
12781	GATAGAAGGC	CCAGTTCTTC	CAGTGTAGCT	CAGTCCCTTA	GGGCAGCCAT	GAGGGAGAAT	
12841	GGGGCTCAGT	GGTAGGACT	GCCTCCGATG	TGGGAGGTCC	TGCCTTGTAT	TCCCTAGCCC	
12901	TGAAACAAAC	AGAAGTGAAA	GAGTGGGAGA	AGGGAGTCTT	TCATGTAAGC	CAGTGACAGA	
12961	CCAGGGCAGC	CAGCAACAAG	AAGGCGAAGC	TTGACTGGGC	AGGAGTGAAG	AGAACAGATT	
13021	TTGTAGGACA	AAGGAAAGGC	TTACCTAAGA	AGAAGGGGAA	CTCCTATGTG	TTTAGGGACT	

13081	GTGACACCTT	TCCTTGAGAG	AAAAGACTAG	GGAAGCAAAG	AAAGGTGGGG	AAAGGCATGA
13141	AAGACAAGAG	CATGGAGGGT	ATGTGTGCGT	GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT
13201	GTGTGTGTAA	GGGTGAAGGA	CCCAGTTTAA	GCAGAAAGCT	TTGATTTTAT	AAAGGACATA
13261	GAGTCCCAAT	GAAGGAACGG	TGCGTGGGCA	CACATGTGCA	CACGGACGCC	AGAGCAGGAC
13321	ATACACTGGG	TCCTGCTCTG	TTGCTGTGTC	TCACAAAGCC	TGGAGCTAGG	TTGGCCACTG
13381	GCAGACTCCA	GCCATCCTCC	GGTCTCCGCC	CTGACCTCAC	ATGTGAGAGG	TGGGGTTACA
13441	GAGGCTGAAC	TCCAGCTTTT	CACACGGGCT	CTGAACTCAG	GCCTCTTGTT	TTGAACAGC
13501	AAGTGCATC	TTCCCAGCCC	CAAGTGAAGG	ATTCTAGCTC	CAGAACTGAT	AGCATCTGAT
13561	GTTTGGTGTT	TCAAAACCTA	ACCAGAGAGA	GGTAGACTGG	TAATTCCTAA	TTACTACACT
13621	CTAAGCATTT	GGCTCTTTAT	TGTTTATTTA	TTTATTTTIT	TGTTTGTGGT	TTGTGGTGTG
13681	TGTGTGTGTG	TGTGTGTGTG	TGTGGTTTTT	GTTTGTAGACA	GGGTTTCTCT	GTGTAGTCTT
13741	GGCTGTCCCTG	GAACTTTCAC	TGTAGACCAG	GCTGGACTCA	AAGTCAGAGA	TCTGCCAGCC
13801	TCTGCCTCCC	TGCTGCTGGG	ATTGAAGGCC	TGGGCCACTG	CCACCCAGCT	TGGCTCTATA
13861	TTTTGTATATC	ACTTTATCTC	AGCAACCTTT	TCTTCGAAGT	GACAGGAAGA	AGGAGGAGGA
13921	AGATAGAAGA	GGCGAAGAGA	AAATGTAATG	AGAGCTGGGG	AACAGCTCGA	CGATACAGTT
13981	CATGCTTAGT	GTTTATACAG	TTTCATGCTTA	GTGTGGGGGA	CTTGATCCTC	AGCAAAATAAG
14041	ACCCACAGCAG	GGATGTATGT	GTCCTCATCTA	CCTGCGATCA	ATGTCTCTCA	AATGTTGATT
14101	TGGGAAGCTT	TGTTACATGT	TCTTATAAGT	TAATTTACTG	GATGTAAAAC	ATTTTTTTCT
14161	CACTGTGGGT	TAAAACAACA	AAACA AAAACC	AACCAGGAAG	GGTGGTACTA	TGAGGTGGAT
14221	GTGTGAGTGC	TCTGGCTTCA	ACAAGAGTCT	AAGGGTCCAC	TGCAGATGGG	GTGTGGGGTT
14281	TCCTACCCAC	GCAGTCTCA	AGCCAGCGTG	AATTCCTCT	GAGTGTCTACG	TACAGCTTGC
14341	CTAAGTTAGT	CGTTCCGAGA	GGGCCTGGAA	GAGAAAGCTA	AACCTACCTG	TGGTGCCTGT
14401	CAAGCACGGC	CCCTCATACA	GGAGGAGATG	AAGCAGAGAG	CAAGCAGGGT	CGTAGGATG
14461	CCAGCGTGCA	ATTCTGCCCA	CCTGAGGCGC	AGGCTAAAAT	GCAGATTCTA	GGCACACACA
14521	CCCCGCCCC	CGGGAATAG	CAAAATGTAAC	TAGGGACGCA	CCTAGGGATC	TGCTTTTGAT
14581	AAAGGCTCAA	ACACGCTGT	CTGAAGCTGA	GAGGATCAGG	GAAGCTCTGC	GAACATTTGG
14641	TGAAGATTAG	CAGGGGCTTG	AGATGACATA	ATGACAATTC	CCTCTAGGAA	GTGCCAGGTA
14701	ACCCAACCTCC	TGCCAGGTCG	TCGACTCTCG	AGAGGCGATG	GATGGCGGGC	GAGGGCCTGA
14761	GGTGAAGGCG	ATCGGCTCAC	TGGGGGCCGG	GCACC CGCTA	TCCGACAGCT	ACACTTGAGA
14821	CTCACCATGC	CAGACTGTAT	TGAGAGATGT	ATAAGCTCTT	TACGGGGCAT	TTACACGTAG
14881	CACTTCTGAT	TGGCTTTAGG	GCAGATCTAT	CACCTATCTG	CTGGGTTAGA	GGTGGGTGGA
14941	AGGTCAATTA	GCACCAATAC	AAAAACACTT	CTCGGCTAAG	CCCTCCTCA	CCCAGCTTCC
15001	TGTTGTCTGT	CATCAACCTC	CTAGCAGGCG	GGACTTTGCC	TCCGCCCAGT	TGGCTAAAAG
15061	ATACGTCGTT	CATTCCCCCG	TTGATTTCCG	GTTGGCGACT	CCTACCGTCT	ATCAATGACA
15121	CTTCACAGTT	CTATTGGCCA	GACCAACGTC	ACTCTAGATG	CTTCCCAATG	ATTGGTTGAA
15181	AATGCTGTCT	CTCCCCAATG	TACCAAGCGT	TCATTGGTAA	AAGCTGCCGT	CGCTCGGGCG
15241	GTGGCCGGCT	CGCGGATTGG	CGGGTGCTAG	CGCGGCGGTG	TCAGGCTCTC	GGTGGCGGGC
15301	GAGCGCGCGG	AGGCCAGGGA	GGAAGATGTC	GTAATGAGCG	GTGGTCTCTG	CGCTCTGCCG
15361	CGCGCTCTTT	CCCGGCCCTC	AAGGGTCCGG	AAGACGGACA	GAGGCTCCGG	GGAGGGGAGG
