



Title	Silk Gland Factor-2, Involved in Fibroin Gene Transcription, Consists of LIM Homeodomain, LIM-interacting, and Single-stranded DNA-binding Proteins
Author(s)	Ohno, Kaoru; Sawada, Jun-ichi; Takiya, Shigeharu et al.
Citation	Journal of biological chemistry, 288(44), 31581-31591 https://doi.org/10.1074/jbc.M113.514471
Issue Date	2013-11-01
Doc URL	https://hdl.handle.net/2115/54948
Rights	This research was originally published in Journal of Biological Chemistry. Shigeharu Takiya(s). Silk Gland Factor-2, Involved in Fibroin Gene Transcription, Consists of LIM Homeodomain, LIM-interacting, and Single-stranded DNA-binding Proteins. Journal of Biological Chemistry. Year;2013 Vol:288(44)pp31518-pp31519. Copyright ©2013 the American Society for Biochemistry and Molecular Biology.
Type	journal article
File Information	takiya.pdf



Silk Gland Factor-2—(SGF-2) Involved in Fibroin Gene Transcription Consists of LIM_-homeodomain, LIM_Domain binding-interacting, and Single-Stranded DNA_-Binding Proteins

Kaoru Ohno¹, Jun-ichi Sawada^{2,5}, Shigeharu Takiya^{1,3,8}, Mai Kimoto³, Akiko Matsumoto², Takuya Tsubota⁴, Keiro Uchino⁴, Chi-chung Hui^{1,6}, Hideki Sezutsu⁴, Hiroshi Handa² and Yoshiaki Suzuki^{1,7}

¹National Institute for Basic Biology, 38 Nishigonaka, Myodaiji-cho, Okazaki 444-8585, Japan

²Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan

³Hokkaido University, North 10 West 8, Kita-ku, Sapporo 060-0810, Japan

⁴National Institute of Agrobiological Sciences, Owashi 1-2, Tsukuba 305-8634, Japan

⁵Present address: University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

⁶Present address: The Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada

⁷Present address: 5651-451 Matsukawa Village, North Azumi County, Nagano 399-8501, Japan

*Running title: Isolation and characterization of SGF-2

⁸To whom correspondence should be addressed: Division of Biological Sciences, Faculty of Science, Hokkaido University, North 10 West 8, Kita-ku, Sapporo 060-0810, Japan

E-mail: takiya@sci.hokudai.ac.jp

Tel.: +81-11-706-3590

Fax: +81-11-706-3588

Keywords: transcription; homeodomain; fibroin gene; silk gland factor-2 (SGF-2); silkworm

Background: SGF-2 is a key factor regulating tissue-specific expression of the fibroin gene.

Results: SGF-2 is a 1.1 MDa heteromeric complex containing Awh, Ldb, Lcaf and fibrohexamerin proteins.

Conclusion: Awh, Ldb and Lcaf interact functionally in SGF-2 to control fibroin gene expression.

Significance: This study provides new insight into the functional role of single-stranded DNA-binding proteins in protein-protein interaction and transcriptional regulation.

SUMMARY

SGF-2 bound to promoter elements governing posterior silk gland-specific expression of the fibroin gene in *Bombyx mori*. We purified SGF-2 and showed that SGF-2 contains at least four gene products; the silkworm orthologues of LIM_-homeodomain protein Awh, LIM_-domain binding protein (Ldb), a sequence-specific single-stranded DNA binding protein (Lcaf), and the silk protein P25/fibrohexamerin (fhx). Using co-expression of these factors in Sf9 cells, Awh, Ldb and Lcaf proteins were co-purified as a ternary complex that bound to the enhancer sequence *in vitro*. Lcaf interacts with Ldb as well as Awh through the conserved regions to mediate transcriptional activation in yeast. Misexpression of Awh in transgenic

silkworms induces ectopic expression of the fibroin gene in the middle silk glands, where *Ldb* and *Lcaf* are expressed. Taken together, this study demonstrates that SGF-2 is a multi-subunit activator complex containing Awh. Moreover, our results suggest that the Ldb/Lcaf-protein complex serves as a scaffold to facilitate communication between transcriptional control elements.

Expression of the silk genes is a trait of terminal differentiation of the silk gland of *Bombyx mori* (1). The fibroin gene encoding the silk fiber protein is expressed only in cells of the posterior silk gland (PSG) (1, 2), while the genes for glue proteins, sericins, are expressed only in cells of the middle silk gland (MSG) (3-5). The cell-free transcription systems using silk genes and crude nuclear extracts derived from silk gland tissues (6, 7) led to the *in vitro* reconstitution of tissue-specific transcription of the fibroin gene and the identification of *cis*-elements important for its transcriptional activity (1, 7-9). In particular, the region between -214 and -180 in the upstream promoter element En I, designated as E site, is essential for tissue-specific transcriptional enhancement (7, 8, 10, Fig. 1A). Recently, Shimizu *et al.* demonstrated that additional enhancer elements further upstream, which contain similar sequences of the E site, are also necessary for

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

full activation of the fibroin gene *in vivo* (11).

We have demonstrated the presence of factors that bind specifically to these elements (10, 12). Silk gland factor-1 (SGF-1) is a fork head (Fkh) protein binding to the proximal upstream region (12, 13). Fibroin-modulator-binding protein-1 (FMBP-1), which contains a novel DNA-binding domain, binds to both En I and the intronic element En II (12, 14). Silk gland factor-2 (SGF-2) binds to the E site with two AT-rich repeat sequences, which resemble the consensus sequence recognized by homeodomain proteins. Of these factors, SGF-2 is specifically detected in PSG (10).

In this study, SGF-2 was purified and its composition was determined. SGF-2 contains at least four components; the silk protein P25/fhx, a LIM_-homeodomain (LIM-HD) protein Awh, LIM_-domain binding (Ldb) protein and a member of the sequence-specific single-stranded DNA_-binding protein (Ssdp) family. By misexpression of Awh in transgenic silkworms, expression of the fibroin gene was induced in the middle silk glands, demonstrating that SGF-2 is a tissue-specific activator complex of the fibroin gene.

EXPERIMENTAL PROCEDURES

Electrophoretic mobility shift assay (EMSA) - Each protein-DNA binding reaction contained 5 to 10 fmol of probe (39-mer oligonucleotide with the E site sequence), 1 µg of poly(dI-dC) (Pharmacia GE Healthcare) and protein samples in a volume of 10 µl (15). After incubation on ice for 30 min, samples were analyzed by the electrophoresis on 3.2 % acrylamide gel containing 2.5 % (v/v) glycerol at 4 °C in 0.25 x TBE (22.5 mM Tris-borate (pH8.0), 0.5 mM EDTA) buffer.

Purification of SGF-2 - Commercial silkworm strains (Kin-Shu x Sho-Wa or Shun-Rei x Sho-Getsu from Kanebo Silk Co., Kasugai City, Japan) of *B. mori* were reared at 27 °C on an artificial diet from Kyodo Shiryō Co. (Yokohama, Japan). SGF-2 was purified from crude nuclear extracts of PSG from V2 instar larvae through six column chromatographic steps (Fig. 1C). Crude nuclear extract (protein, 80.0 g; volume, 2,040 ml) from 40,000 pairs of PSG from V2 instar larvae was prepared as described previously (6, 7, 9), and subjected to following purification steps (Fig. 1C). The nuclear extract was diluted 5-fold by adding TEMGTK₀ buffer. TEMGTK buffers contain 20 mM Tris-HCl (pH 7.9), 0.1 mM EDTA, 12.5 mM MgCl₂, 10 % glycerol, 0.1 % Tween 20, 0.1 % PMSF; the number following TEMGTK denotes the concentration (mM) of

