Development 136, 3451-3461 (2009) doi:10.1242/dev.027714

Smicl is required for phosphorylation of RNA polymerase II and affects 3′-end processing of RNA at the midblastula transition in *Xenopus*

Clara Collart[1], Joana M. Ramis[1], Thomas A. Down[2] and James C. Smith[1,3,*]

Smicl (Smad-interacting CPSF 30-like) is an unusual protein that interacts with transcription factors as well as with the cleavage and polyadenylation specificity factor (CPSF). Previous work has shown that Smicl is expressed maternally in the *Xenopus* embryo and is later required for transcription of *Chordin*. In this paper we search for additional targets of Smicl. We identify many genes whose onset of expression at the midblastula transition (MBT) requires Smicl and is correlated with the translocation of Smicl from cytoplasm to nucleus. At least one such gene, *Xiro1*, is regulated via 3′-end processing. In searching for a general mechanism by which Smicl might regulate gene expression at the MBT, we have discovered that it interacts with the tail of Rpb1, the largest subunit of RNA polymerase II. Our results show that Smicl is required for the phosphorylation of the Rpb1 tail at serine 2 of the repeated heptapeptide YSPTSPS. This site becomes hyperphosphorylated at the MBT, thus allowing the docking of proteins required for elongation of transcription and RNA processing. Our work links the onset of zygotic gene expression in the *Xenopus* embryo with the translocation of Smicl from cytoplasm to nucleus, the phosphorylation of Rpb1 and the 3′-end processing of newly transcribed mRNAs.

KEY WORDS: *Xenopus*, MBT, Smicl, Polyadenylation, RNA polymerase II phosphorylation, CCCH zinc finger

INTRODUCTION

Smicl is a protein that interacts with transcription factors and forms a complex with the Cleavage and polyadenylation specificity factor (CPSF) (Collart et al., 2005a). We have previously shown that Smicl regulates the expression of *Chordin* in the early *Xenopus* embryo, but *Chordin* is unlikely to be the only target of Smicl, because ectopic expression of *Chordin* cannot rescue completely the effects of an antisense morpholino oligonucleotide (MO) directed against Smicl (Collart et al., 2005b). In this paper we use microarray analysis to search for additional targets of Smicl, both in an effort to understand its role in early development as well as to obtain more insights into its mode of action. This analysis has identified many genes whose onset of expression at the midblastula transition (Newport and Kirschner, 1982a; Newport and Kirschner, 1982b; Schier, 2007) requires Smicl activity and whose activation is correlated with the translocation of a tagged form of Smicl from cytoplasm to nucleus. Analysis of one of these targets, *Xiro1*, shows that Smicl influences 3′-end processing of the *Xiro1* transcript.

The effects of Smicl on zygotic gene expression are first detected at the midblastula transition (MBT), which in *Xenopus* occurs after 12 cell divisions and involves many coordinated changes in cell behaviour. In particular, the cell cycle loses its synchrony and, with the introduction of G1 and G2 phases, becomes longer (Kimelman et al., 1987; Lemaitre et al., 1998). In addition, several proteins translocate from cytoplasm to nucleus (Dreyer, 1987), maternal mRNA is degraded (Bushati et al., 2008; Giraldez et al., 2006), the apoptosis programme is engaged (Peng et al., 2007), cells become motile (Kimelman et al., 1987) and karyomeres (individual chromosomes surrounded by a nuclear envelope) disappear (Lemaitre et al., 1998). Perhaps most significantly, the MBT marks the onset of bulk transcription in the embryo, driven by RNA polymerase II (RNA polII) (Newport and Kirschner, 1982a; Newport and Kirschner, 1982b).

RNA polII consists of two large subunits and several smaller subunits. The largest subunit, Rpb1, contains an unusual C-terminal domain (CTD) that contains repeats of the heptapeptide YSPTSPS, which can be phosphorylated on serines 2 and 5 (Hirose and Ohkuma, 2007). The CTD acts as a docking platform for factors involved in co-transcriptional processes, and the positions of phosphorylated serines determine which transcription factors and RNA processing factors are recruited (Egloff and Murphy, 2008). Importantly, before the MBT in *Xenopus*, serine 2 of the CTD heptapeptide is not phosphorylated, but it becomes hyperphosphorylated thereafter (Palancade et al., 2001). Serine 2 phosphorylation is important for overcoming an early elongation block in *Xenopus* (Peterlin and Price, 2006) as well as for splicing and polyadenylation (Hirose and Manley, 1998; Hirose et al., 1999).

CPSF is required for pausing and termination of transcription, and it does this by binding through its 30 kDa subunit (CPSF30) to the body of Rbp1 (Nag et al., 2007). Because Smicl affects 3′-end processing and has a zinc finger domain similar to that of CPSF30, we asked whether it could form a complex with Rpb1. Our work shows that Smicl interacts with Rpb1 and is required for phosphorylation of the Rpb1 CTD, at least between MBT and mid-gastrula stages. These observations link the onset of zygotic gene expression in the *Xenopus* embryo with the translocation of Smicl from cytoplasm to nucleus, the phosphorylation of Rpb1 and the 3′-end processing of newly transcribed mRNAs.

MATERIALS AND METHODS

Constructs and cloning

The open reading frame (ORF) of Venus was cloned 5′ of Smicl-HA in pCS2 (Collart et al., 2005b) using restriction enzymes *Cla*I and *Eco*RI, to generate a full-length Smicl construct with N-terminal Venus and C-terminal HA tags.

[1]Wellcome Trust/CR-UK Gurdon Institute and Department of Zoology, and [2]Wellcome Trust/CR-UK Gurdon Institute and Department of Genetics, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. [3]Medical Research Council National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

*Author for correspondence (jim.smith@nimr.mrc.ac.uk)