15421	CCGGAACCGG	GGGTGCCCCC	CTGGGACTGG	GACTGGCGCC	GGGGCCCCGG	TACGGGAGGG
15481	GCGGGGAGG	AAGGATTTCC	GCCCCGAGGA	GGAGCGCGCC	TGGCCGGCCG	GGCGGAGGCA
15541	CCCGGAAGGG	GGCGTTAGAC	TCCCCGCCCC	TCCGAGCCCC	GGCGGGGAGG	GCGGCGCTTC
15601	CGGGGCTGGC	AGCCGGCGCG	GGCAGCAGGG	CGAGCCTGAT	AACCCCTTTC	GCCCCCTCAG
15661	CTTGGGTCCG	GCCGGGCCCTT	CAGCCTGCGC	CCGAGTTTGC	TCCACGCAGG	CGGCCGACGC
15721	CCCCTCGGGC	TCGTCTGTGT	GGCAGCCCAG	GCTTGGGACC	CAGGTGGTGC	CAGTCAGATG
15781	GGCCCTGTCT	GTCTGAACATG	CAGCCTGTCTG	CGCGGGCCAC	CGCAGTCCCA	GAGGGTGGCA
15841	GGATTGGAGG	GACAGCCCTA	GGCAGATGAC	TGCTGAACGA	AAAACCGACT	TTTAATGTTT
15901	ATACCTGGAT	CCTTGAGCCT	GCGGCTAGTA	TCTGCAAGCT	C TTGGAATTT	TAGAAGTTTG
15961	GAGAACTTGT	GGACAGAGCT	CCCAGGCTAT	GGCCCGATGC	CCCTTGCGGG	NNNNNNNNNN
16021	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16081	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16141	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16261	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16321	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	TGTGTAAGTG	CAGGAACAGA	CTACAGCTGT
16381	TCCATGCTGT	GCCCTAGAAG	GCCCTTGGCA	TTACAGAGGA	CAAACCCCTG	AGTTGAGACA
16441	CCCCCCCCAC	CCCCCGTGGT	GATGAAGGTT	GAAATCACAA	ATGTTGGACA	AGCACCTCTA
16501	CGGCTGAGCT	GTCTTGGAGG	GCTAGAAAAG	TTAGGGGTGAG	CCAGCTCACA	GTAAGCTGGC
16561	ACAGGGGATG	ACAAGAGCCC	AGACCACACA	GTAGCATTTT	CCATTATTGA	CACAAAGAAG
16621	TCCATTGTCT	GTAGCGATGA	GTATAACTGC	AGTGTCTTGT	TTTCTTTTAA	ATCTTAAAG
16681	ACAATGAACA	GCGCGGATCT	CAGAACCTCC	CAATAGGGCT	TATGTCAAGT	TTGCAGTGGT
16741	TTCTTATTTA	GGTGTATAGG	TGTTGGAGCC	TGTTCTTGCT	GTGAAGCCCT	GGCTGGCGTG
16801	TGTTTGTCTCT	GTGATCTGTG	TAGGCCTTGA	ACTTGAGATC	TCTATGCTCC	AGCCTCAGCC
16861	TCCCAAGTGC	CGAGGTATG	GGTGTGTACC	AATTTGCTTG	GTTTCTAGTG	ACCAGGCAGA
16921	CGTCAATCAA	TACCAAGTAC	CATGAGTTAT	CTTTGGCTGA	AGTGACGGAT	GAGCTGTCTA
16981	GTTCCAGTCC	TCCTACTTGC	CTCTTTGAGT	AAAATCAACA	GCCATTGGGA	AAAATAAGCT
17041	GCTCTGACAC	CATTGTGAGC	TCCGCAAGCT	AGGCTGCTAA	AGACACAGGC	TCTTTGCCAA
17101	AGAAAATAAC	CTAGGATTTG	TCCATAACCG	GATCTTCCCC	TGGGCACCCA	GCTGCTCTGA
17161	GGGGAAGAGT	TTGACATCCC	TGCTGATACT	ACATAGATGA	CCTCTACTTA	GATTTGATTT
17221	AAACATTGGG	CCAGGATCAC	TGAGAAGGCC	CAGTGGGTAA	AGACACTTGT	TGCCAAGTCT
17281	GATTACCTGA	GTCCCTGTGT	CTCACCTGGA	AGGAAAGAAA	CAACTTCAAT	ACTGTTGCAC
17341	ACAAAAGATAC	ATACACACAG	AGGGAGGGAG	GGAGCACAAA	ACCAATACAT	GAATTAGAAG
17401	ATATCAAAAA	ATGTGGGCCCT	GTGGAGGCCCT	GTTAAAGAGA	AGAGAAGCAG	TAAAAAGATT
17461	ATACGGCAGA	TGTTTGGTTA	AAAAGCCCAT	CCATGCTGGG	CGTGGTGGCG	CACGCCTTAA
17521	ATCCCAGGAG	GCAGAGGCAG	GTGGATTTCT	GAGTTGGAGG	CCAGCCTGGT	CTATAAAGTG
17581	AGTTCCAGGA	CAGCTAGAAA	TACAGAGAAA	CCCTGTCTCA	AAACAACAAA	ACAAAACAAA
17641	AAAAGCCCTA	TCCATTCCAG	CAGGTAGAGA	AGTTAGTATT	TTGGTCTTGA	CTATCTTAGT
17701	TCTTCTAAGT	GGTTAAGTGT	TTTGGAGAAG	TTATAATCCC	TTGTAGCAAG	CAGTGTCAAG
17761	TCTGACCGTT	TGACCATATG	TGGCAATCCC	CAAGCTTTCT	TTAAGCAAGT	GTGCCCCAAC
17821	CATGGCCCTT	AGAACGACTT	TTAAAGTTTC	TTGCAGCCAG	CTAGTAGACT	CTAGCAGTTT
17881	TCTAGACCCT	GAGCACTCCA	CTGCTGTA AAA	AATGGGACTT	TTTTTTTCTA	AGTTTCTTTC
17941	AGTCCGTCTC	TTCTTTGTTC	TATCTACACT	CCAAGTGGAA	GCTTTGCTTG	CAGATATTGA
18001	AAGTATCAGT	CATTATTTTG	GAACAAGTGG	ATAATGGGCA	CAGAGAGAAA	GAGGCAGAGA
18061	GACCTTATTG	TGAAGGTGTT	GTGGAGTGT	GTGAGAGCCT	CTGGGTTGAC	AGCATTATGT