KCl. After 30 min stirring at 4 °C, the sample was centrifuged with a JA-10 rotor (Beckman) at 10,000 rpm for 1 hour at 4 °C. The supernatant was filtered through a Y020A047A membrane filter (ADVANTEC, Tokyo), and sequentially applied to a 500 ml of SP Sepharose Fast Flow resin column (Pharmacia GE Healthcare) equilibrated in TEMGTK₂₀. The column was washed with 1,000 ml TEMGTK₄₀ and then eluted with TEMGTK₁₂₀. The eluate (volume: 1,000 ml) was loaded onto a 40 ml Source 30Q resin column (Pharmacia GE Healthcare). The column was washed with 200 ml TEMGTK₁₄₀, and the SGF-2 activity was eluted with TEMGTK₁₉₀. The SGF-2 fraction (protein, 511 mg; volume, 983 ml) was diluted with TEMGTK₀ to adjust the KCl concentration to 100 mM. The diluted eluate underwent DNA-affinity purification using Dynabeads (Dyna-I) on which 4 µg poly(EW) DNA was immobilized per mg beads. Poly(EW) DNA contained tandem repeats of SGF-2 binding sequence derived from the E site of fibroin promoter and was 0.2 to 1.0 kbp length. For one round of DNA-affinity purification, 190 ml diluted eluate was incubated with 10 mg heat-denatured salmon sperm DNA and 10 mg poly(dI-dC) on ice for 10 min, and then mixed with 10 mg of the beads. After the binding reaction at 4 °C for 30 min, the beads were collected using a magnetic stand and the supernatant was removed. The beads were washed 8 times (50 ml, 20 ml, 10 ml, 5 ml, 2 ml, 1 ml, 0.5 ml, 0.2 ml, respectively) with TEMGTK₁₀₀ batch-wise, and bound proteins were eluted with 1 ml TEMGTK₁₀₀₀ twice. The total eluate (0.15 mg, 2.1 ml) was dialyzed against TEMGTK₁₀₀ and loaded onto a column of BioSilect 250 (Bio-Rad) equilibrated with TEMGTK₁₀₀. Fractions containing SGF-2 activity (71 µg, 11.8 ml) were pooled and applied to a 0.1 ml Mini S column (Pharmacia GE Healthcare, SMART) equilibrated in TEMGTK₀. The column was washed with TEMGTK₄₀ and proteins were eluted with TEMGTK₀ containing 6 M urea, and then with TEMGTK₀ containing 6 M guanidine-HCl. The SGF-2 activity was not detected in the TEMGTK₄₀ wash fraction. The eluate with 6 M urea was dialyzed against TEMGTK₄₀ and applied to a 0.1 ml Mini Q column (Pharmacia GE Healthcare, SMART) equilibrated in TEMGTK₄₀. The column was washed with TEMGTK₄₀, and bound proteins were eluted with TEMGTK₀ containing 6 M guanidine-HCl.

Amino acid sequencing of SGF-2 - Purified proteins were resolved by SDS-PAGE, and the bands of interest were subjected to in gel tryptic

書式変更：フォント：斜体（なし）

書式変更：フォント：斜体（なし）

書式変更：フォント：斜体（なし）

digestion, as described previously (13). The generated tryptic peptides were fractionated with a reverse-phase column, and the resolved peptide peaks were subjected to automated Edman degradation on an ABI Procise 477A protein sequencer (Applied Biosystems).

Isolation of cDNAs for SGF-2 subunits – cDNA library prepared using poly(A)⁺ RNA from V2 PSG was screened with a random primed probe made from RT-PCR products amplified using primer sets designed on the basis of the results of amino acid sequencing (Table 1). The positive clones were sequenced. The accession numbers of these cDNA clones of SGF2 subunits, p36 (Awh), p47B (~~LDBLdb~~) and p48/p47G/p45 (Lcaf) are AB687553, AB687554 and AB687556, respectively. During the cDNA cloning of SGF2 p47B and p48/p47G/p45, the other clones, named Ldbβ (AB687555) and Lcafβ (AB687557) were also obtained, which is derived probably from alternatively spliced mRNA.

Preparation of recombinant proteins – Recombinant proteins were produced in Sf9 cells using the baculovirus expression system. DNA fragments encoding proteins of interest were cloned into pFastBac donor plasmids, and recombinant baculoviruses were obtained using Bac-to-Bac Baculovirus Expression Systems (Gibco Invitrogen). For expression of recombinant proteins, Sf9 cells were infected by the virus(es), and cultured at 27 °C for 60 ~~hours~~. The infected cells were lysed in 10 mM HEPES-KOH (pH7.9), 10 mM KCl, 1.5 mM MgCl₂, 0.05 mM ZnSO₄, 1 mM DTT, 0.5 % NP-40, and protease inhibitors (Roche Applied Science). The lysates were centrifuged at 15,000 x g for 10 sec at 4 °C. The precipitate was resuspended in 20 mM HEPES-KOH (pH7.9), 0.38 M (NH₄)₂SO₄, 17 % glycerol, 0.2 mM EDTA, 4 mM MgCl₂, 0.05 mM ZnSO₄, 1 mM DTT and protease inhibitors, and centrifuged at 120,000 x g for 1 ~~hour~~ at 4 °C. The recombinant proteins with His-tag were purified using Ni-NTA agarose (QIAGEN). Each fraction was analyzed by SDS-PAGE and Western immunoblotting with anti-His6 (Babco), anti-HA 12CA5, anti-FLAG M2 (Babco) monoclonal antibodies.

Plasmid construction – Expression plasmids for the yeast two-hybrid assay and yeast one-hybrid assay were constructed using pLexA/NLS, which were LexA-fused protein expression vectors carrying the *TRP1* gene, and pGAD424 for GAL4-AD-fused protein expression vector with the *LEU3* gene (16).

Interaction assay by yeast two-hybrid system – Qualitative interaction assay was performed to measure *HIS3* gene expression (16). A pair of

fusion gene plasmids was introduced into yeast strain L40 by the standard Lithium acetate transformation procedure. Transformants were plated on an SD agar plate containing 10 mM 3-amino-1, 2, 4-triazole (3-AT) without histidine, leucine, tryptophan, lysine and uracil, and incubated overnight at 30 °C.

In vivo transcriptional activity assay by yeast one-hybrid system – The quantitative yeast one-hybrid assay was performed to measure the expression of the β-galactosidase gene under the control of four tandem repeated LexA-binding sequences. Equal amounts of logarithmic-growing yeast transformants expressing each LexA hybrid protein were subjected to β-galactosidase activity assay (17).

Preparation of transgenic silkworms – The Awh ORF was amplified by using primers 5'-agtctagaatgaagacggagcaccgcac_-3' and 5'-agtctagatcagacttcactctgcatgc_-3', and inserted into the *BlnI* site of the pBacUASMCS vector (18), which has a [3xP3-AmCyan] screening marker. The plasmid was injected into *w1-pnd* embryos to obtain the UAS-Awh strains. The established strains were crossed with the hs-GAL4 strain (19).

RESULTS

AT-rich sequences of E site in En I are essential for SGF-2 binding – Our previous results showed that SGF-2 binds to both C and E sites in the upstream enhancer element En I of the fibroin gene, with a stronger preference for the E site (Fig. 1A; 10). Further investigation of the sequence important for binding to SGF-2, EMSA was performed using a series of mutant E site (Fig. 1B). Two AT-rich sequences (boxed in Fig. 1B) are critical for SGF-2 binding, overlap with the protected sequences in an *in vitro* footprint assay using V2 PSG extract, and contain homeodomain protein-binding sequences. A similar AT-rich sequence is found in the C site. The importance of these regions for preferential transcription of the fibroin gene in the PSG extracts has been demonstrated repeatedly previously (7-9).

Purification of SGF-2 – SGF-2 was purified from V2 PSG extract through six ~~chromatographic~~-chromatography steps (Fig. 1C). Figure 1D depicts a silver-stained SDS-PAGE gel containing active fractions from the 3rd step of the purification using DNA-immobilized beads. In this step, we used not only EW oligonucleotide with an intact E site but also with two mutant oligonucleotides, EgCW and EgCM. EgCW contains mutations but maintains SGF-2 binding activity, while mutations in EgCM completely abolish SGF-2

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

binding. The elution profile of SGF-2 activity in the 4th step of purification (size exclusion chromatography using Biosilect 250) correlated with EW oligo-specific polypeptides visualized on silver-stained SDS-PAGE (Fig. 2). The native molecular mass of SGF-2 activity was estimated as about 1.1 MDa by gel filtration chromatography.

In the 5th step of Mini S ion exchange chromatography purification, the bound proteins were eluted by 6 M urea and 6 M guanidine-HCl (Gu-HCl) (Fig. 1C). The urea-eluted fraction contained 6 polypeptides of 48, 47, 45, 33, 32 and 30 kDa (Fig. 2C, lanes 5-8 and 7'), and the Gu-HCl-eluted fraction consisted of 5 peptides with molecular weights of 73, 55, 50, 47 and 36 kDa (Fig. 2C, lanes 9 and 10). In the 6th step, the urea-eluted fraction was further fractionated by Mini Q ion exchange chromatography into 3 polypeptides of 48, 47 and 45 kDa in the Gu-HCl-eluted fraction and 3 polypeptides of 33, 32 and 30 kDa in the flow-through (FT) (Fig. 1C).

cDNA cloning of SGF-2 components - To identify the components of SGF-2, amino acid sequence analysis of the purified polypeptides were performed (Table 1). This analysis revealed that p33, p32, and p30 are derived from the silk protein P25/fhx, which was identified as a fibroin-associated protein (20, 21), and other proteins represent novel *Bombyx* gene products. The peptide sequences from p55, p50B ("B" indicates light-brown protein bands in the silver stained gel) and p47B were mostly identical, and so were those from p48, p47G ("G" indicates gray protein bands) and p45. These results suggest that p55/p50B/p47B and p48/p47G/p45 might represent products from two distinct genes by alternative splicing, respectively.