Accepted 13 August 2009

Table 1. Primers used in this study

Gene	Accession number *X. laevis*	Forward primer	Reverse primer
SRFP2	AF136184	AACAAGGTCTGCTCCTTCCA	ATGGTGTCTCCACCTCCTTG
cyclin D1	BC041525.1	CCAGACATTTGTTGCCCTCT	GTTGTGTTGCTGCTGTGCTT
SOX2	AF005476	GTCAAGTCGGAATCCAGCTC	TTCTGCCCCAGGTAGGTACA
Unknown	n	TTTCGTGGTGATCGGTAACA	TGCTCGGAGGAGACAAGTTT
SP5	AY062264	GTGTCCCAACTGCCAAACTT	AGCTTTTGCCACAGAACAGC
Cerberus	BC081277	GAATGGAGCCCCACAGAATA	TTGCTGATTTGGAACATGGA
GATA4	DQ096869	GCTTAAAACTCTCGCCACAGA	TGCTTTAAGCTAAGACCAGGTTG
TFAP2A	M59455	AGAGCTGCCTGACCCACTT	TTGTCCATTGCTTTGAGTGC
Cadherin 1	L29057	CCTGCCAATCCAGATGAAAT	GAGCCACTGCCTTCGTAATC
cyclin B1	AJ304991	TTGTCCATCAATTGGTGACTTT	GTGCACCATGTCATAGTCTGTC
Xiro1	AJ001834	AAACCTGAGTTCCCTCAGCA	TGTTTTCACGAACAGCTTGG
BTG4	BC088789	AAGCAGCTGAGCAAGCAGA	CGGAATGCTTGTCCTTTTGT
ROR2	AB087137	AACTGATGGGAAACTCCATGC	ACCCCATAGGACCAAATATCTGAA
PDGFRA	M80798	GAATGGCAAAACCTGACCAT	GCGAGTAACTGCAGGGTGAT
FBXO43	AF176353	AATCCATCCATTCAGCAAGC	TCTGCTGCAAGTCCCTCTTT
H2BFS	XLHISH3A	GATACCGGCATCTCTTCCAA	ATGGTGGAGCGCTTGTTGTA
RPN2	BC046727.1	ATCCGCTCCAATGTTGACTC	GTGAGCAAGGCTTCAATGGT
KRT24	BC043901.1	AGTTCTGCAGGTGGTTTTGG	GCAAGACGGTCATTGAGGTT
Unknown	BC097911.1	GCCCCAGAACCACTAAGTAAC	CCTGGACCACCATCTCTGAA
YWHAH	BC075238	AGGCCTTTGAAATCAGCAAA	CTTGTCCACAGGGTCAGGTT
WDR20	BC070814	CATGAATGCCACAAGTCCAC	GCAAACTTGCTGACACCAGA
CEPB1	BC077702	AAAATGTCCAGTCGCAGGAT	AGAGCCCCCACAAATACAGT
VLDLR	AB006906	CTTGCCAACCAGACCAATTT	CATCACTCCAGTCCTTGCAG
FOXI1	BC042303	AGTCCTCAGCGACCTTCAAA	GATATGGAGGAGGTGTGATGGAA
CSNK1E	AF183394	TGCACATCGAGAGCAAATTC	ACCAAGCCCCATTAGGAAGT
RBP1	BC068742	GATGATCGCAAATGCATGAC	AAACCTGTTTACAGACAACATCTCC
LOC494796	BC082922	TATGAAGCCGTGCATCAGAC	CTCAAGAAGACATCAAGGCAATC
PGK2	BC077781	TTCCATGTGGAGGAAGAAGG	GCTCTGTGTGCAGTTCCAAA
C15ORF20	BC097805	GCTGTGCGTGGTAAAGATGA	TCTGTACATCTGCCAAGTCG
SYTL2	BC082628	TCTCAAAATGATTGATCTCAGCTT	CACGTTCCTCCTCTGCTTTC
ECT2	AY487422	GCATCAAGGGCCATAAAGAA	CTTCGGCCATCTGCATTTT
MCM10	BC070548	CCCAAAACAAGGTGCAGAAT	TGCGTTTAATGCTGTTTGGA
KCTD14	BC068871	CAGCGAGAACATTGAGGTGA	TCATGGCACTTGCCTACATC
PI4K2B	BC077943	CAAGGCAGTGTGGCTTGTAA	GCTTTCTGGTATTTTCAGGTAAGG
FBXO33	BC099041	CCAGTTGCCAACACCATAGA	CTTGCCATTTCTGCTGTGAA
VRK1	BC054213	GTTCATGGAAGCAGTTAAAGAAC	TAGGACTGCCACTGGCTTTT
TNRC17	BC070738	AGCTGGAGTTCCCTGCCTTA	AGACTGGCTCCACAACTGGT
BTG1	BC090221	GTCCTACAGGAGGCCATGAA	TCAAAAGGGTCAACCCAAAG
RBM5	BC077408	CGCCTTCGTGGAGTTTTATC	CGAAAGTTGTACAGGCCACA
cyclin D1	BC106631	AAATAGTGGCCACCTGGATG	TCAGTGGGATCGTTTCCTTC
Unknown	X92851	TGGTTCAACAGGGATTCTCTA	TGTGGGGCACAAGGTATGT
DHRS3	BC073416	AACCATTTTGGTGAACAATGC	CAATGTGGCCATTCTGAAGTT
Claudin 4	BC099009	TTCATTGGCAACAACATCGT	GAAGGGCCAGCAGAGAGTC
ATP2A2	BC098958	AGATGTCAGTCTGCAGGATGTT	GACCATCATATTGGTGGCATT
TRIM29	BC044714	TCCTTCTGTGAAACCCATCT	TGTTTTGCTGTGGAGCTGAC
ELAVL3	U17598	ATTGAGTCCTGCAAGCTGGT	GGGTGTTTATTGCTTTATCTGC
SLC2A2	BC070704	GTCTGGCATCAGGACTGGTT	TCCCTGTGACAATAGCCAACT
Unknown	AJ009297	CACAACATCTTGGCAGTTGG	GGGGAATAGAGTTCGGCTTC
POLE3	AY271302	CTTGGAAGTTTACAGGCAGGA	CTTGTCCTGATCCTCAGAGTCG
HUNK	AY318878	ATGAGACGTCAAGATTCCCAC	ACTCTGTACTGGTGATTTGGG
DLX6	D10259	ATCCGGTTTAATGGGAAAGG	AGTGCCAGGTACTGGGTTTG
Epiplakin 1	BC082957	ATCAGGAAAGGCAGGTGATG	GCACTGCAATGACCTCTTCA
SLC2A2	BC070704	GTCTGGCATCAGGACTGGTT	TCCCTGTGACAATAGCCAACT
LMO3	BC106431	GCCGCTTGCAGTAAACTGAT	CTTCAGGAAAAACTTGTCCCC
MX11	DQ104100	GCCTCCACATCCTAATCCAA	GCGGCCCGATACAATAACTA
NGFR	BC068640	AATGCAAAGAAACCGAGGTG	TTTGGCCAACTGCTTCTTCt
Unknown	BC108804	CCATTCGCCTTTGTGGAAT	TTGGAAACTCAACACGCAAG
ALCAM	BC074313	ATGGAAGCAAGTGCAAGGAC	CGTTTCTCCATAAACTGCATTG
EFS	BC084637	ATGTGATGCTAGTTCTGGAGC	GCCTGATACTCGCTTTCCTG
GATA3	BC110754	CACCACCCGGCTCTATTAAA	ATTGAAGAGTACATCCATGTC
LRRN1	AB014462	AGAACCCTATTCGCAAGATCC	CGATCTACAGACACCAACTCTGC
PTAFR	BC046657	AGCCAGTGAAGTCGAGGAAA	GAGTCCAAGGGCCATCTACA
HTRA1	BC087471	GGGGATACGGATGATGTCAC	CTTCAAGCCAGCCTCTTCAG
IER5L	BC081197	CGGATATTTGCACCAGGACT	GAGCCGAAGATGGAGATCAG
PDCD4	BC056125	GGGCAGGTCTATGGAGATCA	ATTCTTGCACCATTGGCTTC
OSR2	BC108579	AGCTCACCAACTACTCCTTCC	TGGCATCCATTAAGCCTTG
SALL1	BC059284	CCAGCATAGGAACAACAGCA	AGATGGATCTGCAAGGCACT

Table continued on next page.

Table 1. Continued

Gene	Accession number *X. laevis*	Forward primer	Reverse primer
Unknown	BC110732	ATCAGCTACAGCCTGTCTCC	AGAGCACAAAAGGTGGCACTA
ZFHX1B	BC084972	TGGACCACTCCAGGAGCA	TAAGTCCAAGGGCTCAGCTT
Unknown	AJ009297	GCTGAGAGGAGGCAGCTAGA	GGGGAATAGAGTTCGGCTTC
RBM12	BC059291	TCTGAAGCGTAACCGAATG	TTAGTTGCTGCATCAAAGGAG
STAG2	AF255018	GAGAATTTGCCATGCTTACCA	CCCAAGACAATATGTCGTAGCTT
HNRPU	BC046700	AGAGCTTTGCACTTCTCTGG	GCAACACCATTGCAACATTT
Unknown	BC045031	TGGAAGCCTGGTACAGAGGA	GTTTCTGCCAAAGTGCCTTC
IGSF4	BC108832	GCTGGTTCAAGGGAAACAAA	CAGCTGGATGGTCAACAAGA
DYNLL1	BC073042	GGAAGGAATTTTGGCAGCTA	GAAGGTTTGTTCCCTTGGTT
NUDT22	BC068937	TACCTGCAAGGAGGTCCTGA	GGCTGATGGACACAGTTCCT
GATA6	BC082349	CCTTTCTGACTTTTGCACAGC	GGCAAAGTCTGTTGGATGGT
MEIS1	BC084920	CGATTCCACCTGTTGGAGTTA	ACCCGGTGTACCTGAGTGAG
SALL1	AF310007	GAACAGCACACCTGCAAGAA	CCAGCCATTGGAGTAATGC
CYP26A1	BC073518	GGAGAGACTCTGCAAATGGTG	GTGCAGGTTGGACAGACAGC
MGC61598	BC084778	GACCAACTGCGAGTTTAACG	GCCCCAAACTTTACCAACAG
ZFP36L1	BC100162	AGCTTGTTTGCTCCAAGCAT	CTGGAACTGCTCAGGTAGCC
ZNF503	BC046863	ACTGTTCTCCCCCTGGATCT	TTTTCTTGGCATCCAGCTCT
KIAA1324L	BC077391	CCAAGGGGACAAAATACTACCA	CCTTTGCTGTCTGATGGGATA
P4HB	BC077772	TGCCAAGATGGATTCTACAGC	CATCTCCCTCCTCCAAATCA

Sequences are shown 5′ to 3′.

Smicl-HA in pCS2 was also digested using restriction enzymes *Cla*I and *Sac*I to generate a similarly tagged Smicl construct containing just the last 338 amino acids of the ORF (which include the zinc fingers). FLAG-Rpb1 and FLAG-Rpb1ΔCTD constructs were as described (Rosonina and Blencowe, 2004).

Xenopus embryos and microinjection

Embryos of *Xenopus laevis* were obtained by artificial fertilisation. They were maintained in 10% normal amphibian medium (NAM) (Slack, 1984) and staged as described (Nieuwkoop and Faber, 1975). Embryos were injected at the one-cell stage with 80 ng Smicl antisense morpholino oligonucleotide (Smicl MO) or control morpholino (CoMO) (Collart et al., 2005b), C-terminally HA-tagged *Xt*Smicl in pCS2 (Collart et al., 2005b) was linearised with Asp718 and sense RNA was generated with SP6 RNA polymerase. RNA injections were performed in embryos at the one-cell stage. Use of a fluorescein-labelled Smicl MO confirmed that injected MOs distribute evenly within the embryo (data not shown).

Microarray analysis

Microarray analysis was carried out essentially as described, using a microarray that represents about one-third of the genes expressed during early development (Chalmers et al., 2005; Ramis et al., 2007). Microarrays were scanned using an Axon 4000B scanner and GenePix Pro software (Axon). The microarray results were imported into Acquity analysis software (Axon) and normalised using Lowess normalisation. Data files were then created as described (Ramis et al., 2007). The complete datasets were deposited in the NCBI Gene Expression Omnibus (GEO) data repository (http://www.ncbi.nlm.nih.gov/geo/), with Accession number GSE4952.