Exon 1
(5' end
unclear)

Homology with
human exon
(cDNA no.
D28588)

18121 AGCCTCATT TAAGCCTTGT GGGCTGTGAA TTTGGAGTAA TCTGTTAGGA GATCATATAC
18181 TTCCTTATAT GTTTACTGGG TGGCATGTCT GGACTGTACT TTTGGGTGAT TTAAAGGGTG
18241 TGTAAACTTC TTCCCAGGAT GTTCTAAACC AGGGGCCAG GCTGGGGATG GTTGATGCAC
18301 GTAATCCTAA CATATTCAAG GTTAAGGCAG GAAGATTGAG AGTTTGAGTC TACCCTAGGT
18361 TACTTATCGA GTTGCAAGCT AGCCTGGGCA TAGCAAGACC TAGTCTCAA TGAACAACA
18421 CAAGGAAATT ATACCACAG GATGGCCCAA ATGAGTGGTT CTGCTGGGAG CTTCACTCAC
18481 CCGAGCCGTT GGTGCTTCTC TGTACTGCTG TGCAGACTCA GGCATGGCCC ATCTCCCTCC
18541 TGTCTTCTTT TCTCCTTTCC TTCCACAGCC AACTCTCAGG TCTTTGTTTC TGAAATGTAG
18601 ATGCAGTCT GTCTCTTAGG GAGCTTTCTT CATTACCAAC CAATGTTTTC ATGAGCAGCT
18661 TTTCCGAAGA CCCCAGGGCA GAAGTTCTTT TTGGAGCCTG GTGGGAAGGT AAAAGTGGTT
18721 GAAAGTATGG GGCACGCCAG GCTAAGCAAG AATGAGGCTG GAGGAGGGGC TGGAGGAGGG
18781 GGTGGGGGTG TCACTCTCGA GAAGCTTCCC TGGTGTCAAG GGTGGACTTG TTCCAGTTGA
18841 CAAATGGTGA AGTAGAAAAC TTTTCTACT GCTTCATGGG ACTCTGAAGG TAAAAACAAA
18901 ACACAGGCGC AAGTTCAGAA TTGGCACCAG CAGCCAGACC TCTCAGTATG CTTCTGGACT
18961 GCTCTCTGTC CCTGGCACA GCATCACTGT GCCTTTCTC GGGGTGCACT CCAGACTGGC
19021 ATGCTCTAAG CTGAGACATT GAGCTTTCTT GGGCTGAGGG TGGCAGAGT ATCCCATCTT
19081 GCTGTTTCTC TGTGTGCCCA GAGGGAGCTC CTGGCAGCAT GGAGACATGA AGCTGTACA
19141 TCACTGTACA AGCTAGCTGA AGCCAGCTG TTGGAGTCCG ATGATGGGCA TCTCATCCCT
19201 CAGCTACCTA TCTTTCTGCA TCAAGACTGC TTGCTGGAC CGCCAGCAG CTGCTGAGTT
19261 ACAGTGCAGA GATTTAGCTT GGGTGGCGAA GGCTAAGTCC TTTGAGAGGC CAGTTTCCCT
19321 CTGTGTGGTG ATTTAGGCC TTCAGGTGCC ATTCATGTCA GCCACAGGGC AAAATTGGGA
19381 AGAATTAGCA GTCATTGCAC GTTTCTAAAT GTCTCTTAG AACTTGTTTG TTGGGGTTGT
19441 CATCTCTCCT GTGGGATTTG GTAAAAATCA CAAAGTGGCT CTTGAGACTG GGAACGGGAA
19501 TTAGGAAATA GCTAGTAGGT CATGCGATTA GAAGTGAAT AATTCTGCCT CATGGTTTAT
19561 GTTTAATTGG AAATTTGGAG AGGAAAAACA CCAAGAAACA GACAAAAGCT TTTCCTGGGT
19621 TTTAGAAGAG TAAATGGAAG CCAGCCTTCC TCTCTTCTC AGGGAATGA AGTCACTATT
19681 CATTTCAGAA GCTGATCCTA GCCTGGGGGA GCTCAAAGAG TGACTCACGA AGACATTGCC
19741 TAAGCAAAA AGGCAGCATG TTTGTGTAGA AGGAGATTCT TGGGGGGGAC TCCTGCTTTT
19801 CTCTATGACG GAATCTGTAG TTCCAGTGT CTTTTTTTTT TTTCTTTTT TAAGATTTAT
19861 TATTTTATTT TACATATATG AGTACTCTAG CTGTCTTCAG ATACACCAGA AGAGGGCATT
19921 GAATCTCATC ACAGATGGTT GTTCTAATT GGAAGGCCT TGCATGGTTT CCTGGGGCTG
19981 GTCCCATCAT GCTTAAAGAG TCCCATCATC CCGGGGCCCC CACATTGATG CCTGTGACT
20041 CAGGCCACAA GCAAACTAGG ACTGACAGCA TGTATGCTTG GACGAAAATA TCCAGAATG
20101 AGACAGACCA AGTGTTCACA TTTAGACGCT CTAGTGTATG GTCTGACCTA ATGTCCCCCA
20161 TTCTCGTGCT AACTGGAAAT GGTAAGAGTG TAGCCTGGAG CTGTAGTGCC GTCTGTGGGT
20221 TGACGTGCTT GCAATGAAGT GTCTGCCCTG CTACAGCTCA CAGGCATGGC CACTTAAAT
20281 GGATTTGGTA TATTTGGATG GAGGCTGAAC CATGCAGATG TGGGAGCAAC AGAAGGACCT GGCACAAGAG
20341 TGAGCCACAC CCCAGCCCTC CTTGGTCTCT AAACCTCCAT TAGCATCACT CCCTCTTTGT
20401 TCCTTCAGGA TTCATTTATT CACAAGTGTA TTAAGCAAAT ATAGAAGGCT CCAACTGTGT
20461 GCTAAGTTCT AGAGGTACAA GAGAAACCCCT TGCCTTAGG AAATTTGCAT TCTTAATAAG
20521 ACAGTAATTA AGTAACTAAT GAGTACTTGC TGGGAAGTGG TAAGTGCTAA GGATAAAAAT
20581 AGGGAGGGAG GGAGGGGAAA TGTGTGGGAG AGGCAGTTAC ACAGGGTGT CAGGGACATA