A 2.2 kb cDNA clone for the 36 kDa protein encodes an LIM-HD protein of 274 amino acids. Since the deduced amino acid sequence is highly homologous to that of the *Drosophila* Arrowhead protein (22) and orthologues in other species (Fig. 3A), we named the protein as *Bombyx* Arrowhead (Awh).

Next, we isolated the cDNA clones for p47B. The predicted protein product, which contains 357 amino acid residues, is highly homologous to mouse Ldb1/NLI/CLIM-2, CLIM-1 and *Xenopus* XLdb1 (23-25). It possesses a LIM domain-interacting domain (LID), which is identical among all Ldb proteins (Fig. 3B, and 26, 27). We designated this protein as *Bombyx* LIM domain-binding protein (Ldb). Supporting our notion that p47B, p50B and p55 are products derived from the same gene, their peptide sequences are found in the predicted

amino acid sequence of Ldb.

Finally, we isolated the cDNA for the SGF-2 components p48, p47G and p45. A 3.0 kb cDNA clone, which encodes a novel protein of 357 amino acid residues containing all peptide sequences from p48, p47G and p45, was isolated and designated as Lcaf (LIM-HD and LDB-Ldb complex associated factor). We searched the DNA database for molecules related to Lcaf and identified sequence-specific single-stranded-DNA-binding protein (SSDP) as the closest relative in the vertebrate (Fig. 3C). The amino acid sequence of Lcaf shows high similarity to that of various vertebrate SSDPs, especially in the N-terminal 92 amino-acid sequence. Interestingly, though SSDP was reported originally as a factor binding to the DNase I hypersensitive region of chicken $\alpha 2(I)$ collagen gene promoter (28), it was also identified as a factor interacting with Ldb proteins (29, 30).

SGF-2 subunits shows restricted or preferential expression in PSG - SGF-2 is detected in the extract of PSG, but not of MSG (10). Northern blot analysis using total RNA derived from the posterior or middle portion of the fifth instar silk glands showed that *Awh* and *P25/fhx* transcripts were only detected in PSG (Fig. 3D). On the other hand, *Ldb* and *Lcaf* transcripts were found in both regions of the silk gland, but preferentially in the posterior portion.

Lcaf forms a DNA-binding protein complex with Awh and Ldb - To examine whether Lcaf forms a complex with other SGF-2 subunits Awh, Ldb and P25/fhx, all four proteins were co-expressed in Sf9 insect cells by using the baculovirus expression system. We constructed recombinant baculoviruses expressing each of HA-tagged Awh (ha:Awh), FLAG-tagged Ldb (f:Ldb), His-tagged Lcaf (h:Lcaf) and Myc-tagged P25/fhx (m:P25/fhx). When cells were infected with baculovirus expressing ha:Awh, f:Ldb or m:P25/fhx individually, the recombinant proteins were insoluble and not recovered well. When cells were infected with the h:Lcaf baculovirus, a 45 kDa protein band together with a minor protein band just above it were detected in the extract-affinity purified fraction (Fig. 4A left panel lane 8). On the other hand, when cells were co-infected with ha:Awh, f:Ldb and h:Lcaf baculoviruses and h:Lcaf protein was purified with Ni affinity chromatography, the 45 kDa protein was co-purified with several proteins in an almost stoichiometric manner. Immunoblotting analysis using anti-His6, anti-HA and anti-FLAG antibodies showed that the proteins co-purified with h:Lcaf were ha:Awh and f:Ldb (Fig. 4A

書式変更: フォント: 斜体 (なし)

right panel). Co-infection of Sf9 cells with m:fhx/P25 baculovirus and the other three baculoviruses was performed, but we did not detect integration of P25/fhx protein into the Awh/Ldb/Lcaf complex.

To examine the possible DNA-binding activity of the h:Lcaf complex, EMSA was performed. As shown in Fig. 4B, DNA-binding activity to the E site was detected in the 60 and 80 mM imidazole fractions of Ni affinity chromatography. The complex migrated slightly faster than native SGF-2 purified from PSG extract (compare lane 14 with lane 15 in Fig. 4B). This DNA-protein complex is super-shifted by the addition of antibodies against HA, FLAG and His epitopes, but not by anti-Myc antibody (Fig. 4C). Most importantly, similar to SGF-2, this complex was specifically abolished by the addition of anti-SGF-2 antibody (Fig. 4C, compare lanes 1 and 2 with lanes 4 and 9). These results clearly illustrate that SGF-2-like complex with specific DNA-binding activity to the E site can be reconstituted by recombinant proteins encoded by *Awh*, *Ldb* and *Lcaf* cDNA.

We also purified the protein complex from Sf9 cells co-infected with f:Ldb- and ha:Awh-expressing baculoviruses using FLAG-tag, and examined its DNA-binding activity. The f:Ldb complex could bind specifically to the E site in EMSA, but the DNA-protein complex migrated much faster than that of the h:Lcaf complex (Fig. 4D lanes 1 and 3). The DNA-protein complex was confirmed to contain both f:Ldb and ha:Awh by super-shift migration using anti-HA and anti-FLAG antibodies, but was not affected by anti-SGF-2 antibody. These results demonstrate that the complex of Awh and Ldb is sufficient for the specific binding to the E site, but is not equivalent to the purified SGF-2.

Induction of ectopic expression of the fibroin gene by Awh in MSG - To investigate whether SGF-2 is a tissue-specific transcriptional activator of the fibroin gene, we generated transgenic silkworms that possess a *UAS-Awh* transgene, in which *Bombyx Awh* was under the control of an UAS promoter. *UAS-Awh* silkworms were crossed with *hs-GAL4* transgenic silkworms, and *hs-GAL4/UAS-Awh* offspring were selected. These transgenic worms were kept at 42°C for two hours on day 1 of the fourth instar, and the expression of the fibroin gene was then analyzed. Strikingly, by misexpression of the *Awh* transgene in transgenic worms, the fibroin gene was induced in MSG (Fig. 5), where *Ldb* and *Lcaf* genes are expressed (Fig. 3D), indicating that Awh protein is a PSG-specific activator of the fibroin gene.

Self-association of Lcaf - To compare the

size of the h:Lcaf complex with native SGF-2, gel filtration chromatography was performed. The elution profiles of the h:Lcaf complex are shown in the top panel of Fig. 6A, in which the majority was eluted in a peak corresponding to a molecular mass of ~800 kDa. Although previous reports showed that the LIM-HD and LDB-Ldb proteins bound to each other to form a heterotetrameric complex *in vitro* (24, 31), it is still possible that Lcaf could oligomerize by itself. When h:Lcaf protein, expressed by the baculovirus system and purified by nickel affinity chromatography, was subjected to gel filtration chromatography, the majority of h:Lcaf was eluted in a peak corresponding to a molecular mass of ~300 kDa (Fig. 6A, bottom panel). This is equivalent to almost six times the predicted molecular mass of a sole h:Lcaf molecule (48 kDa). The self-association ability of the Lcaf protein was confirmed by a yeast two-hybrid system using the GAL4 activation domain (GAL4-AD) and LexA as a DNA-binding portion. We prepared expression plasmids for Lcaf fused to an N-terminal Gal4-AD, called G:Lcaf, along with two Lcaf truncated mutants fused to an N-terminal LexA, which contained the amino-terminal region (1-150 aa) or the carboxyl-terminal region (101-357 aa) of Lcaf, named L:LcafΔC151 and L:LcafΔN100, respectively. As shown in Fig. 6B, G:Lcaf interacted with L:LcafΔC151, but not with L:LcafΔN100, in yeast. These findings indicate that Lcaf protein can form a homo-oligomer through its N-terminal 100 amino acid sequence, which may contribute to the formation of a huge h:Lcaf complex with ha:Awh and f:Ldb.