Real-time RT-PCR

Differential expression was validated by real-time RT-PCR using the LightCycler 480 (Roche). Reverse transcription was carried out using Transcriptor First Strand cDNA Synthesis Kit (Roche) followed by real-time PCR using the LightCycler 480 SYBR Green I Master kit (Roche) following the manufacturer's instructions. Primers specific for *Ornithine decarboxylase* (*ODC*) and *Chordin* were as described (Piepenburg et al., 2004). Primers to amplify *Xiro1* intron sequences are: forward, 5′-CAGCTAAGTTCAGCCCAAGG-3′; reverse, 5′-GCGTTTATCGGACAACGATT-3′. Primers to amplify *Chordin* intron sequences are: forward, 5′-CACTGTTGAAGCCAAGCAAA-3′; reverse, 5′-GAGGCTGCATTGCTCTTCTC-3′. Details of other primers are provided in Table 1. The normalised target concentrations (in arbitrary units) were calculated from the real-time PCR efficiencies and the crossing point of the target and of the reference gene *ODC*, as described (Pfaffl, 2001). Correlations between microarray and real-time RT-PCR results were performed using the Windows SPSS version 11.5 software (SPSS, Chicago, USA).

Polyadenylation assays

An aliquot of 4 μg of total RNA was ligated for 30 minutes at 37°C with 0.4 μg of a 3′ amino 5′ phosphorylated oligonucleotide P1 (5′-P-GGTCAC-CTTGATCTGAAGC-NH_2-3′) in a volume of 10 μl using T4 RNA ligase (New England Biolabs). The reaction was placed at 70°C for 15 minutes. The whole 10 μl ligation reaction was used in a 50 μl reverse transcription reaction using Superscript III (Invitrogen), according to the manufacturer's directions using 0.4 μg primer P2 (5′-GCTTCAGATCAAGGTGAC-CTTTTT-3′) as described (Graindorge et al., 2006). Of the resulting cDNA, 1 μl was used in a 50 μl PCR reaction. As a reverse primer, P2 was used in all reactions. As forward primers, we used P3, specific for our positive control *Eg3*: 5′-AAGTGACTATGCAATTTGAGCTAGAAGTAT-3′ (Graindorge et al., 2006), and designed primers specific for *MCM10* (P4: 5′-ACAGCATTGCAGAATCATGG-3′), *FBXO43* (P5: 5′-TCTTGCACCT-GATGTTTGTG-3′) and *Xiro1* (P6: 5′-CCGTGTTCCATTTCAGACCT-3′). The PCR reaction mixture contained 1X buffer (Invitrogen), 0.2 μM of each primer (forward and reverse), 200 μM dNTPs, 1U of Platinum Taq Polymerase (Invitrogen) and 1.5 mM $MgCl_2$, in the presence (P4) or absence (P3, P5, P6) of 10% DMSO. The amplification programme consisted of a preincubation step for denaturation of the template cDNA (95°C for 5 minutes), followed by either 40 (P3), 45 (P4) or 50 (P5, P6) cycles consisting of a denaturation step (95°C for 30 seconds), an annealing step (56°C for 30 seconds for P3, P4, P5 or 59°C for 30 seconds for P6) and an extension step (72°C for 30 seconds), and then one final extension step (72°C for 7 minutes). The amplified products were digested with *Pvu*II (P3), *Bst*UI (P6) or *Psi*I (P4 and P5) to verify the specificity of the PCR products. PCR reactions and digests were separated on a 12% non-denaturing polyacrylamide gel and stained with ethidium bromide.

Co-immunoprecipitation experiments

Transiently transfected HEK293T cells were solubilised in lysis buffer containing 1% NP40, 150 mM NaCl, 20 mM Tris pH 7.5, 2 mM EDTA, 50 mM NaF, 1 mM sodium pyrophosphate, supplemented with protease inhibitors (Roche Molecular Biochemicals). Cell lysates were cleared by centrifugation and precipitations were performed by overnight incubation at 4°C with anti FLAG M2 agarose affinity gel (Sigma). Unbound proteins were removed by washing four times with cold lysis buffer. Bound proteins were harvested by boiling in sample buffer, and they were resolved by SDS-polyacrylamide gel electrophoresis. Flag-tagged and HA-tagged proteins were visualised after western blotting using rat monoclonal anti-HA-peroxidase-

coupled high-affinity antibody (3F10) (Roche) and goat polyclonal anti-FLAG-peroxidase coupled antibody (Abcam) and the SuperSignal West Dura Extended Duration Substrate kit from Thermo Scientific (Pierce).

Whole-mount antibody staining

Embryos were fixed overnight in MEMFA (3.7% formaldehyde, 100 mM MOPS, 2 mM EGTA, 1 mM $MgSO_4$, pH 7.4) at 4°C, and the vitelline membrane was removed. After gradually dehydrating and rehydrating the embryos in methanol, they were washed in phosphate buffered saline (PBS) and bleached in 2 × SSC with 2% formamide and 1.5% H_2O_2. Embryos were washed 2 × 10 minutes in PBS and 2 × 30 minutes in PBS with 0.1% BSA and 0.2% Triton X-100 (PBSbt) at room temperature. They were blocked for 1 hour at room temperature in PBSbt with 10% serum (PBSbts) and incubated overnight at 4°C in a 1/250 dilution of the anti-HA-peroxidase-coupled rat monoclonal antibody (Roche) in PBSbts. Embryos were washed for 1 hour in PBSbts, 4 × 1 hour in PBSbt and 10 minutes in PBS and incubated in Tris-buffered saline (TBS) with 0.066% 3, 3′-Diaminobenzidinetetrahydrochloride (DAB) (SIGMA) for 10 minutes. H_2O_2 was added to a final concentration of 0.024% and staining was observed after 1 minute.

Anti-Rpb1 antibodies

To detect Rpb1 and phospho S2 Rpb1, we used mouse monoclonal antibody 8WG16 (ab817) and a rabbit polyclonal antibody (ab5095) from Abcam.

RESULTS

Isolation of novel Smicl targets

Smicl maintains expression of *Chordin* in the early gastrula of *Xenopus laevis* (Collart et al., 2005b), but it is unlikely that this gene is the only embryonic target of Smicl, because injection of *Chordin* mRNA cannot completely rescue the phenotype of embryos injected with a Smicl antisense MO (Collart et al., 2005b). Further insight into the role and mode of action of Smicl requires the identification of additional targets, and to this end we performed microarray analysis on RNA derived from control embryos at the early gastrula stage and from embryos injected with Smicl antisense MOs. The microarray slides used in our experiments were designed from transcript sequences derived from a large-scale *Xenopus tropicalis* EST project (Gilchrist et al., 2004). These *X. tropicalis* microarrays also work in *X. laevis* (Chalmers et al., 2005; Ramis et al., 2007).

Xenopus laevis embryos from three different spawnings were injected at the one-cell stage with either Smicl MO, which inhibits Smicl function, or the control CoMO (Collart et al., 2005b) and were cultured to the early gastrula stage before RNA isolation. Some embryos were cultured to later stages to confirm that their development was impaired as described previously (Collart et al., 2005b). RNA from each spawning was hybridised with dye-swapped technical replicates, making six microarray slides in total. Oligonucleotides were considered to be differentially expressed when: (1) they showed at least a twofold difference (sample versus control) in expression levels in four out of the six microarrays; and (2) they were significantly different (q=0). In embryos in which Smicl was downregulated, 95 oligonucleotides fulfilled these criteria (Table 2): 33 genes were upregulated and 62 were downregulated.

The *X. laevis* homologues of the *X. tropicalis* cDNAs recognised by the oligonucleotides were identified by BLAST searches (Table 2). Genes were manually classified according to the annotation of their human homologues (NCBI databases, http://www.ncbi.nih.gov/), because gene ontology annotation for *Xenopus* species is not yet available (Fig. 1A). Apart from the unknowns, the largest group is involved in regulation of transcription (20%).

Microarray results were validated by real-time RT-PCR using the same samples that were used for microarray experiments. Primers were designed for the *X. laevis* homologues of all differentially expressed transcripts, with the exception of nine cDNAs for which *X. laevis* homologues could not be identified (Table 2). Of the genes tested, 76% (66) were confirmed as being differentially expressed (Table 2). Bilateral correlation analysis of the results obtained by microarray hybridisation and those obtained by real-time RT-PCR showed a Pearson correlation of 0.909 ($P<0.0005$) (Fig. 1B).

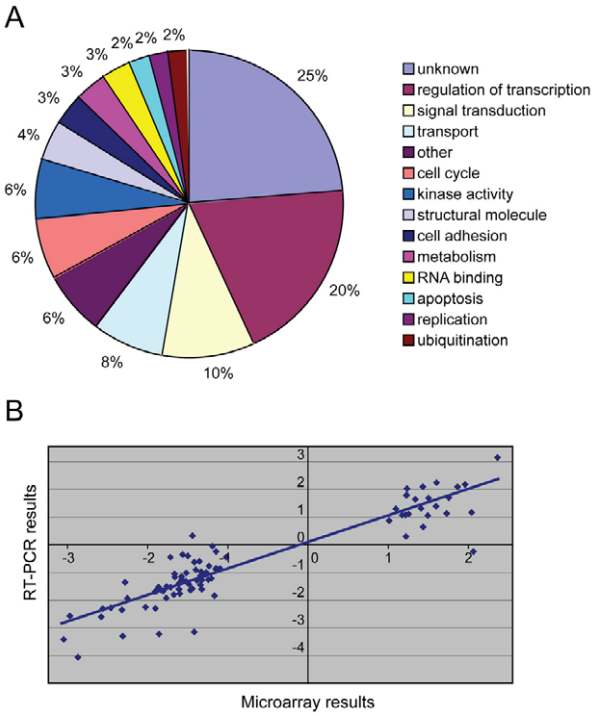

Fig. 1. Microarray experiment searching for targets of Smicl. (**A**) Classification of Smicl targets according to the annotation of their human homologues. (**B**) Correlation between microarray and PCR results. The *x*- and *y*-axes represent the $\log_2$ (sample/control) for the microarray and RT-PCR, respectively.