20641 GTGAGATGGG GACTTCATGG AACCTGAAA GTGACGGGCA TTGAACCAGT TCTGTCTAGA
20701 GTGAGGGGAG GAGGCTGAAC CATGCAGATG TGGGAGCAAC AGAAGGACCT GGCACAAGAG
20761 CCCAGAAGCT CTGGGGAGGA CAGGAAGTCA GATCACATAG GCCTGTGCGT GATGGTTGGA
20821 GCGTCGGCTG TGTGATGGGA TGCTATTGAC CCCTCAACAG AGTGATGATG TGGCTTCATT
20881 TATGTTTGTG GTGTTGAGAA TAAACTGACA GATCAGTGCC GTTCTCAGCT GCCCAGGAAC
20941 AGACATGGAT TAGGTGATGG CGTCAGGGTG AACGCCTCAT GTTTGGAATG TGTTTTGAAG
21001 TTAGAGCTAG AAGTGTTTGC TTCCACATGT GGAIAAAAAA AAAAAACAAA AATCAAAGAC
21061 AATTCCAAGG TCTTAGGGCC GAGCAGCTGG GTGGATGTGG TTAGCCTTCG CAGTGGGGAA
21121 TACCTGAAAA AGCATGGAAT CAGGAGCCAT CGGAACCGTG ATTCAGGGCT GCTTACATTC
21181 AAGGTGCCCA CTGCTGTAAA AGAGAAGATG CCGGCAGGAG GTGGTCTGGA GATAGAACT
21241 CAGGTCTCTA TTATTCAAGT TCTGTGAAGC ATTTGGCTTTG AAAGCCATGA GAGCTGGCAA
21301 AATCACCAAG AAGCAGAGAA GACAGAAAATG TAGTCTCTGA AATGGATTCT GGAGCTTAGA
21361 GTTTAACTGC CTAGAATTTG ACAAGAGGGG AAAAAACCACC AGAGATTTAG AAGAAACCTG
21421 CAGAGCAGGA GGCTGTCTTG GGCACATAAA ATAAGTTTAT CCAGGACCAA GAACAGTCA
21481 CTCTCAGATC ACAAGACAAG GGCTCAGGTC TCGGAGTTAG CAGTTAGATT CAGAAACATG
21541 AGGCCTGGTA GCAATGGCAG TTTGGGGAGG ATGAATGACG TGTACAGCCT GATTGGAGCC
21601 ACGCTCAGGA ATGATTTGTT AGAGTAATGA GGAGCAGCAA GTGCACAGCC TGCTGGGCTT
21661 AGGGTAGAGC TGTCTTTTGC ACAATTTGTT GTTTGTTTGT TTGTGTGTT TTCAAGAAA
21721 GGATTTCCCTA GCTGTCTTGG ACTTGCTTTG TAGGCCAGCC TGGCCTCAA CTTATAGAAA
21781 TCTGCCTGCC TCTGCCTCCT GAGTGTCTGG ATTTAAAGCG TCGCCACCA CTGCTCAGCA
21841 AACTGCTCAG CAAACATGCA TTTTAATATA CCTAGCCTTC CCAATATTAC AGCAAATGTT
21901 ACACTGGTTT AAAAAACAGG TGTAAGTATG CACATCTGTA TTCACAGCAC TCCAGAGGTG
21961 GAGGCTGGAG GACCAGGAGT TAAAGCCAT GCTTGGATCC ATAGTGAATT TGAGGCCAGC
22021 TTGAACTCCA TGAGACCCCTA TCTCAGAAAC CAAAGACTAC TTAGGTTTGC TGAGAGTCA
22081 ACAAAATCAT TTTCCACAAC GTGTTTTATA ACAAAGTGTG TATCATGTAA CTTACTGGAT
22141 ACTGTACTAA ACTTAGAAAA CATAATTATG AGTGTGCTTT CATGTACTAT GACATTGAAA
22201 AATGAAGTTG GAAGCCTGCC GGGTAGCTCA GTGCTTCAAG GTGTTAGCCT GTCATGTCAG
22261 GCTGAGTCTG ACACCACGGA TTCAACCCCA GCACCCACAT GGGGGAAGGG GAGAAAACAAC
22321 TTGTAAAAGT TGCTCTCTGA CCTCATGTTG GCCATGGCAT GTGTGTAGCC ATACACAGAT
22381 GTCCACACAT ATACACACAA ATAAAATATG TTTAAGGTGA GAAAAAAGA AAGTTAAGTT
22441 GAACCAGAGG AGTAAGGAGT CTCCTTGGTG ACTGAGGAGT TAGAGAGGTT AGGAATCCCG
22501 GCTCAGAAAG GAAGGGTAAAC TGCTTGATAA AGTCTGTGTG GTATCTCAGG GTCATGGCCG
22561 AGCGGTGGGG TCAGGCTTGG GCTGGAGAGG GATCTGCTC TCTCTCTTG CCCTGTGGCG
22621 TTAGCAGCCT CTGCTCAGAG TTCTGCAATG AAAGGCATGA ACTAAATCTT CCTGTGTGTG
22681 GTTGCTGGGT TCAGTTTTCC GGACAGTTTG GAGCCTGTCC CAGCTCAGCT TGCTCCCTCC
22741 ACCCTCCATC CAAGCTGATT CATTCACTAG TTTCCCTAGA GGTCCAGTTT TGTCTTTTAA
22801 AACCACTGTG AGCTTGGCTT TTCCCTACCA GCTCTGAGTT TTTCCAGGCT CTTCCAGGCT
22861 GAAAGCCTGG CCTTAGAAAC TGCTTTACTG GCTTCTGGCC CAGCCAGACT GACCAAAATA
22921 GCCACCAGAA GGCATGTTTT TGCAGAGTGA AATTCCTGCT GGGGAAATAC CATAGTGGAG
22981 GGCTAAGGAA GCAATAGTTG GGAAGGGTTT AAAAAAATAA AAAAGCCAGC TAGCCTAGGG
23041 TCTAAGTGTG GACTTTGAGC CCCGTTCCAT CCTAGGATAA CCTGATAATC AGGGCTCAGG
23101 GTCCTTCTTA TTATGTTCTC CCAACCCAAT CGGCCAGAGA AAGCCACCCC CCTCCAGGCT