Lcaf interacts with Ldb - Co-purification of ha:Awh and f:Ldb proteins with h:Lcaf suggested possible direct interactions of Lcaf with Awh and Ldb. The yeast two-hybrid system was used to examine this possibility. We constructed expression plasmids for Ldb as LexA fusion L:Ldb and Awh as GAL4-AD fusion G:Awh (Fig. 7A). Yeast transformants co-expressing L:Ldb and GAL4-AD did not grow on an SD agar plate containing 10 mM 3-amino-1,2,4-triazole (3-AT) without histidine. When L:Ldb was co-expressed with G:Awh or G:Lcaf in the reporter yeast strain, both transformants were able to grow under the same conditions (Fig. 7B).

To define the regions on Ldb involved in the interactions with Awh and Lcaf, two LexA hybrid proteins called L:LdbΔC257 and L:LdbΔN256 that contain the amino-terminal region (1-256) and the carboxyl-terminal region (257-376) of Ldb, respectively, were examined.

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

Our results showed that G:Lcaf ~~□~~ interacts with L:LdbΔC257 (i.e. the amino-terminal portion of Ldb), while G:AwH binds to L:LdbΔN256 (i.e. the carboxyl-terminal portion of Ldb) in yeast (Fig. 7B). These observations suggest that Ldb interacts with AwH and Lcaf through distinct binding domains. To determine which portion of Lcaf is necessary for the Lcaf-Ldb interaction, we used two truncation mutants of Lcaf, as described previously. While G:LcafΔC151 hybrid protein maintained its interaction with L:Ldb in the reporter yeast strain, G:LcafΔN100 lost this ability under the same condition (Fig. 7C), suggesting that the amino-terminal 100 amino acid sequence of Lcaf is necessary for Lcaf-Ldb interaction.

AwH and Lcaf contribute to transcriptional activation - To examine which SGF-2 subunits identified here play a significant role in transcriptional activation, we performed experiments using the yeast one-hybrid system. Yeast transformants expressing L:AwH or L:Lcaf grew well on media containing 10 mM 3-AT without histidine, and were also positive for β-galactosidase activity (Fig. 7D). The β-galactosidase activity of the L:Lcaf expressing transformant was about 8-fold stronger than those expressing L:AwH. On the other hand, yeast transformants expressing L:Ldb or L:P25/fhx did not grow under the same condition and exhibited little or no β-galactosidase activity. These results indicate that both AwH and Lcaf possess intrinsic transcriptional activation ability.

DISCUSSION

SGF-2: A transcriptional activator complex of the fibroin gene - SGF-2 was originally identified by EMSA in extracts of PSG on the basis of the binding activity to the fibroin En I element, and is thought to be a key transactivator for the fibroin gene (6-10). We identified four proteins, AwH, Ldb, Lcaf and P25/fhx, as components of SGF-2. Several lines of evidence support that these proteins constitute SGF-2 and promote transcriptional activation of the fibroin gene. 1. Recombinant AwH, Ldb and Lcaf proteins formed a complex with specific DNA-binding activity to the E site. 2. The protein complex of recombinant AwH, Ldb, and Lcaf was recognized by anti-SGF-2 antibody. 3. AwH, Ldb, Lcaf and P25/fhx are specifically or preferentially expressed in PSG. 4. The purified complex of recombinant AwH, Ldb and Lcaf exhibited almost the same mass as purified SGF-2. 5. AwH and Lcaf showed transcriptional activation activity in yeast one-hybrid system. 6. Misexpression of AwH, which

is normally restricted to PSG, induced ectopic expression of the fibroin gene in MSG of transgenic silkworms. Another indirect evidence supporting the possible contribution of AwH to tissue- and developmental stage-specific transcriptional activation of the fibroin gene is that in *Drosophila* transgenic lines carrying the fibroin promoter fused to the β-galactosidase gene, reporter gene expression is restricted to anterior cells of the larval salivary gland, where the *Drosophila* AwH gene is specifically expressed (22, 32).

The silkworm AwH and Ldb are members of the LIM-HD family of transcription factors and the Ldb protein family ~~DB~~, respectively. The LIM-HD ~~•~~LDB complex appears to be a critical regulator during development and functions as a transcriptional activator (23, 24, 33, 34). Our finding that SGF-2 contains AwH and Ldb is consistent with the notion that ~~LDB~~-Ldb proteins are a requisite component of many transcriptional regulatory complexes involving LIM-HD factors.

P25/fhx is known to be a component of the 2.3 MDa secretory elementary unit of silk fibroin (35). Since P25/fhx is not expressed in MSG and misexpression of AwH can induce fibroin gene expression in MSG, P25/fhx appears to be a non-essential component in the SGF-2 complex. However it will be intriguing to speculate that P25/fhx might play a role in fine-tuning the molecular ratio of fibroin to P25/fhx by regulating the transcription of the fibroin gene through SGF-2. P25/fhx is a glycoprotein (36). The N-linked oligosaccharide chains of P25/fhx are important for maintaining the 2.3 MDa complex of fibroin elementary unit (35). However, recombinant P25/fhx protein in Sf9 cells seemed not to be glycosylated on the mobility on SDS-PAGE. It is possible that glycosylation of P25/fhx is important for its integration into AwH ~~•~~Ldb ~~•~~Lcaf complex.

Lcaf: Additional component of LIM-HD ~~•~~LDBdb complex - The present study identified another protein, Lcaf, a member of the SSDP family, which could transform the AwH/Ldb protein complex into a larger protein complex due to its oligomerization activity. The N-terminal amino acid sequence (~100 amino acids) of Lcaf is almost identical to that of sequence-specific single-stranded-DNA-binding protein (SSDP), which was originally identified as a nuclear protein that binds to the single-stranded pyrimidine-rich element in the chicken ~~α~~-2(I) collagen gene promoter (28). This promoter element is well conserved among different mammalian species and located in a region that is DNase I hypersensitive only when

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

the promoter is active. These observations have led investigators to believe that SSDP might be involved in the transcriptional regulation of the $\alpha 2(I)$ collagen gene. The remarkable conservation of the N-terminus between Lcaf and mammalian SSDP (37) suggests that silkworm Lcaf-protein is a functional homologue of vertebrate SSDP. The conserved domain of Lcaf was necessary not only for its self-oligomerization ability but also for Ldb interaction.

Previous studies in flies and vertebrates have revealed the importance of LIM-HD and ~~LDB~~ Ldb proteins in tissue patterning and differentiation (30). However, the molecular mechanisms in which LIM-HD ~~LDB~~ protein complex functions as transcriptional regulator are not fully understood. Genetic experiments imply that Chip, a *Drosophila* homologue of ~~LDB-Ldb~~ proteins, may mediate communication between enhancers and promoters (34, 38, 39). Given that the SSDP/Lcaf protein family is a requisite interaction partner for ~~LDB-Ldb~~ proteins, they might play a role in long-range enhancer-promoter communication by cooperating with ~~LDB-Ldb~~ proteins. In this scenario, the potential sequence-specific single-stranded DNA -binding activity, interaction activity with ~~LDB-Ldb~~ proteins and the self-oligomerization activity of the SSDP/Lcaf protein family could facilitate enhancer-promoter communication by gathering together

sequence- and structure-specific *cis*-elements scattered throughout certain gene loci and by organizing transcriptional regulatory elements on chromatin to form particular higher-order structures that support transcriptional regulation. From this point of view, it is important to stress that besides the En I region of the fibroin gene, a key region in the further upstream enhancer element from -1659 to -1590 detected *in vivo* and localized near a DNase hypersensitive site also possesses an SGF-2-binding sequence (11, Takiya unpublished results). The mouse ~~LDB~~ Ldb protein was found to occupy numerous DNase I hypersensitive sites on chromatin across a region of ~130 kb in the ~~mouse~~ mouse α -globin locus (40). Long-range genomic interaction via Ldb1 and GATA1 was reported also in mammalian β -globin gene locus (41). It would be interesting to investigate whether the SSDP/Lcaf protein family can co-occupy the same positions as the ~~LDB-Ldb~~ proteins, and if so, how it could contribute to the regulation of developmental gene expression, such as the formation of intra-chromosomal loops and histone modification (42). Recently, Brandt and his coworkers (43, 44) reported that SSDPs regulate the activity of Ldb-containing complex through stabilization of Ldb proteins by interfering proteasomal degradation. SSDPs may be a multifunctional component in transcriptional regulation.