The temporal regulation of gene expression by Smicl starts at the MBT

Smicl is expressed both maternally and zygotically during *Xenopus* development (Collart et al., 2005b). Direct targets of Smicl are likely to be among the first to be up- or downregulated by loss of Smicl function. To find such targets, we studied the temporal expression patterns of 78 of the genes identified in our microarray analysis in control embryos and in embryos injected with a Smicl MO. We found that the genes could be divided into four categories, an example of each of which is shown in Fig. 2; for the complete results, see Fig. S1 in the supplementary material. In category 1 (41 genes), normal embryos exhibited low maternal mRNA levels that increased at the onset of zygotic transcription (between stages 7 and 8.5). Loss of Smicl function caused a reduction in zygotic transcripts (Fig. 2A,E). In category 2 (19 genes), maternal transcript levels normally decreased after the MBT, but the degradation of these gene products was delayed in Smicl-depleted embryos (Fig. 2B). Category 3 (four genes) comprise genes with expression levels in Smicl-depleted embryos that were first upregulated and later downregulated compared with control embryos (Fig. 2C). And category 4 (14 genes) consists of the genes with RNA expression levels in this experiment that proved to be only slightly affected by depletion of Smicl (Fig. 2D). We do not yet understand the transient

upregulation observed in the four category 3 genes, which include *Chordin*. One possibility is that loss of Smicl causes the downregulation of a repressor of *Chordin* expression.

Smicl translocates from the cytoplasm to the nucleus at MBT

The expression profiles of the genes in our categories 1 and 2 indicate that absence of Smicl affects the time of onset of zygotic gene expression and of the degradation of maternal mRNAs. These events usually occur at the MBT, and, interestingly, we observed that an HA-tagged form of Smicl translocated from the cytoplasm to the nucleus at precisely this time (Fig. 3A-F). Similar behaviour was observed in dissociated animal pole cells using a Venus-tagged form of Smicl and confocal microscopy: at the MBT the distribution of Smicl changed from diffuse and cortical to well defined and nuclear (Fig. 3G,H). A similar transition was observed in a construct consisting of just the Smicl zinc finger domain (Fig. 3I).

To ask whether the nuclear translocation of Smicl at the MBT requires new transcription, we injected embryos with α-amanitin as well as with RNA encoding HA-tagged Smicl. Such embryos divided normally to stage 9, but the transcription of *GS17* and *Siamois* was prevented, confirming that transcription was blocked (Fig. 3J). This inhibition of transcription had no effect on the nuclear translocation of Smicl after the MBT (Fig. 3K,L). This is discussed below.

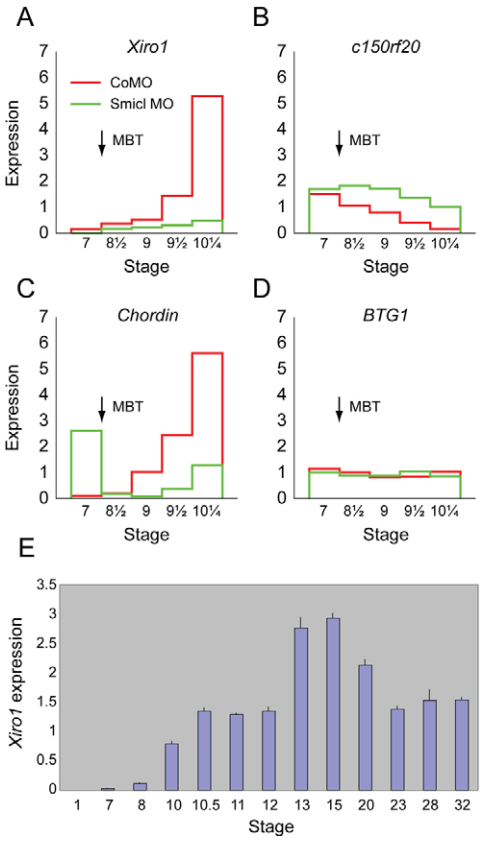

Fig. 2. Smicl targets fall into four categories. (**A-D**) *Xenopus* embryos were injected at the one-cell stage with 80 ng CoMO or Smicl MO. They were harvested at stages 7, 8.5, 9, 9.5 and 10.25, and the RNA expression levels of putative Smicl target genes were measured by quantitative RT-PCR. RNA expression levels relative to *Ornithine decarboxylase* (*ODC*) in CoMO embryos are indicated in red, and levels in Smicl-MO-injected embryos are indicated in green. We define four categories of Smicl targets and one example of each type of target is shown here. The others are presented in Fig. S1 in the supplementary material, with the above stages represented as numerals 1 to 5. (A) In normal embryos, RNA expression levels of category 1 genes are upregulated when zygotic transcription starts. Their transcription is downregulated in Smicl-depleted embryos. (B) RNA expression levels of category 2 genes diminish after MBT during normal development. Their decline is slower in embryos lacking Smicl. (C) In the course of normal development, category 3 genes are upregulated at the onset of zygotic transcription. Depletion of Smicl causes an early elevation of expression that is followed, in comparison to normal embryos, by a decrease in transcription. (D) RNA expression levels of genes in category 4 are similar in coMO- and Smicl-MO-injected embryos. The time of the MBT is indicated by an arrow in the four graphs. (**E**) The temporal RNA expression pattern of *Xiro1* assayed by quantitative RT-PCR and normalised with respect to *ODC*.

Smicl affects 3′-end processing of at least one of its target genes, *Xiro1*

Our previous work in tissue culture identified Smicl as a molecule that may be involved in 3′-end processing (Collart et al., 2005a). To address this point we studied three Smicl targets identified in our

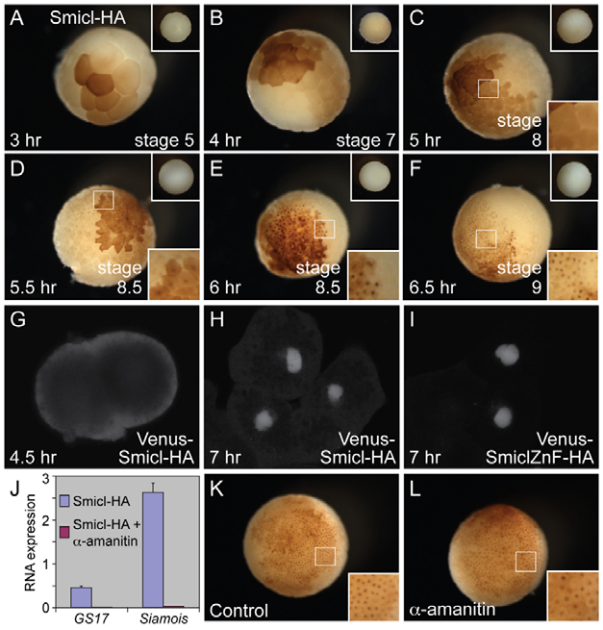

Fig. 3. Subcellular localisation of Smicl before and after the MBT. (**A-F**) *Xenopus* embryos were injected at the one-cell stage with 1.5 ng RNA encoding HA-tagged XtSmicl or with H_2O (inset). Embryos were harvested 3 (A), 4 (B), 5 (C), 5.5 (D), 6 (E) and 6.5 (F) hours after fertilisation, with the MBT occurring at about 5.5 hours after fertilisation at room temperature, at the onset of transcription of *GS17*. Developmental stages are indicated at the bottom right and higher-magnification insets show nuclear translocation between 5 and 5.5 hours after fertilisation. Overexpressed Smicl protein was detected by whole-mount antibody staining. (**G-I**) Confocal microscopy images of Venus-tagged full-length Smicl (G,H) and a version comprising just the zinc finger domain (I). Note that nuclear staining is not visible at 4.5 hours after fertilisation (G) but is detectable at 7 hours (H). The zinc finger domain is also nuclear at 7 hours (I). (**J-L**) Nuclear localisation of HA-tagged Smicl 6.5 hours after fertilisation in embryos injected with α-amanitin (50 μg/embryo). (J) Quantitative RT-PCR shows that this concentration of α-amanitin inhibits activation of *GS17* and *Siamois*. HA-tagged Smicl is nuclear in control embryos (K) and also in embryos injected with α-amanitin (L).