23161 AGAGCCATCT ATGCTCCCTG CTCTGTGGTG AGGCCAGCAT TGCCAAGGGT GCTACAGGAG
 23221 TGTGCACTGG CACCCAACAA GGTTCATCTGC TTCTGGAGAG TAAGTGAGAA ACAACCTCTA
 23281 GGGATTATAA GTGGCCAAAG GAGAGCGCTG GCTTCAGGCT CCCAGGAGAG AGAATGGCAG
 23341 GCGAGGAGGG AGGAGGCTGG GTGCTTGGTC CCTCTAAGCT TATCAGAAGA GCTTCCGAAG
 23401 AAAACTATGG GGGAGGTGTG GTCAGGGGAG AGGGTAGAAG GGTCTTGGGT TTGCAAAGAT
 23461 AGTCTTCCAA AAGTGAAC TGCTGGATAAA GGGAAAGTGG CCAGAGGCTG AAAGCTGCCC
 23521 TGGAGGCTGT GGATGGGAGC TCTGCCTGAG CCTGAGGGGT CATTAAAGCA ATGCCCTGGG
 23581 ATTACTGAAC AGCAAGGAGA GAACTTGTGA CCAAGAATGA GGAGGACCA GTTACCCACC
 23641 AGGAGGGTGA TGCTCAGGGT CCTGCTGAGA TTGCAACTAC ATATCCAAA GACTCTATTC
 23701 TCCCCACTCT TGCGGAAC TGCTCAGTGC CATCTCAGTC CAGAAAGGAC CCAAGGAGAC GATATACCTG
 23761 TCAAAGAGCA GGCCCTTACT CCCATCCCTT AGGCTGCTCT CACTGTAATG GGCCTGGAT
 23821 GCTAGCAAAA GTCCCTGACTT CATCTCAGTC CAGAAAGGAC CCAAGGAGAC GATATACCTG
 23881 TTACAAAATG GCTGTCAATA AATTTTGTGA GGGAGGAAAA TGAAAAGTGG CCTAAAGGTG
 23941 CAAGCATCAG CTGGA AAAAGT CAGAAGAGCC CTAAGCAAGG GAAGGGTAAC AGAGATAAGC
 24001 TCCTGGGATG CAAAGGGAGG GCTGCGGAGG CTGGGGAGGG GAAAGAACA ACAGGCCAGG
 24061 AGGGAATGCT CTGGGTGGG AGTGTCTGGC TGACCCAGTG AAGGGCCACA GAGGGCTGGG
 24121 CTGAACACAG CCCTGGGAGC GAGCCTGTGG AAGCAGTCTG AAGCCACGTG AACAGACACA
 24181 GTCCCCAAG GAAAGAGGTC TTGGGTGAGA AGGGTCGTTT GTTGAAGAAT ATCAGTATTT
 24241 CAGTGATGGA GTGAATTTGG AGGGGGCGGC AGCCTCGAGA GAGTCTGGT GACCTGGTGA
 24301 AAGCAGCCCT CTGAGTGTGG AGGGAGGTGT GGAGGGGACG GCAGCGTGAG GCAGGAGGAG
 24361 AAGGCTGTAC TGGAGGGCAA GAGATGGAGG CGGATGCTCT TAAGAGTCTG GACATTAGGT
 24421 GCTTTTCTCT CTGTCTTGGG TTCTGTTCGG CCTCAGTTA CTTGATACTC CTTCTCCCT
 24481 GCTGCCTTGT TCACACCCTC AGGACTTTTC CTTAGGACAG AGATGTGCCG GTCATTATG
 24541 CTTGTATCAA AGATCACTGC CACTTTCAG CTTAGAGACT TGGGGACGTT TTAGTGAAA
 24601 TGAATGATAA AATCCAGATT CTACAACCTG GAGACACGCA CTTTTTCTCT TCAAGAAAT
 24661 TTAGTAAAGT TGTTGATAT GGCTCAGCAG ATAAGGTGAT TGCTGCCAAA CTTGATAACC
 24721 AGGGTTCAGT CCCTACAGAA CCCACGTGGT GGAAGAAGT GGTTCAAAAT TGTCTGTGA
 24781 CTTTCAATATG TGTCAAACA CACAAAAAT AAATAAATGT AAAAAAGGAG AAAAAAGCAG
 24841 TAATCCTTAT TACCATATT GCTAAGTTGT AGGTTTCAGT GACCACATAC TACGGTACTA
 24901 TTA AAAATTTG TGAGCTGTGC TGGAGATGGC TCAGGGGTAT AGCATTGTCT GTGCGTGTG
 24961 GAGGCCCTGA GTCAAACCCC TAGAACAGGA AGAAGACCGT GTAGCATAGT GGAAGTGGTG
 25021 TGAAGTGATT ACAGAATCAG AGCTGAGCCC CTCTGTGGCT TTCTATCCCA CTCTGCAGTG
 25081 CTCTCAGTGA CTGGAGGAG CATCAAGCAT GACTTACATC ACATTACACA CGTAAACAC
 25141 AGGGTTCAG GTGGAGTGAG GCTGCTCGGA GAATCTGTGG GTACGTGTCT GCCTCTCCT
 25201 ATCCCATCC ACTGCCTCAC ACCAGCCAGA AGTGTCTCTC TCTTTCTTCT TCTTTCTCT
 25261 TCCAGAGTAG AATTTCTGAG TGGCAGGGAA ATGTTTCTCC CATTTTGGCC CAGGGGCCAC
 25321 ACAAACTCTC CAGTGACAAT AGGAGACAGC CATAGTGGTT TAGACTAAA GAAGTATTT
 25381 ATGAGCCAGC TGGGAAGAA AATGAAACTA GGATGAGGTG TGCAGCCGGA AGCAGCAGCC
 25441 CGCAGCCTCC GCAGTAGGCT CTGTCTTACC CTCTGACCCA GTGGCTCTCA ACCTCTCTAC
 25501 CACTGCGACC AGCTGATACA GTTCCTCGTG GAGTGACCCG CAACCATAAA ATATTTTGT
 25561 TGCCACTTCA TAATGTAAAT TTTGCTATTA TGACTCGCAA TACAAAATATC TGATATGCAG
 25621 AATATCTCAT ATGTGGCTCC TACAAAAGGG CCGACTGACA TCCCTTCAGG GGTTCATGACT
 25681 CACAGGCTTA AACTCTGGAG TTCAACCAGG TTTCCACACC GATCGATCCA GGGCTACAGA
 25741 GGGAGAGCGA AGCGGTAGG AAGTGTGGT CTGTCTTTGT TCCACTTCTC TTACTCTGC
 25801 CCACTGCAAG TCTCTGTGTT TCGTAATACA GGAAGGACAT CATTCACTAA GTTTTGTAA
 25861 GTAAACTTGA ATCCTTGAGA GCCTCCTAAG AGACCTTTGT GGGGCTTTTG TTTTGTTTT
 25921 AAAGATTTAT TTATTTTATT GTATGTGAGT ACACTGTAGC TGTCTTCAGA CACACCAGAA
 25981 GAGGGCATCA GATCTCATT CGGATAGTTG AGCCACCATG TGTTGCTGG GAATTGAAC
 26041 CAGGACCTCT GGAAGAGCAG GCAGTGCTCT TAACCGCTGA GCCATCTCTC CAGCCCCAAC
 26101 CTTTGTGTTT TCATCAGCAC AACCTGGACT GAAAATTTTT TGTGTGTTTT GTTTTACTGT
 26161 GCTAGTATG TAACCAGGGA CCTTGACATC GTGAGGCAGG GTCTCTCCCA CCTGATCCAC
 26221 ACCCTTGCGC CTCAGGATAA ACATTTGAAA GAAAATGTAG TCCAACTTTT CACCCACCT
 26281 CTATCCCATC TCCAACAATA TACATCTTGT CTCTGGTTAA ACAATTTTGA AGAAAGGGAG
 26341 ACATACCCCC CTCTGAGACA ACTGGTTTCAG TCTTTCAAAG GTATCTTTGA GTTAAACAGA
 26401 AACCTGGCTT CTTGAACATC TACACGGGTG GCTTGCTTTT GTCTTTTGA CTTACAGAGT
 26461 GTGGTCAGAA GCTTGAAGT CTCTCGTTGG AAATACCCCT TAGAATGGGC CCAAGACTGA
 26521 ATATAACCCA GTGTGGGGAG TAATATATAG GGAATATTAT TTTTCTTGT TATAATTGT
 26581 ATTTTTTAGT TTTTTTAAA ATAGTTTTTG TTAGCCAAAC ATGGTGGTGC ATGACTTAAA
 26641 TTCCAGCAGT CGGAGGCAAA GGCTAGACTA ATGTGATATT CAGTGAGTAA ATAGTTAATT
 26701 AATTA AAAATG ACAATTTACCA TAAGAAAAA AATGTCAGTC AAAAGACCCA AGATGCCTGG
 26761 CCATGGTGGT GCACGCCTTT AATCCAGCA CTTGGGAGGC AGAGGCAGC GATTTCTGA
 26821 GTTCGAGCCA GCCTGGTCTA CAAAGTGAGT TCCAGGACAG CCAGGGCTAC ACAGAGAAAC
 26881 CCTGTCTCAA AAAACCAAAA ACAACAAAAC AAAAAAAGAC CCAAGATGCC TGTGGCCCT
 26941 GGTCTGTGAG CAACTGTGAG TTGGATACCT TGGCTCGTAC CCACCAACC TTAGCCCCAA
 27001 ATGATATGAT CACCCTTCA TTCTCTCCTC TCTCTCCATA AATACTATAA GCACTACAAT
 27061 GTCAGTGAA ACAGAAATGT TGGTGCATGT CTTTTACAGA TGGGCCATC CAGAACCTTT
 27121 AAAAGCAAAT GGCTGTGCTT ACAGGACTAG TTTTGTCAAT TGCATGCTTT TTTCTGTAGT
 27181 TGTTCCTTCA TAACTGCAAT GATTTTACAG GGATCCGTTG CTGTAATATC CTAGAATCTA
 27241 TGTGTCCACT GCGCAGAACC ATCCACTGGC CTCCAGTTG GCAGCATCGC AGGATCACCC
 27301 TTGTTCTGCT GCATCTCATG GCTTGGGGCC TGGTGACCAC TGAAGGCAC GTGTTGCTTC
 27361 CTGGCCAGCC AGCCTGCCTG TTTCTCTTTC TTTACCTTCT ACCTTTCAGG GGCCTGAAA
 27421 AAGCCATGAA TGGTGGTACT TCTGTCCCTCA CCAAACATA GTTCTGTTTC TGCTGAGCT
 27481 CTGTCTCAT AGACTCCTGA CTTGGGGTCA GTTCTGAGC TTCTGTCCCT TGAGTCTTCC
 27541 CCATCTCTCT GTTTCCTGGA ATTTGCTTTC TTAATGTCTA AGATACCAGC CTGACTTCCC
 27601 AGTTGCTTTG CTTTAGTGAT TTATGACTCT TACCATGGCA GAGTCTTAC CCAGAGCTCT
 27661 TAGGGCTTCA TTTTCTTTTC TTTCTTCTTT TTCTTTCTTT TTTTGTGTTG TTTTGTGTTG
 27721 TTTTTTGGAT TTTTTTTAAT GGGCTCACTA TGTAGACCAG GCTGGCTTCA AACTCCTAGA
 27781 GATCCAGTTT CCCCTGCCTC CCATGTGTCT AGATTAAGG CGTGTGCCAC CACATCCAGC
 27841 CTCTTGCTTG CCTTTATTTT GTTTTAAAATG CTATGTGTAT GGTTTTTCCC CTGCATGTAC
 27901 GTTTTACACC ACTTGATAT GTGCATCCTG CACTGGAGGC TAGAGGAGGC CAGCCCTTGG
 27961 AACTGGAGTT ACAGAGGATT GTGATGACCA TGTGGGTGCT GGGACTTGA CATGGGTCTT
 28021 CTGGAAGAAT AGCCAGTACT CTTAGTTTCT CTCTTATCAA AAGTAAGCGG CTGGCCCTGG
 28081 TGGCACACAC CTGTAATCCC AGCCCTTGAG AGACTAGTAA ACAGTCCCA GGCCACCTGG
 28141 GGATATGGTG CAAGATTCTT GTCTCAAATA AAGTAAAAT AAGTTTAAA AATTATAGGA