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

書式変更: フォント: 斜体 (なし)

REFERENCES

1. Suzuki, Y., Takiya, S., Suzuki, T., Hui, C.-C., Matsuno, K., Fukuta, M., Nagata, T., Ueno, K. (1990) *Molecular Insect Science*. Ed by Hagedorn, H.H. *et al.* Molecular Insect Science. Plenum Press, NY, pp.83-89.
2. Suzuki, Y. and Suzuki, E. (1974) *J. Mol. Biol.* **88**, 393-407
3. Okamoto, H., Ishikawa, E. and Suzuki, Y. (1982) *J. Biol. Chem.* **257**, 15192-15199
4. Couble, P., Michaille, J.-J., Garel, A., Couble, M.-L. & Prudhomme, J.-C. (1987) *Dev. Biol.* **124**, 431-440.
5. Takasu, Y., Hata, T., Uchino, K. and Zhang, Q. (2010) *Insect Biochem. Mol. Biol.* **40**, 339-344.
6. Tsuda, M. and Suzuki, Y. (1981) *Cell* **27**, 175-82.
7. Suzuki, Y., Tsuda, M., Takiya, S., Hirose, S., Suzuki, E., Kameda, M. and Ninaki, O. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9522-9526.
8. Tsuda, M. and Suzuki, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7442-7446.
9. Takiya, S., Hui, C.-C. and Suzuki, Y. (1990) *EMBO J.* **9**, 489-496.
10. Hui, C.-C., Matsuno, K. and Suzuki, Y. (1990) *J. Mol. Biol.* **213**, 651-670.
11. Shimizu, K., Ogawa, S., Hino, R., Adachi, T., Masahiro, T. and Yoshizato, K. (2007) *Insect Biochem. Mol. Biol.* **37**, 713-725.
12. Takiya, S., Kokubo, H. and Suzuki, Y. (1997) *Biochem. J.* **321**, 645-653.
13. Mach, V., Takiya, S., Ohno, K., Handa, H., Imai, T. and Suzuki, Y. (1995) *J. Biol. Chem.* **270**, 9340-9346.
14. Takiya, S., Ishikawa, T., Ohtsuka, K., Nishita, Y. and Suzuki, Y. (2005) *Nucleic Acids Res.* **33**, 786-795.
15. Suzuki, T. and Suzuki, Y. (1988) *J. Biol. Chem.* **263**, 5979-5986.

16. Vojtek, A. B., Hollenberg, S. M. and Cooper, J. A. (1993) *Cell* **74**, 205-214.
17. Savoyesky, E., Mizuno, T., Sowa, Y., Watanabe, H., Sawada, J.-i., Nomura, H., Ohsugi, Y., Handa, H. and Sakai, T. (1994) *Oncogene* **9**, 1839-1846.
18. Sakudoh, T., Sezutsu, H., Nakashima, T., Kobayashi, I., Fujimoto, H., Uchino, K., Banno, Y., Iwano, H., Maekawa, H., Tamura, T., Kataoka, H. and Tsuchida, K. (2007) *Proc. Natl. Acad. Sci. USA* **104**, 8941-8946.
19. Uchino, K., Imamura, M., Sezutsu, H., Kobayashi, I., Kojima, K., Kanda, T. and Tamura, T. (2006) *J. Insect Biotech. Sericol.* **75**, 89-97.
20. Couble, P., Moine, A., Garel, A. and Prudhomme, J.-C. (1983) *Dev. Biol.* **97**, 398-407.
21. Chevillard, M., Couble, P. and Prudhomme, J.-C. (1986) *Nucleic Acids Res.* **14**, 6341-6342.
22. Curtiss, J. and Heilig, J. S. (1997) *Dev. Biol.* **190**, 129-41.
23. Agulnick, A. D., Taira, M., Breen, J. J., Tanaka, T., Dawid, I. B. and Westphal, H. (1996) *Nature* **384**, 270-272.
24. Jurata, L. W., Kenny, D. A. and Gill, G. N. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 11693-11698.
25. Bach, I., Camiere, C., Ostendorff, H. P., Andersen, B. and Rosenfeld, M. G. (1997) *Genes Dev.* **11**, 1370-1380.
26. Breen, J. J., Agulnick, A. D., Westphal, H. and Dawid, I. B. (1998) *J. Biol. Chem.* **273**, 4712-4717.
27. van Meyer, D. J., O'Keefe, D. D., Thor, S., Jurata, L. W., Gill, G. N. and Thomas, J. B. (2000) *Development* **127**, 1823-1831.
28. Bayarsaihan, D., Soto, R. J. and Lukens, L. N. (1998) *Biochem. J.* **331**, 447-452.
29. Chen, L., Segal, D., Hukriede, N. A., Podtelejnikov, A. V., Bayarsaihan, D., Kennison, J. A., Ogryzko, V. V., Dawid, I. B. and Westphal, H. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 14320-14325.
30. van Meyer, D. J., Thomas, J. B. and Agulnick, A. D. (2003) *Development* **130**, 1915-1925.
31. Hobert, O. and Westphal, H. (2000) *Trends Genet.* **16**, 75-83.
32. Nony P., Prudhomme J.-C. and Couble P. (1995) *Biol. Cell* **84**, 43-52.
33. Rincon-Limas, D. E., Lu, C.-H., Canal, I. and Botas, J. (2000) *EMBO J.* **19**, 2602-2614.
34. Morcillo, P., Rosen, C., Baylies, M. K. and Dorsett, D. (1997) *Genes Dev.* **11**, 2729-2740.
35. Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K. and Mizuno, S. (2000) *J. Biol. Chem.* **275**, 40517-40528.
36. Tanaka, K., Inoue, S. and Mizuno, S. (1999) *Insect Biochem. Mol. Biol.* **29**, 269-276.
37. Bayarsaihan, D. (2002) *Biochem. Biophys. Acta.* **1599**, 152-155.
38. Dorsett, D. (1999) *Curr. Opin. Genet. Dev.* **9**, 505-514.
39. Morcillo, P., Rosen, C. and Dorsett, D. (1996) *Genetics* **144**, 1143-1154.
40. Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W. G. and Higgs, D.R. (2004) *EMBO J.* **23**, 2841-2852.
41. Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory P. D. and Dean, A. (2012) *Cell* **149**, 1233-1244.
42. Dean, A. (2006) *Trends Genet.* **22**, 38-45.
43. Xu, Z., Meng, X., Cai, Y., Liang, H., Nagarajan, L. and Brandt S. J. (2007) *Genes Dev.* **21**, 942-955.
44. Wang, Y., Klumpp, S., Amin, H. M., Liang, H., Li, J., Estrov, Z., Zweidler-McKay, P., Brandt S. J., Agulnick, A. and Nagarajan, L. (2010) *Oncogene* **29**, 3044-3053.

Acknowledgments - We are grateful to the members of Y. Suzuki, H. Handa and H. Sezutsu laboratories and associated groups of Drs. H. Kokubo, C. Sawa, K. Morohashi and Y. Nagahama, and Ms E. Suzuki, M. Sasaki and M. Masuda for their technical assistance and encouragement throughout this study. We also thank Dr. Kikawada for the preparation of total RNA from silk glands and Dr. Hollenberg for providing the yeast two-hybrid system.

⁸The abbreviations used are: SGF-2, silk gland factor 2; LIM-HD, LIM-homeodomain; LDB, LIM-domain binding; SSDP, sequence-specific single-stranded DNA-binding protein; LID, LIM-interacting domain; PSG, posterior silk gland; MSG, middle silk gland; EMSA, electrophoretic mobility shift assay; PAGE, polyacrylamide gel electrophoresis; Awh, Arrowhead; 3-AT, 3-amino-1, 2, 4-triazole; ONPG, o-nitrophenyl-γ-D-galactopyranoside; GAL4, galactose 4; UAS, upstream activation sequence

FOOTNOTES

This work was partially supported by a Grant-in-Aid for Scientific Research on Priority Areas from The Ministry of Education, Science, Sports and Culture, and by a Grant from the Ministry of Agriculture, Forestry, and Fisheries (MAFF) of Japan, a Research Grant from Core Research for Evolutional Science and Technology (CREST) of the Japan Science and Technology Corporation (JST), Research Fellowships from the Japan Society for the Promotion of Science for Young Scientists, and a grant for R and D Projects in Cooperation with Academic Institutions from New Energy and Industrial Technology Development Organization (NEDO).