Table 2. Analysis of the genes identified by microarray experiments

Gene	Accession number		Log_2 (sample/control)			
	X. tropicalis	*X. laevis*	Microarray	RT-PCR	Array and RT-PCR	Category
ELAVL3	NM_001030326	U17598	–3.040	–3.435	+	1
SP5	CT025179	AY062264	–2.963	–2.582	+	1
KRT24	CX745012	BC043901	–2.862	–4.071	+	1
MX11	CR942348	DQ104100	–2.662	N.D.	–	N.D.
TRIM29	CR926196	BC044714	–2.572	–2.620	+	1
Unknown 43	CT485755	BC045031	–2.552	–2.321	+	2
SOX2	CR760314	AF005476	–2.453	–2.293	+	1
Epiplakin 1	No homology	BC082957	–2.315	–2.366	+	N.D.
KIAA1324L	CT025403	BC077391	–2.303	–3.308	+	1
Unknown	BX689202	AJ009297	–2.277	–1.371	+	N.D.
cyclin D1	BQ522031	BC041525	–2.247	–1.944	+	1
cyclin D1	BQ522031	BC106631	–2.021	–2.260	+	N.D.
SLC2A2	CT030568	BC070704	–1.899	–1.703	+	1
Cerberus	NM_203515	BC081277	–1.897	–2.308	+	1
DLX6	CX836573	D10259	–1.876	–1.677	+	1
DHRS3	CR848157	BC073416	–1.860	–1.534	+	3
HTRA1	BC073416	BC087471	–1.853	–3.231	+	1
NUDT22	CR942377	BC068937	–1.827	–1.657	+	1
TFAP2A	CR848274	M59455	–1.799	–1.689	+	1
Unknown 122	CX516192	AJ009297	–1.761	–1.936	+	1
Claudin 4	NM_001016663	BC099009	–1.759	–1.518	+	1
ATP2A2	DT418145	BC098958	–1.709	–0.466	–	1
SLC2A2	CT030568	BC070704	–1.706	–1.550	+	N.D.
GATA6	CT030595	BC082349	–1.667	–1.815	+	1
NGFR	BQ393897	BC068640	–1.657	–1.071	+	1
PTAFR	No homology	BC046657	–1.656	–1.179	+	1
MEIS1	CR760178	BC084920	–1.604	–1.418	+	1
DYNLL1	CR761866	BC073042	–1.603	–1.619	+	3
ROR2	CF219705	AB087137	–1.592	–1.781	+	1
SALL1	No homology	BC059284	–1.572	–1.167	+	1
PDCD4	CF151949	BC056125	–1.572	–1.264	+	4
IER5L	CX363316	BC081197	–1.559	–0.368	–	1
Unknown	NM_001016654	BC110732	–1.520	–1.332	+	N.D.
ZNF503	CF219041	BC046863	–1.512	–1.034	+	N.D.
EFS	CR761735	BC084637	–1.494	–0.416	–	4
SRFP2	CR761702	AF059570	–1.478	–1.446	+	3
PDGFRA	CR761598	M80798	–1.463	–1.656	+	1
LRRN1	CX374253	AB014462	–1.438	–1.597	+	1
IGSF4	DR833927	BC108832	–1.435	0.313	–	4
LMO3	CX874767	BC106431	–1.413	–3.157	+	1
ZFHX1B	DT447181	BC084972	–1.413	–0.640	–	1
CYP26A1	NM_001016147	BC073518	–1.394	–0.918	–	4
Cadherin 1	CF217799	L29057	–1.382	–1.296	+	1
Xiro1	DT434125	AJ001834	–1.370	–1.289	+	1
STAG2	BX775788	AF255018	–1.349	–0.609	–	1
SALL1	CX490174	AF310007	–1.336	–1.465	+	N.D.
GATA4	NM_001016949	DQ096869	–1.334	–1.136	+	1
RBM12	DN075293	BC059291	–1.325	–0.981	–	1
HNRPU	CR855462	BC046700	–1.319	–1.617	+	1
MGC61598	CX372678	BC084778	–1.316	–1.285	+	1
Unknown 115	No homology	X92851	–1.276	–1.093	+	1
POLE3	NM_001017264	AY271302	–1.264	–1.018	+	4
OSR2	DN038055	BC108579	–1.239	–1.278	+	1
ALCAM	AL797115	BC074313	–1.229	–0.776	–	1
HUNK	CX465369	AY318878	–1.198	–1.228	+	1
ZFP36L1	CT025463	BC100162	–1.188	–0.059	–	4
GATA3	CF221701	BC110754	–1.159	–1.850	+	1
P4HB	BX695354	BC077772	–1.144	–0.897	–	4
BTG1	BQ526679	BC090221	–1.142	–0.865	–	4
Unknown 132	CF241734	BC108804	–1.140	–0.260	–	1
RBM5	NM_001017278	BC077408	–1.090	–0.874	–	4
H2BFS	NM_001017251	X03018	–1.015	–0.465	–	1
FBXO43	CT010561	AF176353	2.366	3.130	+	2
YWHAH	CR848634	BC075238	2.066	–0.259	–	4
cyclin B1	DT407498	AJ304991	2.043	1.157	+	2
Unknown	BX772260	No homology	2.020	N.D.	–	N.D.

Table continued on next page.

Table 2. Continued

Gene	Accession number		Log_2 (sample/control)		Array and RT-PCR	Category
	X. tropicalis	*X. laevis*	Microarray	RT-PCR		
WDR20	CR848483	BC070814	1.963	2.165	+	2
CEPB1	NM_001017330	BC077702	1.869	2.074	+	2
VLDLR	CR760336	AB006906	1.759	1.684	+	2
FOXI1	BQ393881	BC042303	1.733	1.112	+	4
BTG4	CR761158	BC088789	1.609	2.228	+	2
CSNK1E	CF376118	AF183394	1.601	1.381	+	2
RBP1	CN076203	BC068742	1.508	1.669	+	2
Unknown	CR760615	No homology	1.507	N.D.	–	N.D.
LOC494796	CF591895	BC082922	1.499	1.037	+	4
Unknown	CX899957;CX899956	No homology	1.453	N.D.	–	N.D.
RPN2	CR848133	BC046727	1.440	0.626	–	4
Unknown	CR761187	BC097911	1.435	2.080	+	2
PGK2	NM_001016545	BC077781	1.415	1.299	+	2
Unknown	CX807382	No homology	1.407	N.D.	–	N.D.
C15ORF20	NM_001016603	BC097805	1.344	1.629	+	2
Unknown	CR760587	No homology	1.328	N.D.	–	N.D.
Unknown	No homology	No homology	1.307	N.D.	–	N.D.
Unknown	CR848499	No homology	1.306	N.D.	–	N.D.
SYTL2	CF781991	BC082628	1.258	1.098	+	2
ECT2	DN094916	AY487422	1.239	2.016	+	2
MCM10	CR848457	BC070548	1.235	1.779	+	2
KCTD14	CT025440	BC068871	1.231	1.772	+	2
Unknown	DN096278	No homology	1.228	N.D.	–	N.D.
Unknown	CT025176	No homology	1.226	0.287	–	N.D.
PI4K2B	NM_001016953	BC077943	1.220	1.055	+	2
Unknown	No homology	No homology	1.178	N.D.	–	N.D.
FBXO33	CT027881	BC099041	1.177	1.057	+	2
VRK1	CR761062	BC054213	1.100	1.283	+	2
TNRC17	DR841565	BC070738	1.017	0.857	–	4

N.D., not determined.

microarray experiments. We chose *Xiro1*, a category 1 gene that is among the earliest zygotically transcribed genes and is therefore likely to be a direct target (Fig. 2A,E), and two category 2 genes, *FBXO43* and *MCM10*, with regulation that is more likely to be indirect. To investigate lengths of poly (A) tails we used an RNA ligation assay (Graindorge et al., 2006) with minor modifications (Fig. 4A). The *Eg3* gene, with RNA that undergoes deadenylation between fertilisation and the 64-cell stage (Graindorge et al., 2006), was used as a positive control. Fig. 4B confirms that the assay could detect the deadenylation of *Eg3* mRNA during early cleavage stages, and Fig. 4C shows that overexpression of Smicl lengthened the *Xiro1* mRNA poly(A) tail whereas depletion of Smicl, by injection of an antisense MO, shortened it. To facilitate comparison of the PCR products produced in the polyadenylation assay, the assay was performed at the very onset of transcription of *Xiro1*, when transcript levels in the three different samples are still similar. This activation of *Xiro1* occurred slightly earlier than that of *GS17*, the onset of expression of which in our experiments defined the MBT. Smicl had no effect on the lengths of the poly(A) tails of *FBXO43* or *MCM10*. These experiments show that Smicl regulates 3′-end processing of *Xiro1* mRNA.

Smicl also affects levels of unprocessed *Xiro1* transcripts

To ask whether the early reduction in transcript levels caused by loss of Smicl is associated with this decrease in polyadenylation, perhaps by affecting the stability of the mature gene product, we studied transcript levels of *Xiro1* using primers specific for the mature mRNA and for the unprocessed mRNA. We have previously shown that levels of *Chordin* are reduced in Smicl morpholino-injected embryos and we therefore used *Chordin* as a positive control.

Our results show that transcript levels of *Xiro1* as well as *Chordin* were downregulated in embryos injected with Smicl MO and, significantly, that this is true for unprocessed as well as processed transcripts at the early gastrula stage (Fig. 5A). These observations suggest that downregulation also occurs at the level of transcription, and that the lower levels of *Xiro1* mRNA observed are not solely a consequence of RNA instability resulting from a shortening of its poly(A) tail.