Start of
 fragment
 "Intron 3"

28201 GGATACAGTG ATCCCTTTGGT ACCCACAGGA GCTCCTGGAG ATTCCAGAAC CCGTAGATGC
28261 TCAAGACTTA TATATAATGT ATAGAAGTTA CCTGTAAGTT GCACATCTGC CGCATACCTT
28321 AAGTCATCTG CAGATTATGT AGCACCAACA TTTCTTGATG CATGGTTGGT TGAATCTGTG
28381 GCTACAGAACT CTGTTTGATAT GGAGGTCAAC TATCCAGTTG TGGGGAAATG AGAAGAGAAA
28441 AATTTGCGCA CTTATTACTG CATATACACA AAAGTGGAAA GTGTTCATTC TTCCATGTAG
28501 TTGTTAGCCC TGCTGAGTCT ATGACCTACA CAGTCAGATC AGTTGTCTGT GGTAAAGGGG
28561 TTAGTCCACT TAAGGCAGCG CTTCTCAGCC TTCCTTACGC AGCAACCCTT TAGTGCAGTC
28621 CCTCATGTTG TGGTGACCCC CAAACATAAC ATTATTTTTT CTGCCACTTC ACAGCCATAC
28681 TTTTGGTACT CTTACGAAGT GTCATGTAAT TACTTTTGGG GAGGTTTGCC AAATGGGTCA
28741 CGACCCATAG GCTGAGAAGT GCTGGCTTAA AGGAGTTGCC GTCTTCTGGT TCCTGGAGGC
28801 CCTTGTGTCT CTGTGTGAAG GTTGTGAAGG TTGAGCAGAC AGACAGCCTT TAGAGTAGGA
28861 CAGCTAGTCT CTGCTGCTCA GGAATGCTGG CCTTGCCTTT TTTCTTTCCC AGACTGGCTT
28921 GCATCTAGCT CCCGCTGGAT TGAGCTGGTA GCCTGCCTTA GCACAAGCCT AGAGGCTCTG
28981 CCCCAATGAA GAGATTCAGA TTTAGAGGGC TACCAGTGTG CACCAGGGGG CACTCTGGAC
29041 TCAGCACAGC CGGACCAGCA CACGAGTCTC CCAGTTCTCC TCATCTTAGA TGGTGTGGGG
29101 AGTGTGGTGA AGGCAGGCC TTAGCTGAATC GATGAGTGTG AGCAAAGGAA AAGCTCACCC
29161 TGGCAAGCCT GGGGTGGGA TGCTCGAGCG GGGCCTTGCA GGAAAAGCTC TTAACCTGTT
29221 GGGAAAAGCA CACAGAGGC AAAAAAATA ATTAATTTAA TTTAAAAAGC GGAGGCCAAG
29281 GGCCCTGGGT TTCTCTGTGG TGTGTTTCTT ACATCATTC ACCTTTCTCC CAGCAGATCC
29341 **ACAGATGAGC ATGGCCGCCA CTGCTGCTGT CAGTCCCAGT GACTACCTGC AGCCTGTGTC** Exon 3
29401 **TGCTACTACC CAGGCAAGTG CACGGTGCTC TCTGAGGGCA GGGCCAGAGC CCGCCGGGTT**
29461 GATGAAGTGG GTTTTCTGCC TTACTCAGAC CCGCTCGCTC TCACTGCAGA CCCCAGGGA
29521 GATTCCTTTG TTCTCATTTC TGAAC TAAGA CATAGACACC TGTCTAAGAT GTCAAAGTTG
29581 GGTCAGGCAA AAAGAGTTGT GTTACCCGTG ACCTCTGGGG CTGGGGCCAT GGCTCTTAGT
29641 AAAGTTGCTT CACATGCATG AGTTCTTGAG TTCAATTCCA AGTTCCCGTG TAAACCCGTG
29701 GGCTGGGCTT GGTAGTACAC ACCTATAATC CTGGCATTTC AGAGGCAGAC AGAGGGGCTC
29761 CACTGGGCTT ACTGGCCAAC TAGTCTTGCC CAGTCAGTGA GCTTCAGATT CAGTAAGAGA
29821 CCCTGTCTCA AAAATGACGG CAAAGGAGGT GGAGAGAGCC AGGCGTGGTG GCGCATGTCT
29881 TTAATCCCAG CACTCGGGAG GCAGAGGCTG CCGGATTCTT GAGTTCGAGG CCAGCCTGTT
29941 CTACAAAGTG AGTTCCAGGG CTACACAGAG AAACCCCTGT TTGAAAAGCC AAAAAAGAAA
30001 AAAGAGGTAG AGAGGTGCTC AGTGGCTTAA GAGCGCTGGC TGCTCTCCA GAGGCCTGGG
30061 TTTGATATGA TAGCTCATA TCCAGGATCA AGTGTCTGTA CACCCTCTCT GGCTCTGAA
30121 AGCACCAGT ACAAGTACTT GAGTAAGGCA TAGACATA TA TGTGGGAAA ACACCCAT
30181 ACATTAATA AATAACAATT TAATAAACAA GGGAGAGTCA AAGAGTGAGA CTTCAATTAAC
30241 ATATCTGGCC TCTACTACT CCACCTATGC TGATTCGGCT ATACCCACAC ACCTCCACA
30301 TGAACACATA CATGTATATA ATACACACT CAGAACA AAA CCAGAAAACCT TTTACCGAGA
30361 GCAAGTTTCAT GTCATCTTCC CTCCCATCA GCTCTCTTGT CCTCTTGTGA CTGAGTTGGT
30421 AGAATTCCAA CTGATACCTG GGTGTGTTGT TATCC**CTGCC** GACTAGGC CAGCAGGGCG
30481 GGTGGAGGGG GGGGAGAACA GCCTGTGTTG TTGTGTGTG GTGTCCAAGG CTGCTTCTGG
30541 GGGGTTTTAC ACTGAGTTCC TTCACTTCAT GGAGAAACA GCAGGAGCA ATGAGCCAAT
30601 GCAGGACACA CCGACGTGTG TGCGCACGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG
30661 TGTGAGTGAT AGCTTGCCAC AGGGGTGGGG AAGCTGCTGA GAAGAGAGGC TCGTGAGGGC
30721 AGAGTGTCTG CCTAGTCAGC CCATCTCACT CCTTCTTCT GTTGATTTTT GCAG**GACTCC**
30781 **CAGCCATCTC CTTTGGCCCT GTTTGTGCG ACATGTAGCA AAATTGGCCC CCCAGCTGTT**
30841 **GAAGTGCAG TGACACCTCC TGCTCCCCC CAGCTACCC CAAGAAACT GTTCCCTATC**
30901 **AAACCCGCTC CTCTCCCTCT CAGTCCCTGC AAGAATAGCT TTAGCATCTT ATCATCTAAA**
30961 **GGAAATATA TCCAGATTCA GGGCTCCCAG CTGAGTACCT CTTACCCTGG AGGGCAGTTT**
31021 **GTGTTTGC AA TTCAGAACC CACCCTGATC ACAAAGGGA CCGGATCAA TGCAAGTATC**
31081 **CAGTACCAAG TCCCAGAT CCAAGGGAAC AGTCCCAGA CCATCCAGGT GCAGCCAGT**
31141 **CTCACC AACC AGATCCAGGT CATCCAGGC ACCAACAAG CCATCATCAC CCGTCAACA**
31201 **TCTGGTCATA AGCCTGTCCC CATCAAGCCA GCTCCTGTCC AGAAGTCAAG TACGACTACC**
31261 **ACCCTGTGTC AGAGCGGGGC CAACGTGGTG AAGCTGACAG CCGGGGGTAG CAACATGACG** Exon 4
31321 **CTCACCTGTC GCCTCAACAA CCTGGTGAAC ACCAGCGATA TTGGGGGCC TGCTCAGCTC**
31381 **CTCACTGAGA GCCCCCCAC CCCACTGTCT AAGACTAACA AGAAAGCTAG GAAGAAGAGC**
31441 **CTCCCTGTGT CACAGCCATC CGTAGCTGTG CGTGAGCAGG TGAGAGCGGT GCTGATAGAG**
31501 **ACCACGGCA ACAACATCAT CCAGGCAGGA ACAAACCTGC GATCGTCCA GAGCCCTGGC**
31561 **GGGGGCCAGC CAGCTGTGGT CCAGCAAGTC CAGGTGGTGC CACCCAAGGC TGAGCAGCAG**
31621 **CAGGTGGTCC AGATCCCCCA GCAGGCACGT CCGGTGGTGC AAGTGCCTC TGCCACCCTC**
31681 **CCCACCGTCC CCCAGAAGCC CTCACAGAAC TTTGAGTACC AGACA ACTGA ACCAACACCA**
31741 **CCCAGGTAC CATAGTCCCT CTGCCCTGCC TCCCCTCCT GGTCACTCT TTTGGCTTG**
31801 TTTTGAAACT GAGGCAGATC ATCGTCAGAT CTTACGGGAA GCTGGAGTCA CCATCATTTG
31861 GGGTGGGCAG GAAATGGCTT CTCCCTGGG GTCTCTCAG GTCTTAGGCC AAGTAGTCTT
31921 GACCTGTAGA AATCAGATGG AACAGGGATA TCTGGTCTGT TTTATTTCTT CCTTATGGGG
31981 TAGCTCTGGG GGCATAATGG GGTAGATCTG TATACTGTA TAATAAAGTA TTTTAAATGT
32041 AAGAGTATAT ATAACAGCAA **GAATGAGAA** TCGAACTAGA ACCATAGAAT CTGCTGTAAA
32101 GTAGGAAAGT TCAAGAGAAA AAAATGCTGG CCTCTGCTAA AGACAAGAGC AATAGCAGTG
32161 GGGCAGCCAG CCCTTTTCTT TTCTGAGACA GGTCTCCCTG TGGTCCCAG GTGGCCTTT
32221 ACCTCGCTGC AGACCCCAAG CTGGCTTCAC TCGCTGCACA GCCCAAGCTG GCTTCAGACA
32281 TGGGACCCTC CTGCTCCAGC CTCAGTGCAG TGCTGGGGTC GCATGTGTGA GCTGTAGCCT
32341 TGCTTCTTTG CACCCCTGTT TATCTCCCC GAGAGGGGGT TGTCTTCTG CTTCCTTAC
32401 TGTGTGTGCA TGTGGCCACT GAGCGTGTGT TTTCCCCTCT AGTCTAGATC CCTCCTAAG
32461 GAGAAGACAT TCGTCTTCTT GAGCCTGCCA TGTTCACCT ACCATGATGA CCTCTGGGTT
32521 CTGGAGTGTG CTCACGGCTG CTTTGTCTTC TGCTGCTGTC AGGGCAGCTG GCCATTCTGC
32581 CAGCTGTCAA CTCTTGTCTT TTTAGCTGCC CTGTGGCCGC CATCTTGCTG CTTTGGTAGC
32641 CTGCCCTGCC GGCTGCTTAG CTCTCTGGTC ACTTTGCACT CTTCTGCTGCT GCTTTTCTT
32701 TCCTTCCCTT TTTTTTTTTT TGTAAGCCAG GCTGGCCTGA GACTCAGAAA TCTGCCTACC
32761 TCTGCCTCTT CAGGGCTGGG ATTAAGGGCG TCATCACCGC CACTCAGCTG TTGCTCCTTT
32821 ATGTTACCAG CTCAGCCCTAC TCTAAAAGCA GACAAGAACA CACAGAGGCC AGTGTGTGCT
32881 AAGACAGACG GGGCAGTAAA GGAGACTCCA GCAGCCATGG GAGTCAAACCT GCTCCATTCC
32941 CCAGTGTCTC TGCTGATTTA CAGGGAAATG GCCATGAGAA GGAGTGCTCA AAATGATCGT
33001 TTAATAACTTG ATGCAGGTGC TCTGGTGCAC ATCTGTAAC ATGCATTGAG GGGCTGAGG
33061 CAGGACTGAC ATGCGTGTGC ACCAGGCCCA GCAATGGGGC TTGAGCGATG GGTGTGGTT
33121 GACTCGCCGC TTTTCCAGA GGTTCACAGG TTGATTCCCA GCACTCACGT GGCTCTCAAC