FIGURE LEGENDS

Figure 1. Purification of SGF-2. **A.** The effects of mutations in AT-rich regions of En I element (-211 ~ -91) on PSG-specific transcriptional enhancement and SGF-2 binding are summarized. The AT-rich sequences (E box) recognized by SGF-2 are boxed. **B.** The binding activity of each mutation construct of the E site to SGF-2 in EMSA was indicated as + and - on the left side. The result of *in vitro* footprint assay performed by Hui *et al.* (10) was also summarized. + at the top indicates positions protected by SGF-2 in footprint assay. The 'consensus' at the top indicates the Homeodomain protein-binding sequences. The nucleotide sequence of the C site is shown on the bottom line. **C.** Column chromatography procedures of SGF-2 purification. **D.** Comparison of peptides purified by DNA-immobilizing beads with intact and mutated E box from Source 30Q-purified SGF-2 fraction. Top represents the sequence of one unit of DNA. EW and EgCW contain the intact sequence of the E box, while EgCM has mutated sequence. These DNAs were polymerized and immobilized onto beads. Bottom represents silver-stained SDS-PAGE gels of fractions bound to each DNA-immobilized bead.

Figure 2. SGF-2 is a 1.1 MDa heteromeric complex. **A.** Chromatography chart of SGF-2 activity passed through BioSilect 250 gel filtration column. **B.** EMSA was performed using the E site probe and each fraction passed through BioSilect 250 column. Source Q fr is the peak fraction from the Source 30Q column, and the protein amount changed in lanes 1-5. **C.** Silver-stained SDS-PAGE gel of BioSilect 250 (lanes 2-4), Mini S (lanes 5-8) and Mini Q (lanes 9-10) fractions. Lane 7' shows a magnified image of lane 7. M in lane 1 is a size marker.

Figure 3. Deduced amino-acid sequences of SGF-2 subunits and their restricted expression in the posterior silk gland. **A.** Deduced amino-acid sequence of p36 protein, *Bombyx* Arrowhead (Awh), is compared with orthologues of other animals. Identical amino acids are indicated with white letter on black. Bm; *Bombyx mori*, Dm; *Drosophila melanogaster*, Hs; *Homo sapiens*, Mm; *Mus musculus*, Xl; *Xenopus laevis*, Dr; *Danio rerio*, Ce; *Caenorhabditis elegans*. **B.** Deduced amino-acid sequence of p48 protein, *Bombyx* Ldb, is compared with orthologues of other animals. **C.** Deduced amino-acid sequence of p45 protein, *Bombyx* Lcaf, is compared with orthologues of other animals. **D.** Northern blot analysis was performed using total RNA derived from middle and posterior silk gland at the fifth instar. cActin A3: cytoplasmic actin A3, used as a control.

Figure 4. Lcaf forms a DNA-binding protein complex with Awh and Ldb. **A.** Left: Silver-stained SDS-PAGE gel of Ni-NTA-agarose purification fractions from the extract of Sf9 cells co-expressed with ha:Awh, f:Ldb and h:Lcaf by the baculovirus expression system. W in lane 1: wash fraction of Ni-NTA-agarose. Numbers at the top indicate imidazole concentration (mM) for elution. In lane 8, the purified fraction from Sf9 cell extract expressing only h:Lcaf. The character of native Ldb protein with light-brown color in the silver stained gel was also observed for recombinant Ldb protein, but the slight slower migrating band than the h:Lcaf band in lane 8 was grey-colored, not light brown; therefore it seemed not to be Ldb. Right: Western blot analysis using Ni-NTA-agarose purification fractions. Each lane corresponds to that of the silver-stained SDS-PAGE gel shown on the left. **B.** EMSA using Ni-NTA-agarose fractions of h:Lcaf complex with the E box DNA probe. Numbers and SQ at the top indicate imidazole concentration (mM) for elution and partial purified native SGF-2 fraction after Source 30Q column, respectively. In lanes In and FT, the extracts and the flow through fraction of Ni-NTA-agarose purification. **C.** EMSA of h:Lcaf complex with antibodies. The h:Lcaf

complex eluted by 80 mM imidazole was used with each antibody indicated at the top. Anti-IgG and anti-c-Myc antibodies were used as negative controls. D. EMSA of f:Ldb complex with antibodies. The f:Ldb complex eluted by 80 mM imidazole was used with each antibody indicated at the top.

Figure 5. Misexpression of Awh induced expression of the fibroin gene in MSG. **A.** Total RNA of MSG or PSG was prepared individually from two larvae (No. 1 and No.2) of the wild type (+/+) silkworms or the silkworms carrying *hs-Gal4* transgene alone (+/hs-Gal4), with (IV1d hs → IV2d) or without (IV2d) heat shock at 42 °C. ~~The Number of cycle numbers~~ of PCR ~~are~~ is 35 for *Awh* and 21 for *H-fib* and *Rp49* control. **B.** RT-PCR was performed with total RNA from two strains of transgenic worms (Tb47-1/hs-Gal4 and Tb47-2/hs-Gal4) with *UAS-Awh* transgene in different locus.

書式変更: フォント: 斜体 (なし)

Figure 6. Self-association of Lcaf protein. **A.** Elution profiles of the h:Lcaf complex (top) and h:Lcaf protein (bottom) in gel filtration chromatography. Numbers at the top indicate fraction numbers of the chromatography. Elution of markers is shown on the horizontal axis. M and IN at the top are the molecular weight marker and input protein fraction on chromatography, respectively. **B.** Self-association of Lcaf in yeast two-hybrid analysis. *HIS3* gene was used as a reporter gene. Top: Schematic structures of Lcaf hybrid proteins. Bottom: Each yeast transformant was inoculated on an SD agar plate containing 10 mM 3-AT without histidine or on SD agar with histidine.

Figure 7. Mutual interaction of Lcaf and Awh with Ldb, and transcriptional activation by Lcaf and Awh. **A.** Schematic structures of hybrid proteins used in the yeast two-hybrid assay. **B.** Lcaf and Awh interact with the amino- (1-256 aa) and carboxyl-terminal (257-376 aa) portion of Ldb, respectively, in the yeast two-hybrid assay. **C.** Ldb interacts with the amino-terminal portion (1-151 aa) of Lcaf. *HIS3* gene was used as a reporter gene. Each yeast transformant was inoculated onto an SD agar plate containing 10 mM 3-AT without histidine or on SD agar with histidine. **D.** Yeast one-hybrid assay was performed using each SGF-2 subunit fused with LexA and nuclear location signal. *HIS3* gene and β -galactosidase gene were used as reporters. Left shows that yeast transformants expressing Awh and Lcaf grew well in media containing 10 mM 3-AT without histidine. Right represents β -galactosidase activity of each yeast extract. All results are the mean $\pm$ S.E. of at least four independent transformants.

Table 1: Amino acid sequences ~~from~~ digested peptides of SGF-2 components

p48	#1	LALYVVEYLLHVGAT ^(g)	p36	#1	<u>MLTEHRTCCACGEPIAD-</u>
p47G	#2	AAQTFLSEIR ^(g)			RFLLEVGGAAWHT ^(a)
p45	#3	NITLGEPPGF ^(g)		#2	TCCACGEPIADR ^(a')
	#4	SSPGGVGGGGPGTP ^(g)		#3	FLLEVGGAAWHTGCLR
	#5	EDSGSGM <u>GDYNLS</u> FGG- PGGDQSDQTESAAIL ^(g)		#4	CCVCAVQLDR
	#6	IKESMQEER ^(g)		#5	HPSCFLR
	#7	<u>FEKDPDHPDYFM</u> ^(g)		#6	QVYCK
				#7	GISSSDWVR
				#8	<u>EQVYHLACE</u>
				#9	QLSTGEQFALHED
				#10	VLCKP
p55	#1	ELIPR ^(e)		#11	VTQVWFQNR
	#2	HTPYFGQPDYR ^(b)		#12	QNQLMSR
				#13	PINLHLTY
p50B	#1	HTPYFGQPDYR ^(b)			
	#2	TLIPR			
	#3	LILEFTFDDLMR ^(c')	p33	#1	GQIPSQYEIPVFQFEIPYF
	#4	SWHMAVR ^(d)		#2	ATYVD
	#5	ELIPR ^(e)		#3	NLI
	#6	LCVILEPMQELMSR ^(f)			
	#7	TTLFQK	p32	#1	EDVVLSEFYIDGSYS
p47B	#1	VALGSL	p30	#1	MLAR
	#2	HTPYFGQPDYR ^(b)		#2	PCYLDDYK
	#3	<u>VYELNK</u>		#3	CIPGR
	#4	VCTEGRLLILEFTFDDLMR ^(c)		#4	NHDQCR
	#5	SWHMAVR ^(d)		#5	TLAQHMSFK
	#6	ELIPR ^(e)		#6	LTTVFDDK
	#7	<u>LCVILEPMQELMSR</u> ^(f)		#7	TAQWLSK
	#8	WQR		#8	EHIFGK
	#9	KGSAGANAAP		#9	NWLAR
	#10	VALG		#10	TLCDFGCQH

Oligonucleotides encoding underlined sequences were used for RT-PCR to amplify partial cDNA fragments of the corresponding SGF-2 components.