The same assay was used to analyse levels of *Xiro1* and *Chordin* after overexpression of Smicl. Although Smicl causes upregulation of *Chordin* (Collart et al., 2005b), it had no effect on levels of *Xiro1* at the early gastrula stage (Fig. 5B). *Xiro1* proves to not be unusual in this respect: of all the Smicl target genes, shown in Fig. S1 in the supplementary material, only *Chordin* and *FoxI1* mRNA levels were upregulated in response to Smicl (data not shown).

Smicl does not bind to CBTF and NF-Y but does interact with Rpb1

We have identified 41 category 1 targets, all of which require Smicl for the onset of zygotic gene expression. To ask whether all 41 are regulated by a common mechanism, we first asked whether Smicl is brought to their promoters by a common transcription factor. To this end we analysed sequences 500 base pairs 5′ of the transcription start sites of the genes and searched for motifs from the JASPAR CORE database (http://jaspar.genereg.net/). We found the best match for each motif in each sequence using the nmscan tool from the NestedMICA suite (Down and Hubbard, 2005) and then compared the distribution of scores for each motif to that from 1000 randomly selected *Xenopus* genes. In this way we found that the NF-Y motif (core sequence CCAAT) was

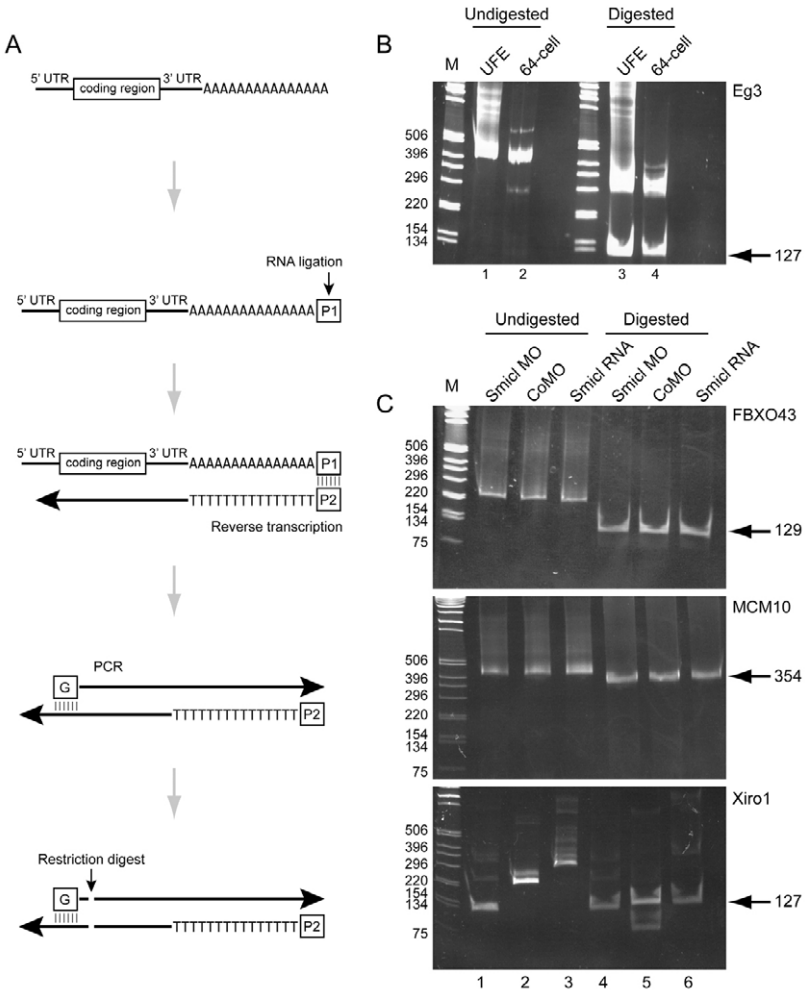

Fig. 4. Analysis of polyadenylation. (**A**) Schematic of the polyadenylation assay. Primer P1 is ligated to the poly(A) tail of mRNAs. cDNA is synthesised with primer P2, complementary to P1, and gene-specific products are obtained by PCR using a forward gene-specific primer (G) and reverse primer P2. The specificity of the products was tested by digestion with a restriction enzyme that cuts uniquely between the gene-specific primer and the beginning of the poly(A) tail (*Pvu*II for *Eg3*, *Psl*I for *MCM10* and *FBXO043*, and *Bst*UI for *Xiro1*). The products are analysed on a polyacrylamide gel. The PCR products from the polyadenylation assay, which contain a piece of the 3′-end untranslated region as well as the poly(A) tail itself, were split in half. On each gel (see B,C), the undigested halves, which are loaded on the left-hand lanes, reveal differences in the lengths of the poly(A) tails. The digested samples, which are loaded on the right-hand side, show fragments of DNA between the annealing site of the gene-specific region and the restriction site indicated above, together with other fragments, which include the poly(A) tail of unknown length. The lengths of the former can be calculated. They confirm the specificity of the PCR products and are indicated on the figure. (**B**) Control experiment to validate the polyadenylation assay. Lanes 1 and 2: products of an *Eg3* polyadenylation assay performed on RNA derived from unfertilised eggs (UFE) and from 64-cell stage embryos. Lanes 3 and 4: restriction digest of the polyadenylation products. M, DNA size marker. (**C**) Lengths of the poly(A) tails of three Smicl targets. *Xenopus* embryos were injected at the one-cell stage with 80 ng of Smicl MO, 80 ng of CoMO or 1.5 ng of Smicl RNA, and harvested at stage 7. The polyadenylation assay described in A was performed on RNA derived from these embryos for the genes *MCM10*, *FBXO43* and *Xiro1*. Lanes 1 to 3 show products of the polyadenylation assays. Lanes 4 to 6 show the restriction digests of these products.

significantly over-represented (P=0.007: an empirical figure based on re-sampling of scores from the random genes). This sequence is the binding site for the transcription factors NF-Y and CBTF in early frog embryos. NF-Y consists of three subunits and is the predominant Y-box binding protein in *Xenopus* oocyte nuclei (Li et al., 1998). CBTF is the main CCAAT binding factor at midblastula stages (Ovsenek et al., 1991). It is expressed maternally, and its activity is regulated by its p122 subunit ($CBTF^{p122}$). This protein is perinuclear during early embryogenesis, but moves from cytoplasm to nucleus at stage 9, before the detection of CBTF activity in the nucleus (Orford et al., 1998). These observations suggest that Smicl might regulate gene expression at the MBT by interacting with NF-Y or $CBTF^{p122}$, but our co-immunoprecipitation experiments have revealed no such interaction under our experimental conditions (data not shown).

In a further attempt to identify a general mechanism by which Smicl might affect gene expression at the MBT, we noted that CPSF was required for pausing and termination of transcription, and that it did this by binding, through its 30 kDa subunit, to the body of RNA polymerase II (Nag et al., 2007). Because Smicl affects transcription as well as 3′-end processing, and because it is similar to the 30K subunit of CPSF, we asked whether it could form a complex with Rpb1, the largest subunit of RNA polymerase II. Co-immunoprecipitation experiments in HEK293T cells indicate that HA-tagged Smicl interacted with full-length Flag-tagged Rpb1 (Fig. 6A). In contrast to the 30 kDa subunit of CPSF (Nag et al., 2007), Smicl did not interact with a tagged Rpb1 construct that lacked the C-terminal domain (Rpb1ΔCTD).

Smicl affects the phosphorylation of Rpb1 after the MBT

The CTD of Rpb1, which is required for interaction with Smicl, is a docking platform for factors involved in various co-transcriptional events (Egloff and Murphy, 2008). Different proteins are recruited to the Rpb1 CTD at different stages of the transcription cycle in response to changes in the phosphorylation state of this domain (Egloff and Murphy, 2008). Before the MBT in *Xenopus*, two forms of Rpb1 are present: one is hypophosphorylated in the CTD and the other is phosphorylated, but not at serine 2 of the repeated heptapeptide YSPTSPS (Palancade et al., 2001). Phosphorylation of serine 2 usually occurs at the MBT, and this hyperphosphorylated form of Rpb1 can be detected on a western blot as a band that migrates slightly more slowly than the embryonic phosphorylated form, as well as by a specific antibody (Palancade et al., 2001).

Phosphorylation of serine 2 at the CTD is required for elongation of transcription and for recruitment of factors involved in cleavage and polyadenylation (Egloff and Murphy, 2008). Smicl is also involved in cleavage and polyadenylation, and it interacts with Rpb1 in a manner that requires the CTD. We therefore asked whether Smicl is required for the phosphorylation of the CTD of Rpb1.