Exon 3

End of fragment
"Intron 3" and
start of
fragment "Exon
4"

Exon 4

End of fragment
"Exon 4", start
of fragment
"Intron 4"

33181 CAGCTGTAAC TCCACTCCCA GAGGATTCCG AGTCCTCCTG TAGTCTCCAT CTGCACATGA
33241 TAAACAAATG TGATAAAGAC ATTAGACAAA GCACCCCCAC ACAGTCAGGT AATAATAAAA
33301 CAAACAGTGC TTAACACCCA CTCAGTACCC ACCAAGGGAG CAGTGGGCCC TCGCCTGACT
33361 CAGGGTGTCT GAGGCTAGCA GAAGAGAAGA CAGAAGGGGC AGGTCTCTCT CTCTCTGCGC
33421 CCCAGTTACT CCCAGCCA TTTACTACCT CAGAGAAAAA ACAAGTACAA CATGGTAAAG
33481 TTTTAAATG GGGTACATAT ACAGAACTGG AAAGGGACCT GAGGCTTCTA TATGTATTAA
33541 GGGATGGCAG GTTCAGATGT TATCCTTTAA CTAAAAGTCT TCAGTATCCA GGAATTAAT
33601 ATAACCCAG TTAATAATCAG TCTTACAAAG GATTTTTTTC TCCTGGATC ATTTAAACAT
33661 AAACCTGCCAC CATGATCCCC CTCGCCACTTG CAAATGCTTT AATFCTCCA CACAGGCACA
33721 GCCCTGCGC ACACCGCAGC CTTCAAGCCT GAAACTCACT AGTGAACCTG TGGTGTCTAT
33781 CGGGGAGGTA CCATTTCAGGT CTCACCACAT GCTTCGGATA GTAATTCCAA CGCTGCCTAA
33841 GGACTCTGTC CACCTTACAT GTAGTTGTGG CATCTCTCTG TCAGATCTC TCTTTCAGAA
33901 AGAATTCCTC GCATCCCCCA CCTCCATGGA GCTGGAAATT AAACCAGGGC TTCGTGGATC
33961 CTAGGCACTT AAGTGTCCCTA CCCTGAACCTA CATCTCAGC CCCACTTCTT CTTAATGTG
34021 GCAGTCAACC TTGTCTGTCTG TGAAGGCCAC AGGCTAGTTG TTTTGTAGAA TGCTCTTAG
34081 TTTGTTTATG TGATGTCCCA TGAGTGGATT GCATTTGCCA TCCCAGTGCT CATACAGAGA
34141 GAGGTGAGAG CAGAAAACAGG AGGACCAGAC AGGAACCAGC CAGCCTGGCA TATGCAGTAG
34201 CGAACAAAGAA GCCCTACTTA CAGGAAGGCA AGGTGTCTCT GTGACTTCCA CAAGTGTGCT
34261 GTAGCACATG CGTACCACATA CTCTCACACA TAAACACATA CACACACACA CATACACACA
34321 CAACATAAAA TTTTTCCTC CTTGTATTTC ATTAGTATT TGAAGAGAGA CATTTTGGGA
34381 TAAGATAAAT ATCGAGTTCC TCTTAAACAA TTTGCTTAGA TTATTTTATT ATATATATAT
34441 ATGTATGTAT ATATANNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
34501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
34561 NNNNNNNNNA GGTTCGCTTG CATATATATG TACGTGACT ATGCCTAAC ATAGTATGTG
34621 AACATGCCTC ATGCCTGAGA AGGTGCGCAG AGGCATCAG CTCCTCTAGA ACTGGAGTTA
34681 GGGATGCTG TAAGTACTG TGTGGATGCT GGGAAATCTG GTCTCTCTGG AAGAGCAGCC
34741 AGTACTGCTA AGCTCAGGCC CATCTCTCCA GCCCTGCTG TTTAAAGATA GTTACTGTG
34801 TGGTTAAGCT GGCTCCACAT CACTTTGTAG CCCATGCTGG TGGTGTGCTT TTTTATTAT
34861 CTTCCAGTTT TGAGATTTAG CAGTTAATAT TTTTCCATGC TAAGCAGATT AAAAGTTTGG
34921 TTGCAAGTCA GTAACCGCT CATGCCGATT GAGTGCAGG CAGCATCGGC CCGTTAAAC
34981 ATGAGTAGTG CACAAGTGTG TTCTTGGGCT GGATGAGATG ACAGAGGTGT CCCAGTTCTC
35041 CCTCGGACAA GGAAGGTCAT GCTGCCATCT GTCTTAGAGC ACTCTCTTTT CTCTGGCCCC
35101 TCAGGTTTAT **ATCCGTACAC GTTCTGTGA GGTACAGACA GGTCTGTCTC AGGACAGCCC**
35161 CCCAGCAACA CTGTCAACCA CATCAACCGT CACCTGTAAC AGCCTGCAC TCCGAGCTCC
35221 CCATCTGAGT GGCACCAGCA AAAAATTC AGCTGCCATT CTCCGAAAAG AGCGGCCCTT
35281 GCCAAAGATT GCACCTGCTG GGAGCATCAT CAGCCTGAA CCCCACAGC TGGCAGCAGC
35341 TGCTCAGGCC ATGCAGACCA TCAACATCAA TGGTGTCCAA TGCCAGGGTG TGCTGTCCAC
35401 CATCAACCAAC ATGGCGGTG AGGGAAAGGC CTGAAGCCAG GAGGGTACTG GGAGCAGGGC
35461 TTTGCATTCA GCATATGGGA ATCTGAGCC GGGCCCTGG GGTTCAGAG GTGAGCCAGG
35521 AGGATGCTCA GCTTGTCTCT GTCCCTGCC AGCCTCAGGT CATATCTGTC TTCACACTTG
35581 GTCCTTACAG TAATACTGTG AGATAGGAAG TCACATGACC ACATGATAAC TGCAGATGTT
35641 GAATAGAGAC TGAGAAAATA GATGTGTCCA AGTAATATAC AAATTCACAT TAATTTGGCT
35701 GGTCACTTGC CCAGTCCAGG TAGCTAAGGG CTGTACCTAA TGTAAGGAA GTCAAGGCTT
35761 CTTTGGGGT TCAAGGGTAG TGTTCGAGA CAGTGATTCT GTATAGTCTT GGCTTTCCTG
35821 GACCCTGTA GACCAGAACA CGCTGTGCC CTTGAGCGCT ATTATTAAG GCATGTGCCA
35881 CCACCCTCCA GCTAAGGACG GCTGTGTGTG AGAGTGGGAG GCAGGGGAGG GCGTCCCTGG
35941 CAGCAGGACC AGCTTACACT GCACCAGAGC TGATGATGAT ACATTCTGCG AGTGCAGCTT
36001 TGAGAAAAAG GGAAGCGAC TTGGGAGATG GCTCAGTGCT CAGCAACACC GGCTGCTCTT
36061 CCAGAGCCTG TGTTCATTCC CCAGCACCCA CATGGCAGCT CACAACAGTC TAGTCCATA
36121 GGATCTAATG CCCTTTCCTG GTATGTAGAC ATACACACAG ACAAACATC CATATACATA
36181 AAATACATAA AAAAATCTTT AAAAAATAA AAAATGAAG AGAGAGGGAA GGAGGGAGGG
36241 AGGGAGAGCT GAGGCTGCCT AGCGAAGCAG GCTGCATCTG CACCTGGGAG TCTGCAGCGC
36301 CCCCGCGGCG GCGGAGAGCT CACGCTTTC CACATCCCTA ATAAACGTCG AGCATCTTAG
36361 GCATTTGCTT GGTCCACAGA ACAAATTTT TTAGGCGAGT TTCCTGCTAG CAGGGGCGA
36421 ATCTCAGGGA GGCCTGTGGC CCTCACTGCC CTCCAGTGTA TTAGGCTCCA TGAGTCTTCC
36481 TCAGCCAGAG AGAGGGAGAG GGAGGGAGAG AGGGATAGCC TGGGCTTCTT TTCCTGAACC
36541 ACACCTTTT CTCTCCACA **GGGCAGCAGC AGCTGACCGT GCAGAACGTT TCTGGAACA**
36601 ACCTGACTAT CAGTGGGCTG AGCCCCACC AGATCCAGCT GCAGATGGAG CAGGCTCTAG
36661 CCGGAGAGGC CCAGCTGGG GAGAAACGGC GGCGCATGGC CTGCACGTTG CCCAACTGCA
36721 AGGATGGGGA GAAGAGGTAA TATGCCGAC CCTGGCCATA CCCACACCC TGTTCACAG
36781 GTCACCTGGG GACAAGTCTT GGCCTGACCT CCGGTGCTTC TCTCCCTTCC CTCTCCCAA
36841 CTCCTCCCC TACTCTGAAC CCAACCCCTC **AGCTTGGGG AGCAGGGCAA GAAGAAGCAC**
36901 GTGTGCCACA TCCCAGACT GCGGAAGACT TTCCGGAAGA CATCTTGTCT CCGGCCCCAC
36961 GTCCGCTGCA ACACCTGGCA GCGGCCCTTT GTCTGCAACT GGTCTTCTG TGGAAAGAGA
37021 TTCACCTCGGA GCGATGAAC CCAGCGACAT GCTCGAACCC CACACAGGTC AGTTCCTCCA
37081 TGCCCTGTG TGTCTTTGT ACACCTGTAC CCATGCCTGT TCGGGCTTC ACCATCCACA
37141 CTACCCAGAC CCAAAATCAA AACCAAAACA AACCAGCA GTGATGAGAA ATGCCAAATG
37201 CTGTGTGTAG CATTTCAGA TAAGTTAGTC CCAGGAAGGC CTTGGAGGGT TACCTCAGTG
37261 CCAGCCACCT GGCTGAGGTT TGGATAAGTG GCTAGTGGTG GGCAAGCTGG GCTCCTACCC
37321 AAGCCCTGGG CTCTCAAGGC CACATACTTT GTAATGCTT TGTTTTAGGA CTTAAGGTTA
37381 CTCTGTCTCC ACCTCTGTAA AGCCCTGTGA ACTAAGCCAC CAAGCTGGCC TCCAGTCTTG
37441 GTTTTGGAGA GAGAGCGATA ACAAGATTGA GCCAAGTGT CAGGCTGCC TCGAGCAGT
37501 CCTGCCTCAG ACTCCAGCT TGGGATTAC AGGTGTGAGT ACCACACCC AACTCTTTT
37561 GTTTGTTTTA AGACAACATT TTGCTGTACA GCCCAGTTT ACCTTAACT CGTGGCACTT
37621 CTCCTATCTC CTTAATGCTG AGATTACAGG CCTGCTCCGA CATGCATGGC CTACTGCAGG
37681 CCTGCCTCCA TGGCAGCCTA AGTTTACTAT TAAGTTGCT ATCACTATTA TGTGCTTTAA
37741 TGCACCCAGT AGAAGGTGCC TTAAGATATG CTTGGGCATG CCCAGTGGTC CACTGGCAGA
37801 TGCTGTGTGA GGGAAATGAT CCCTGTGAG CAGACACTGA GGTGGGGGAC ATGCCTCAGC
37861 TCCTGGTGAG TCACACCCAT CGCAAGCAGA AGCATAACA CCTGATACA GGTGGGTCG
37921 CTGTGGTCCA GACTCAGGGG CTCAGGGCTG AGATGTCTTC CTGTAAGGAG GAAGGTGTAA
37981 AGTCCCTGGG ACACAGTCCA GAATGAAGCA GGTATGCAGA GGCTCAAGGA CAAGCAGCCT
38041 GGGCAAAGAC CAGAGTGTAC CCAAGAACGG AGTGTAGAGA GGTAGATTAG ACGGGGACG
38101 TGGGGCAGAG AGGTGCGACC AATAGACGAC AGGAGGAGGC TGTAGTGGGA GTGCAGAGAC
38161 AAGGCCAAGT TGGCCTTTCC AGGGGCTCTC ACTTCTTGGA GCTTCTCTT TCTGGGACTG

End of fragment
"Intron 4"