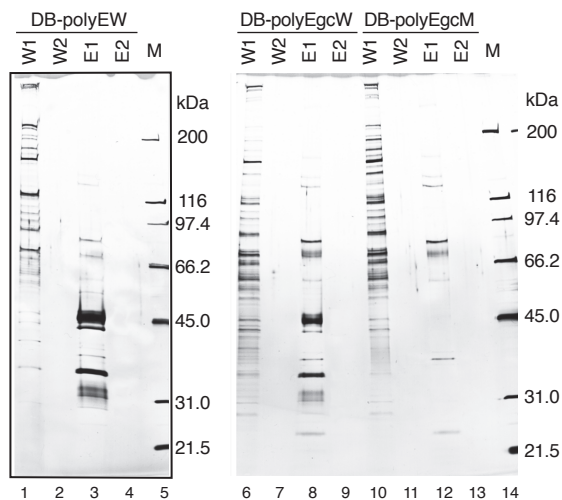
(a) sequence was from undigested p36 protein

(a') sequence was included in (a)

(b) and (d)-(f) were the same sequence

(b) includes (c') sequence

(g) sequences were found in every proteins of p48, p47G and p45



A

```

Bm      MKTE-H
Dm      MKTEL
Hs      MSEE---CGRTT-----ALAAGRTRKGAGEGLVSPEGAGDE
Mm      MAQPGSGCKATTRC-----LEGTAPPAMAQSDAEALAG-ALDKEG
XI      MDIATGPESLDRCFTRGPSDCAKMLDTI
Dr      MYWK---NEQTT-----SKES-PSKNTAKTQDLF-----EEV
Ce      MDAHLVQAKKTSTASELSDSSLTFPFIGDYLSSPSLTTSDDYSDCSNLVTEGVPVPAQEF

Bm      RT-----CCACGEPTADRFLLEVGGAAWHT
Dm      RS-----CAACGEPTSDRFLEVGGSYHA
Hs      SCSSSAPLSPSSS-----PRSMASGSGCPPGKVCVNSCGLEIVDKYLLKVNDLCWHV
Mm      RASPCPTSPSPVC-----SPPSAASSVPSAGKNICSSCGLEILDRYLLKVNNLTHV
XI      RMEDHPLRTGTATLG-----VLLGSECQHQAVCEGCRPTSDRFLMRVNEASWHE
Dr      SCSSSLSLSPSSS-----PHTMT-----EKAICTSCGTEIVDKYLLKVNDMCWHV
Ce      SSSDESSVYISSALRLADYAFTPDNIRIKPDIAVICTQCQHQIQDKFFLSDGRNYHE

Bm      GCLRCCVCAVOLDRHPSCFLDRDQVYCK-----ODYAKSFGAKCSKCCRGISS
Dm      HCLRCCVCMCPIDRQSCFTREROVYCK-----ADYSKNFGAKCSKCCRGISAS
Hs      RCLSCSVCRTSLGRHTSCYIKDXDIFCK-----LDYFRRYGTRCSRGRHIST
Mm      RCLSCSVCRTSLRQONSQYIKNKEIYCK-----MDYFSRFGKGARCGRQIYAS
XI      ECLQCTVCQQLPTT--SCYFRDKLFCCK-----DYQQLFAAKCSGCMKIAPT
Dr      RCLSCSVQTSLSGRHTSCYIKEKEIFCK-----LDYFRKYGTRGHCGRNTHSN
Ce      NCLQCTCENPLSN--KCFYKDKTFYCKGCFYRTHVTSTASSCRELCPKASCDRTIQAT

Bm      DWVRKAREQVYHLACFACDAAGRQLSTGEQFALHEDRVLCRPHYLETLDDGSISSDDG--
Dm      DWVRKARELVFHLACFACDAQGRQLSTGEQFALMDRVLCRKAHYLETVEGGTSSDEG--
Hs      DWVRKAGNVYHLACFACFSCGRQLSTGEEFALVEEKVLCRVHYDCMLDNLKREVENGN
Mm      DWVRKARGNAYHLACFACFSCGRQLSTGEEFGLVEEKVLCRTHYDTMIENLKRAENGN
XI      EFVMRALQVYHLSCFCCCVGERQLRKGEFVLKEGQLLCKSDYEKEKDLLSGSPDDSD
Dr      DWVRKAGNTYHLACFACFSCGRQLSTGEEFALVDERVLCRVHYDCMLDNLKRAMENK-
Ce      DWVRKARNVYVHLACFSCNQCKRQLSTGEEYALQEGNLLCKQHIFELVEG-----DSG--

Bm      -----CDSEGYHKSIAKRVRTTFTEEQQLVQLQANFQDSNPDG
Dm      -----CDGDGYHKSITKRVRTTFTEEQQLVQLQANFQDSNPDG
Hs      -----GISVEGALLTEQDVNHPKPAKRARTSETADQLQVMOAQFAQDNNPDA
Mm      -----GLTLEGAVPSEQDS-QPKPAKRARTSETAEQLQVMOAQFAQDNNPDA
XI      SVKSDDEEGDVKPGKGRVNGQKGSDDCKDPRRPKRPRITILITQORRAFKASFEVSSKPCR
Dr      -----GVNVEGAVPSEQEVNQPKPAKRARTSETADQLQVMOAQFAQDNNPDA
Ce      -----VSS---QKAKTKRVRTTFADQLSVLQTYFNDRSNPDG

Bm      QDLERIAQVTGLSKRVTVWFONSRAROKKHQHTGKGNOLMSREGDAVGFGRPINLHL
Dm      QDLERIASVTGLSKRVTVWFONSRAROKKHITAGKNK-----IREPEGSSFAHINLOL
Hs      QTLQKLAERTGLSRRVTVWFONCRARHKHVSPNHSSSTPVTAVPPSRLSPPMLEEMAY
Mm      QTLQKLAEMTGLSRRVTVWFONCRARHKKHT-POHP--VPPSGAPPTRLPSALSDDIHY
XI      KVRETLAAETGLSVRVTVWFONQRAIKKLARRHQQDEQONSQRLGQEVMSRMEGM
Dr      QTLQKLAERTGLSRRVTVWFONCRARHKHVSPNHSSAAPVSSLQSAHLSPLIDELQY
Ce      ADLEKIASMTGLSKRVTVWFONSRAROKKWHQKSEED-----NGDSQRSSVG-----PS

Bm      TYSFQNMPPFVPIIDTGS-----FTDSSMDE
Dm      TYSFQNNAQNPMHLNGSKAGLYP-----THESMDE
Hs      SAYVPQDGTMLTALHSYMDAHS-----PTTLGLQP
Mm      SPFSSPERARMVTLHGYESHP-----FSVLTLP
XI      TSYAPLAPSQQQLVIMDQNSYSTDPFQQGLTPPQMPGDHMPYGNDTIFHDIDSQTLTS
Dr      TAFAPVDTPMLTALHSYMDVHS-----PISVLQP
Ce      SPSSQKSDSSSEMYPIS-----VTTISVED

Bm      LSEDSS--IHCMSQEV
Dm      LSQDSS--VHCMPSEV
Hs      LLPHSM--TQLPISHT
Mm      LAHLSMGTTLQLPLSR
XI      LDCFLASSEVTSMOARVGNPIDRLYSMQSSYFAS
Dr      LLPHSM--TQLPISHA
Ce      AIPDSI--VILGSLQFD

```

B

Bm -
Dm MNRRGLNAGNTMTSQANIDDGSKAVSEGGSM LPA SN SAVLNPDG SNQSGFAQGGLPYNS
Hs -
Mm -
Xi -
Dr -
Ce -

Bm MPVALGS-----
Dm AGNPYPAGQSSPAGNQSI VFQNSNQPGSNT PQYTSSAPS GSSTPGVGAQNI PGNY PQ
Hs -
Mm -
Xi -
Dr -
Ce MCVVIFLKFLFSDNGNNQHGTAAFGRL