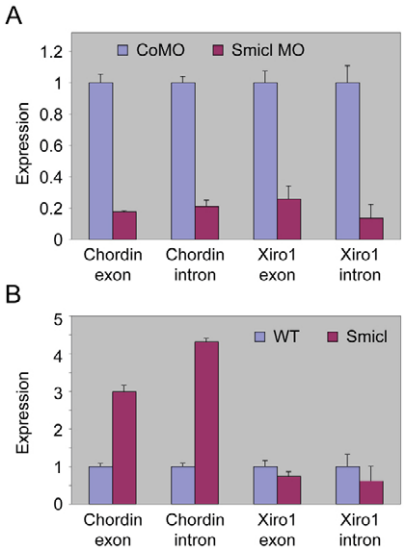

Fig. 5. Regulation of *Xiro1* and *Chordin* by Smicl. (**A**) Quantitative RT-PCR using primers specific for exon or intron sequences in *Xiro1* and *Chordin* show that downregulation of Smicl causes loss of immature transcripts as well as mature mRNA. Embryos were injected at the one-cell stage with 80 ng CoMO or Smicl MO and then cultured to stage 10.5. RNA was prepared from these embryos, and cDNA was synthesised by reverse transcription using a mixture of oligo(dT) primers and random hexamers. Intron and exon sequences from the Smicl targets *Xiro1* and *Chordin* were amplified by quantitative RT-PCR as indicated. (**B**) Transcriptional regulation of *Chordin* and *Xiro1* by Smicl differs. Embryos were injected at the one-cell stage with H_2O or RNA encoding HA-tagged Smicl. The assay described in A was repeated. Note that Smicl upregulates expression of *Chordin* but not of *Xiro1*.

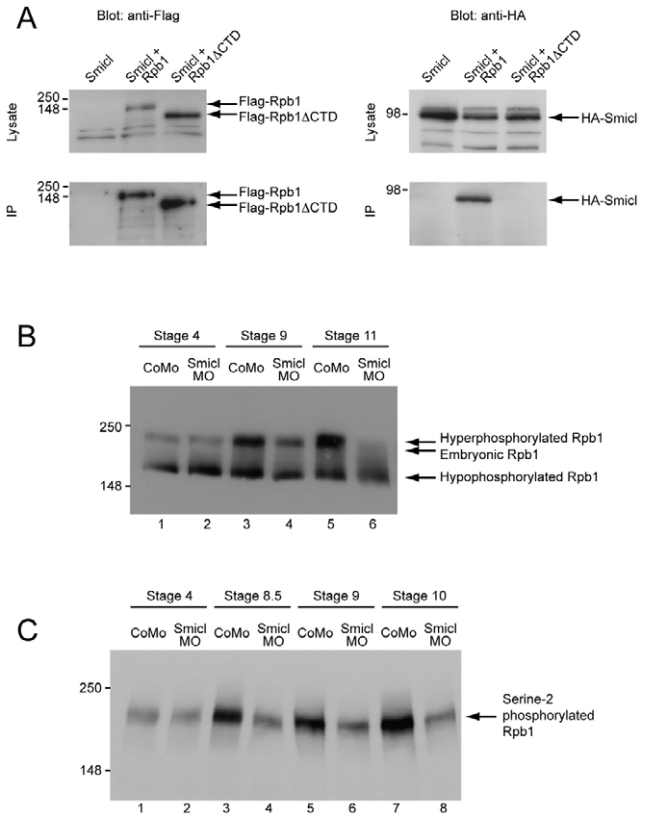

Fig. 6. Smicl interacts with Rpb1 and is required for its hyperphosphorylation at the MBT. (**A**) Smicl interacts with the largest subunit of RNA polymerase II (Rpb1), but not with Rpb1 lacking the C-terminal domain (Rpb1ΔCTD). HA-tagged Smicl and Flag-tagged Rpb1 and Rpb1ΔCTD were overexpressed in HEK293T cells as indicated. Samples were immunoprecipitated with anti-Flag coupled beads and the presence of Flag-tagged Rpb1 constructs and of HA-Smicl was determined by western blotting. Note that Smicl interacts with Rpb1 but not with Rpb1ΔCTD (lower right panel). (**B**) Smicl is required for the hyperphosphorylation of Rpb1 that occurs at the MBT. *Xenopus* embryos at the one-cell stage were injected with 80 ng CoMO or Smicl MO. They were harvested at stages 4, 9 and 11. Proteins derived from whole embryo lysates were separated by SDS PAGE and phosphorylation of Rpb1 was analysed by western blotting using an antibody that recognises all forms of Rpb1. Lanes 1 and 2 show the presence of the hypophosphorylated form and the embryonic form in embryos at stage 4 in CoMO- (lane 1) and Smicl-MO- (lane 2) injected embryos. Levels of the hyperphosphorylated form increase after MBT in CoMO-injected embryos at stages 9 and 11 (lanes 3 and 5) but not in Smicl-MO-injected embryos (lanes 4 and 6). (**C**) A similar assay to that described in B, but embryos were harvested at stages 4, 8.5, 9 and 10, and the western blot was probed with an antibody that recognises Rpb1 that is phosphorylated on Ser2 of the C-terminal heptapeptide. Lanes 3, 5 and 7 show increased levels of this phosphorylated form of Rpb1 in CoMO-injected embryos at stages 8.5, 9 and 10. This increase does not occur in Smicl-MO-injected embryos (lanes 4, 6 and 8).

Injection of our control antisense MO proved to have no effect on the phosphorylation of the Rpb1 CTD, but injection of our Smicl MO inhibited serine 2 phosphorylation at the MBT (Fig. 6B,C). This and our other results are discussed below.

DISCUSSION

Smicl (Smad-interacting CPSF30-like) is an unusual protein that not only interacts with transcription factors (Collart et al., 2005a; Collart et al., 2005b) but also with the CPSF complex. In addition, it has a zinc finger domain that is similar to that of CPSF30 (Collart et al., 2005a). These observations suggest that Smicl may control gene expression by modulating 3′-end processing. In an attempt to test this idea we used microarray analysis to identify novel Smicl targets and have shown that at least one of these, the homeobox-containing gene *Xiro1*, is indeed regulated by 3′-end processing. In the course of searching for a general mechanism by which Smicl might function, we discovered that it can form a complex with Rpb1, the largest subunit of RNA polymerase II, and that it is required for serine 2 phosphorylation of the Rpb1 CTD. This modification is required for the elongation of transcription, for splicing and for polyadenylation (Egloff and Murphy, 2008); in the *Xenopus* embryo it normally occurs at the MBT, and it coincides with the translocation of Smicl from cytoplasm to nucleus. We discuss these observations below, and speculate on the mechanism by which Smicl affects the Rpb1 CTD.

Smicl affects gene expression at the MBT

Our previous work identified *Chordin* as a target of Smicl, but it seemed unlikely that this gene was the only Smicl target because overexpression of *Chordin* does not completely rescue the loss of Smicl function (Collart et al., 2005b). We therefore used microarray analysis to compare gene expression in wild type and in Smicl-depleted embryos in an effort to identify additional Smicl targets (Fig. 1). Quantitative RT-PCR analysis at different stages of development revealed that these putative Smicl targets fell into four categories (Fig. 2A-D). Most fell within category 1, in which Smicl prevented or reduced the onset of gene expression at the MBT, or

within category 2, in which loss of Smicl caused a delay in the degradation of maternal transcripts at the MBT. Interestingly, the four criteria we define can also be used to classify zebrafish genes that respond to depletion of the TATA-binding protein (TBP) (Ferg et al., 2007), and indeed 27% of our Smicl targets are present in the list of 1927 genes that are regulated by TBP. This observation is consistent with the fact that transcription and 3′-end processing are coordinated processes.

We suspect that our list of Smicl targets is not complete, and indeed *Xlim5*, classified as a target in the course of additional experiments (not shown), was not identified in our microarray experiment. This failure to identify all Smicl targets might have occurred because our microarray does not represent all genes expressed during early development, because our selection criteria were too stringent, or because our Smicl MO does not completely inhibit phosphorylation of Rpb1, and different genes may show different sensitivities to levels of phosphorylation.

Consistent with the observation that Smicl regulates gene expression at the MBT, we note that Smicl becomes concentrated in the nuclei of blastomeres at this stage (Fig. 3A-I), although we cannot exclude the possibility that there is some protein present in the nucleus before the MBT. We do not know how this nuclear accumulation occurs, but we have demonstrated that it does not require new transcription, and can therefore be uncoupled from this manifestation of the MBT (Fig. 3J-L).

Smicl affects 3′-end processing of *Xiro1*

To investigate the ability of Smicl to regulate 3′-end processing, we turned to *Xiro1*, one of the earliest transcribed category 1 Smicl targets and therefore one more likely to be a direct target of Smicl. In our first experiments, use of a modified polyadenylation assay (Fig. 4A,B) indeed showed that depletion of Smicl shortens, and overexpression of Smicl lengthens, the poly(A) tail of zygotically expressed *Xiro1* (Fig. 4C). These results suggest that Smicl might stabilise *Xiro1* transcripts around the MBT.

We also investigated two category 2 genes, *FBXO43* and *MCM10*, in which loss of Smicl activity causes a delay in the degradation of maternal transcripts at the MBT. For the mRNA products of these genes we saw no effect of gain or loss of Smicl on the lengths of their poly(A) tails. The delay in degradation of their maternal transcripts may therefore be indirect, resulting from a failure to activate one or more category 1 genes: zygotic transcription is required for degradation of maternal mRNA (Bushati et al., 2008; Giraldez et al., 2006).