Exon 5

Exon 6

Exon 7

38221 GGGCAGTGAT ACTGCATGCC CTCTTTGGTA AAAGATATTT CTTCTTCCCA GGGGACAAGC
38281 GCTTTGAGTG TGCCCAATGT CAGAAGCGCT TCATGAGGAG TGACCACCTC ACCAAGCATT
38341 ACAAGACCCA CCTAGGCACG AAGGGCTTGT AAGGCCAACT GAGGCAGGAG GCCCTGAAGA
38401 TGGAAACACCT CCCATCTGAT TGGCCCTGGG TCCATGATGG ACAGGTGCCC ACGACTGCCC
38461 TGGGCAGCCC TCACCCCACT CCGTGTCTGT GGCTGTGTCC CTGCAGAGAG GGACTCTGTT
38521 CCTGTGTTCT CCTCCTTCCA AAGGCTCTGC CCTCTCACC GTGAGCTCCT GGCTGCCCA
38581 GGCCATGGAC ACTGGCCATG CCCAATGAGA CGTTCTAACC CAGGACGCGT GGGGACCCTT
38641 ATTTCCAAAG GAAAAACATA CATTTCACCT CGTCGAGGAG CAAAGTGAGC CCCCCATCCC
38701 CACCCCATCC CCTTCTCAA CACTGCCGGA GTCCCATTTG GCCATGCCCC CTCCCCCACC
38761 GTGCCTCCTT GTCCCCCTCA ATTAACCTCT CCTGGGCCC TAGGGACACC CAGGGGCCCT
38821 GTACCCAGCC TCCCCTGAC ACCCCCCACC GGGGAGGTGG GAGGCTGCCT CCACTCCACT
38881 GCCCAGCCCT GCCACGGCTC TCCCTCAGAC ATTCCAACCT CCTATGAAAG GTCAGACGCC
38941 GAGGACCTGC CCTCCCTCTT TTCTACGGAG AACGTTGCCT TATACTCACT TCAGATGATG
39001 ACACTGTGGA CGTGTGTGCT TTAACCACAT TTTATTTAAT TGTCCCTTC CTCTCTTGC
39061 TAGACGCCCT CCCCTTGCTT GCATTCACTT TAGGGTTTTG GGGGCAATGC TGAGCATAGG
39121 AGGGTCTTTT TTTTTTGTCT ATTTTTTTTC TCACTTTAAC AATCTTTTC TAAATGACAC
39181 ATTTCAACTC AAATCTGGAC TCACCTGAGA ACACCTTCCA TCCTGACCCC GCCACGCCCT
39241 CATCTCCCC TGCCCAAGCT CCACTGTGTA CAGTGATAT TGTATAATAG ACTATTGTGT
39301 CTAAGACATG TTTTAAAAAC ATATTGCTTG TTATTTTTGA GGCTTTTAAA TTAAACAGAA
39361 ATCCAACCTT ATTTTCGTTG TAACCTGCTG AGGATGTTTT ATTAATTAAG TGAAGATTG
39421 TTATCCTTTA TTAACGATGT ATTTTGTTGG TCAGTACTGG GCTGACAAA AGTTTCTTGC
39481 TAATAAATTT AGTTGCCCAA GGCAAAGTCT GCGTTTGCA GTTGGGCTC CTTGTCTGT
39541 GATGTCAGAG GTCTTACAGA AGGATTCTTG AGGCGGGAAC TGTGGACTCA ATTTACCCTC
39601 CTGGTCTTGG TCCCAAAGGA GACCGGCTCT CAGAGCACCT GGGGGCCTC GGAGCCAGAG
39661 CCCAGTGCAG TCTCTGCTTT TTTCTCCCAC TCCAGTCTT GCTTCAGCCA CAGAGGCGAA
39721 CCCTAGAAGA ACTTCTGAG CTGGCCGACT GTTGTACCTC CCACCCAGC CCCGAGGAAG
39781 GCACTGCCTC CCAGTCTGTG AGAAGCTAGA GGAGTCAAAA GCCTTAGCC ACAAGGTTGG
39841 GTTTGCAGAA AATACTGTTT CCACAGGGTG GGAGGGAAGT CTGGGACCAA GCAGGTTCTT
39901 CAGGGTCGGC TCTGGGTTTG AGTAGAGTTG CAGAGTGGAA ACTGTATTGG GGGGCACCTT
39961 TTCATTTTGG CAGAACTACA GAGAAAGAAG TCAGGGATTG TGACTTACTC AGATTTAAAG
40021 CCACAAGCGT TGGGGCCTG AAAAGCAGCA GCCAGGCTGG AGGAAGAGGC CAGTGTGCCT

Exon 8

Mouse Sp2 cDNA sequence; NCBI access. no. BC021759

1 ggcggcgag | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100
 101 gctgctacta ccaaggact ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc
 201 ctgctcccc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc ccaagctacc
 301 aggaataata ctccagatac aggcctccca gctgagatcc tctaccctcg gctgagatcc tctaccctcg gctgagatcc tctaccctcg gctgagatcc
 401 agccgataca atgcaagtat ccaatcacca gccatcacca cccctcaga tccaaagggaa ccaatcacca gccatcacca cccctcaga tccaaagggaa
 501 tcatcccaag caccacaaca gccatcacca cccctcaga tccaaagggaa ccaatcacca gccatcacca cccctcaga tccaaagggaa
 601 caccctctg cagagcgggg ccaactggtt cctcactgag agccccca tgcctgata gaccacggca ccaactggtt cctcactgag agccccca
 701 atgggggccc ctgctcagct ggctgagcag gtgagacggg tccagcaagt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt
 801 ccgtagctgt ggctgagcag gtgagacggg tccagcaagt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt
 901 cggggggccc cagctgtggt tccagcaagt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt
 1001 caagctgctc ctgcccacct cccacacggt tccagcaagt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt
 1101 cacctctctg tgaggtacag acagtctctg tccagcaagt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt
 1201 tccccatctg agtggcaca gcaaaaaa tccagctgct cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt cccacacggt
 1301 aacgcccga cagctgccc gctgacgct agctgctcag gccatgacga ccatcaacat cctgactatc ctggaacaca cctgactatc ccatcaacat
 1401 ggcagcaga gctgacgct agctgctcag gccatgacga ccatcaacat cctgactatc ctggaacaca cctgactatc ccatcaacat cctgactatc
 1501 cggagagccc cagctgggg agaaacggcg agaaacggcg agaaacggcg agaaacggcg agaaacggcg agaaacggcg agaaacggcg agaaacggcg
 1601 cactgtgccc acatcccga ctgcccggag ctgcccggag ctgcccggag ctgcccggag ctgcccggag ctgcccggag ctgcccggag ctgcccggag
 1701 actggttctt ctgtgaaag agattcactt ggaagatga actccagcga ggaagatga actccagcga ggaagatga actccagcga ggaagatga
 1801 gaagcgttc atgaggatg accactcac caagcttac aagaccaccc caagcttac aagaccaccc caagcttac aagaccaccc caagcttac
 1901 gaacacctcc catctgattg gccctgggtc catgatggac aggtggcccc catgatggac aggtggcccc catgatggac aggtggcccc catgatggac
 2001 gcagagaggg actctgttc tgttctctcc tccctccaaa ggtcttggccc tccctccaaa ggtcttggccc tccctccaaa ggtcttggccc tccctccaaa
 2101 caatgagacg tctaaacca tgcctgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc
 2201 cccatcccc tcccaaac tgcctgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc ccactgagtc
 2301 ggacaccag gggccctgta cccagctcc cactgacacc cccacacggg cccacacggg cccacacggg cccacacggg cccacacggg cccacacggg
 2401 tccaaacatt ccaacttct atgaaaggtc agacgcccag gactgcccct cccacttctc tacggagaac gttgcccctat cccacttctc tacggagaac
 2501 ctgtggagct gtgtgcttta accaatctt atgaaaggtc agacgcccag gactgcccct cccacttctc tacggagaac gttgcccctat cccacttctc
 2601 gcaatgctga gcataggagg gctctttttt ttttctat ttttctca ctttaacaa atgacacatt tcaactcaaa tctggactca tctggactca
 2701 cctgagaaca ccttccatcc tgacccccgc agcctctcat ctccccctgc ccaagctcca ctgtgtacag tgtatatgt ataatagact attgtgtcta
 2801 adacatgqtt taaaaacata ttqcttqta ttttttaqgc ttttaaatga acaaaaaa caactttaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

I= exon-intron transitions; putative start codons (in-frame ATGs) and stop codons are depicted in red

6.3 Acknowledgements

First of all, I would like to thank my supervisors in Marburg and Rotterdam, Prof. Dr. Guntram Suske, Dr. Sjaak Philipsen and Prof. Dr. Renate Renkawitz-Pohl for giving me the opportunity to do this work.

In parallel, I would like to thank all the people from Guntram's and Sjaak's group for their manifold help and support: Alexandra, Bastian, Bernhard, Daniel, Eva, Grigore, Harald, Hetty, Imme, Iris, Jun, Laura, Nynke, Rita, Roy, Tamar and Trientje.

I am specially grateful to Iris and Nynke, the best technicians, who helped and motivated me so much. Thanks!

Thanks also for all the collaborations: Bastian for the help during antibody generation, Grigore for cell immunostainings, Nynke for all the ES cell works and mouse genotyping (and all the work I didn't realize), Imme for Southern Blot analyses of the targeted ES cells, John Kong a Sang for blastocyst injections, Sjaak for organization and mouse works, Jun for the statistics.

Special thanks to Matthijs Uyl and Patrick Rodriguez for the time they spent to help me!

I am also grateful for all the support, motivation and distractions coming from outside the biological field: from my parents and my family, Mareen, Volker and family, Otto and Brigitte, "I Rimbombanti", the "Spranz-Quartett", the "Hildrizhausen-Ensemble", "Organetto" and Co., as well as from Jenni, Markus and all the people from the "Fourth Floor". Thanks to all of you! Also special thanks to everyone being close to me, although I could not meet you very often during the last three years!

Finally, I want to thank Nicki for her endless patience, encouragement and love! This is your work!

6.4 Erklärung

Ich versichere, daß ich meine Dissertation mit dem Titel "**Transcription factor Sp2: Molecular characterization and generation of Sp2 gene targeted mice**" selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 30.06.2005

(Frank Baur)