Smicl affects expression levels of *Xiro1*

Consistent with the idea that Smicl regulates *Xiro1*, we note that the expression patterns of the two genes overlap, both being activated most strongly in dorsal tissues and later in the neural plate (Collart et al., 2005b; Gâomez-Skarmeta et al., 1998). As discussed above, this regulation is likely to occur through 3′-end processing of *Xiro1* mRNA (Fig. 4), although our data also show that our Smicl MO causes the loss of unprocessed *Xiro1* transcripts as well as mature *Xiro1* mRNA (Fig. 5A). This, together with data showing that Smicl affects serine-2 phosphorylation of the CTD of Rpb1 (Fig. 6), which is required for overcoming an early block in transcriptional elongation (Peterlin and Price, 2006), suggests that Smicl may indirectly affect transcription of *Xiro1*. However, it is possible that Smicl is required to prevent degradation of the unprocessed *Xiro1* transcript, and we also note that overexpression of Smicl does not elevate expression of *Xiro1* (Fig. 5B). Indeed, of all the genes with expression controlled by Smicl, only *Chordin* and *FoxI1* are upregulated in response to its overexpression (data not shown). We continue to investigate the role of Smicl in transcription.

Smicl interacts with, and is required for phosphorylation of, the CTD of Rpb1

In searching for a general mechanism by which Smicl regulates all its category 1 and 3 target genes, we first reasoned that Smicl might be brought to the promoter of its targets by a specific transcription factor, influencing 3′-end processing by substituting for the 30 kDa subunit of CPSF. It proved that the binding sites of the transcription factors CBTF and NF-Y are over-represented in the promoters of category 1 Smicl targets, but Smicl did not co-immunoprecipitate with either protein (data not shown). However, we then asked whether Smicl interacts with RNA polymerase II itself and were able to show that it forms a complex with Rpb1, the largest subunit of RNA polII (Fig. 6A). In contrast to CPSF30, which interacts with the body of Rpb1 (Nag et al., 2007), it is the tail of Rbp1 that is required for complex formation with Smicl (Fig. 6A).

Smicl is also required for phosphorylation of the Rpb1 CTD at serine 2 of the repeated heptapeptide YSPTSPS, a modification that normally occurs at the midblastula transition in *Xenopus* (Fig. 6B,C). By regulating phosphorylation of the Rpb1 CTD, Smicl changes the docking platform for proteins involved in the elongation of transcription and RNA processing, and it therefore influences, albeit indirectly, 3′-end formation at the MBT. We do not yet know how Smicl causes the phosphorylation state of Rpb1 to change, although we note that the activity of RNA polymerase II can be regulated by small non-coding RNAs (Barrandon et al., 2008). Smicl binds and degrades small RNAs with a stem loop secondary structure (Collart et al., 2005a), and it is possible this activity is related to phosphorylation of the CTD and the modulation of RNA polymerase activity.

Acknowledgements

We thank our colleagues for helpful discussions through the course of this work, and especially James Smith, Martin Roth, Mike Gilchrist and Rick Livesey for their advice concerning microarray construction and analyses. We also thank Mike Chesney for his comments on the manuscript. This work is supported by the Wellcome Trust and the EU Network of Excellence 'Cells into Organs'. Deposited in PMC for release after 6 months.

Supplementary material

Supplementary material for this article is available at http://dev.biologists.org/cgi/content/full/136/20/3451/DC1

References

Barrandon, C., Spiluttini, B. and Bensaude, O. (2008). Non-coding RNAs regulating the transcriptional machinery. *Biol. Cell* **100**, 83-95.

Bushati, N., Stark, A., Brennecke, J. and Cohen, S. M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. *Curr. Biol.* **18**, 501-506.

Chalmers, A. D., Goldstone, K., Smith, J. C., Gilchrist, M., Amaya, E. and Papalopulu, N. (2005). A Xenopus tropicalis oligonucleotide microarray works across species using RNA from Xenopus laevis. *Mech. Dev.* **122**, 355-363.

Collart, C., Remacle, J. E., Barabino, S., van Grunsven, L. A., Nelles, L., Schellens, A., Van de Putte, T., Pype, S., Huylebroeck, D. and Verschueren, K. (2005a). Smicl is a novel Smad interacting protein and cleavage and polyadenylation specificity factor associated protein. *Genes Cells* **10**, 897-906.

Collart, C., Verschueren, K., Rana, A., Smith, J. C. and Huylebroeck, D. (2005b). The novel Smad-interacting protein Smicl regulates Chordin expression in the Xenopus embryo. *Development* **132**, 4575-4586.

Down, T. A. and Hubbard, T. J. (2005). NestedMICA: sensitive inference of over-represented motifs in nucleic acid sequence. *Nucleic Acids Res.* **33**, 1445-1453.

Dreyer, C. (1987). Differential accumulation of oocyte nuclear proteins by embryonic nuclei of Xenopus. *Development* **101**, 829-846.

Egloff, S. and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. *Trends Genet.* **24**, 280-288.

Ferg, M., Sanges, R., Gehrig, J., Kiss, J., Bauer, M., Lovas, A., Szabo, M., Yang, L., Straehle, U., Pankratz, M. J. et al. (2007). The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. *EMBO J.* **26**, 3945-3956.
Gâomez-Skarmeta, J. L., Glavic, A., de la Calle-Mustienes, E., Modolell, J. and Mayor, R. (1998). Xiro, a Xenopus homolog of the Drosophila Iroquois complex genes, controls development at the neural plate. *EMBO J.* **17**, 181-190.
Gilchrist, M. J., Zorn, A. M., Voigt, J., Smith, J. C., Papalopulu, N. and Amaya, E. (2004). Defining a large set of full-length clones from a Xenopus tropicalis EST project. *Dev. Biol.* **271**, 498-516.
Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J. and Schier, A. F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75-79.
Graindorge, A., Thuret, R., Pollet, N., Osborne, H. B. and Audic, Y. (2006). Identification of post-transcriptionally regulated Xenopus tropicalis maternal mRNAs by microarray. *Nucleic Acids Res.* **34**, 986-995.
Hirose, Y. and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**, 93-96.
Hirose, Y. and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J. Biochem.* **141**, 601-608.
Hirose, Y., Tacke, R. and Manley, J. L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* **13**, 1234-1239.
Kimelman, D., Kirschner, M. and Scherson, T. (1987). The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle. *Cell* **48**, 399-407.
Lemaitre, J. M., Geraud, G. and Mechali, M. (1998). Dynamics of the genome during early Xenopus laevis development: karyomeres as independent units of replication. *J. Cell Biol.* **142**, 1159-1166.
Li, Q., Herrler, M., Landsberger, N., Kaludov, N., Ogryzko, V. V., Nakatani, Y. and Wolffe, A. P. (1998). Xenopus NF-Y pre-sets chromatin to potentiate p300 and acetylation-responsive transcription from the Xenopus hsp70 promoter in vivo. *EMBO J.* **17**, 6300-6315.
Nag, A., Narsinh, K. and Martinson, H. G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat. Struct. Mol. Biol.* **14**, 662-669.
Newport, J. and Kirschner, M. (1982a). A major developmental transition in early Xenopus embryos: I. Characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675-686.
Newport, J. and Kirschner, M. (1982b). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell* **30**, 687-696.
Nieuwkoop, P. D. and Faber, J. (1975). *Normal Table of Xenopus laevis (Daudin)*. Amsterdam: North Holland.
Orford, R. L., Robinson, C., Haydon, J. M., Patient, R. K. and Guille, M. J. (1998). The maternal CCAAT box transcription factor which controls GATA-2 expression is novel and developmentally regulated and contains a double-stranded-RNA-binding subunit. *Mol. Cell. Biol.* **18**, 5557-5566.
Ovsenek, N., Karn, H. A. and Heikkila, J. J. (1991). Analysis of CCAAT box transcription factor binding activity during early Xenopus laevis embryogenesis. *Dev. Biol.* **145**, 323-327.
Palancade, B., Bellier, S., Almouzni, G. and Bensaude, O. (2001). Incomplete RNA polymerase II phosphorylation in Xenopus laevis early embryos. *J. Cell Sci.* **114**, 2483-2489.
Peng, A., Lewellyn, A. L. and Maller, J. L. (2007). Undamaged DNA transmits and enhances DNA damage checkpoint signals in early embryos. *Mol. Cell. Biol.* **27**, 6852-6862.
Peterlin, B. M. and Price, D. H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297-305.
Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45.
Piepenburg, O., Grimmer, D., Williams, P. H. and Smith, J. C. (2004). Activin redux: specification of mesodermal pattern in Xenopus by graded concentrations of endogenous activin B. *Development* **131**, 4977-4986.
Ramis, J. M., Collart, C. and Smith, J. C. (2007). Xnrs and activin regulate distinct genes during Xenopus development: activin regulates cell division. *PLoS ONE* **2**, e213.
Rosonina, E. and Blencowe, B. J. (2004). Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3′-end cleavage. *RNA* **10**, 581-589.
Schier, A. F. (2007). The maternal-zygotic transition: death and birth of RNAs. *Science* **316**, 406-407.
Slack, J. M. (1984). Regional biosynthetic markers in the early amphibian embryo. *J. Embryol. Exp. Morphol.* **80**, 289-319.