Bm -----LG-----GEYRDYGHHLSPHTPHPPHIQY-----HEYAPPPP
Dm SATAGNFNGPVGGPFGSPSSGLQFSRPASSGTFNSGQAGHFSSTPTVFSVGQFNPMPP
Hs MSSTPHD-----
Mm MSSTPHD-----
Xi MSSTPHD-----
Dr MSNPPHD-----
Ce AQS YRIQDLLFIIVIFFRAIHIMMFHRPPSARPLNHVDQLKSLQATALG-----

Bm P-----PIYHQIPDPYFRRHTPYFGQPDYRVYELNKRLQORT----EDSD
Dm ASPFGHGGNHMMGGPOOMERIDGERRHNSYFSHTEHRYVELNKRLQORN----EESD
Hs PFYSSPFGPFYRRHTPYMVQPEYRIYENKRLQSRIT----EDSD
Mm PFYSSPFGPFYRRHTPYMVQPEYRIYENKRLQSRIT----EDSD
Xi PFYSSPFGPFYRRHAPYMVQPEYRIYENKRLQSRIT----EDSD
Dr PFYSSPFGPFYRRHSPYMVQPEYRIYENKRLQSRIT----EESD
Ce -----KPGMRSQATEPQPIGNTVSPLEFRIHDMNRRLYIFSSTGVSEN DQ

Bm NFWWDAFATEFFEDD ATLTLSFCLE----DGPKRYTIGRTLIPRYFRSIYEGGVSELYYT
Dm NQWWSFITEFFEDDARLTLSFCLE----DGPKRYTIGRTLIPRYFRSIYEGGVSDLYFQ
Hs NLWWDAFATEFFEDD ATLTLSFCLE----DGPKRYTIGRTLIPRYFSTVFEGGVTDLYII
Mm NLWWDAFATEFFEDD ATLTLSFCLE----DGPKRYTIGRTLIPRYFSTVFEGGVTDLYII
Xi NLWWDAFATEFFEDD ATLTLSFCLE----DGPKRYTIGRTLIPRYFSTVFEGGVTDLYII
Dr SLWWDAFATEFFED ATLTLSFCLE----DGPKRYTIGRTLIPRYFSTVFEGGVTDLYII
Ce QQWWDASFHEFFDDCKLWFVIGSEPVAFASRERYTINRQFIKFERSTFDSGMRELQYV

Bm MRQPKESFHNSTITLDCDHCTMVTHHGKPMFTKVCTEGRLLILEFTFDDLM--RIKSW
Dm LKHAKE--SFHNSTITLDCDQCTVITQHGKPFFTKVCADARLLILEFM--YDDYM--RIKSW
Hs LKHSKE--SYHNSTITVDCDQCTMVTHHGKPMFTKVCTEGRLLILEFTFDDLM--RIKSW
Mm LKHSKE--SYHNSTITVDCDQCA MVTHHGKPMFTKVCTEGRLLILEFTFDDLM--RIKSW
Xi LKHSKE--SYHNSTITVDCDQCTMVTHHGKPMFTKVCTEGRLLILEFTFDDLM--RIKSW
Dr LKHSKE--SFHNSTITVDCDQCTMVTHHGKPMFTKVCTEGRLLILEFAFDDLM--RIKSW
Ce LRGPSRECTLANGQAYENENVLQITRYDQSSQFVNTEGKLYVEFAPFDEVNRYRIKAW

Bm HMAVAHARELIPROAV---HPPDHAALDQLTKNITROGITNSTLNYLR LCVILEPMQELM
Dm HMTIKGHRELIPRSVIGTSLPPDPMLLDQITKNITRAGITNSTLNYLR LCVILEPMQELM
Hs HFTIRQYRELIPRSILA-MHAADPQVLDQLSKNITRMGLTNFTLNYLR LCVILEPMQELM
Mm HFTIRQYRELIPRSILA-MHAADPQVLDQLSKNITRMGLTNFTLNYLR LCVILEPMQELM
Xi HFTIRQYRELIPRSILA-MHAADPQVLEQLSKNITRMGLTNFTLNYLR LCVILEPMQELM
Dr HFNIRQYRELIPRSILA-MHAADPGVLEQLSKNITRMGLTNFTLNYLR LCVILEPMQELM
Ce TLELKRSEFVYNQNTA---DYRVEAQNPQENKPRMGFFKSTFNLMTMLKILDPMSIM

Bm SRHKAYALSPRDCLKTTLFQKWQR-----MVAPP-----
Dm SRHKAYALSPRDCLKTTLFQKWQR-----MVAPPG-----
Hs SRHK-TYNLSPRDCLKTQLFQKWQR-----MVAPPA-----
Mm SRHK-TYNLSPRDCLKTQLFQKWQR-----MVAPPA-----
Xi SRHK-TYNLSPRDCLKTQLFQKWQR-----MVAPPA-----
Dr SRHK-TYNLSPRDCLKTQLFQKWQR-----MVAPPAGHPQNFKEIFPTNH
Ce SSAKSA PATTPREV MKRTL FQHHQVRQNM RQQQLNQMMITPAPE-----

Bm ---ESORPASKRRKRKGS---AGAN--AAPPAPAKRSP--GPNFSLASQ-----
Dm -KKDPORPPNKKRRKRKGSNS--GGGNSNTPPVTNQKRSPSGPSFSLSSQ-----
Hs --EPTROTITKRRKRKNSSTSSNSAGNANSTGSKKTTAANLSLSQV-----
Mm --EPTROTITKRRKRKNSSTSSNSAGNTTNSAGSKKTPAASLSLATQ-----
Xi --EPTROTITKRRKRKNSSTNNASNSNAGNNATSAYNRKKVPAASLNSQV-----
Dr KKEPTROTITKRRKRKNSASSASNSLGN---SAGGKRSPANNFSLASQ-----
Ce --PEKPKPARKRQRKPAANPRGSKKATAAAAAAATNGVPITVPTASANNQQFPPNP

Bm -----DVMVVGEPSLMGGEFGDEDERLITRLENTQYEGE-----
Dm -----DVMVVGEPTLMGGEFGDEDERLITRLENTQYDGTNAVEHDNHTGFGHAD
Hs -----PGLGAIPNCSLNPGRDGD-----
Mm -----GLGAIPNCSLNPGRDGD-----
Xi -----PDVMVVGEPTLMGGEFGDEDERLITRLENTQYDAANGMDDEED-----
Dr -----DVMVVGEPTLMGGEFGDEDERLITRLENTQYDATNGLDDDD-----
Ce MTSQFQQMSYPDVMVVGEPSVMGSEFGENDERITISRVENSQYDPNAMQMQLSGQGNSSM

Bm -----GVEWSAPPASP-----AKTTPGTH
Dm SPISGSNPWSIDRAGATPASPNGGAAPQNNANISDIDKKSPIVSQ
Hs -----LCHSTAVTPSGQFKEKH
Mm -----LCHSTAVTPSGQFKEKH
Xi -----FNSSPALGNNSPWNKPPNAETKSDNPTQQA SQ
Dr -----FTSSPALANNSPWNKPPPTGQDSKPESTASQSQ
Ce N-INGRNMNQHPGMQPPGQGHMPPHSMGSMPTSMHNPMMHPPGSMQGHGGMPPMPS

Bm -
Dm -
Hs -
Mm -
Xi -
Dr -
Ce TMANQMPPNLPPTHMSNQMPNSRMPPMQGMPPSGMPGQMSNPNMMSGMPMSMSSQM

Bm -
Dm -
Hs -
Mm -
Xi -
Dr -
Ce PGSMSPNMPNQMPGGMQNMQMPPNYSQYTGPPPPQWPPPN SAMITG

C

```

Bm MYAKGKGSTVPSDAQAREKLALVYVEYLHVGAIAAAQTFLEIR-----WEKN
Dm MYGSKTSAVPSDAQAREKLALVYVEYLHVGAIAAAQTFLEIR-----WEKN
Hs MFAKGGKSAVPSDAQAREKLALVYVEYLHVGAIAAAQTFLEIR-----WEKN
Mm MFAKGGKSAVPSDAQAREKLALVYVEYLHVGAIAAAQTFLEIR-----WEKN
XI MYAKGKGSTVPSDAQAREKLALVYVEYLHVGAIAAAQTFLEIR-----WEKN
Dr MFAKGGKSAVPSDAQAREKLALVYVEYLHVGAIAAAQTFLEIR-----WEKN
Ce MPPQVIQQQQSLASEMTARDRLTSYIYVEYLQQTGASKTAETTFKEVLESTNPAAGLAAAN

Bm ITLGEPPGFLHSWWCVFDWLYCAAPERRDT-CEHSSEAKAFHDYGFVNSGYGVNGTIG---
Dm ITLGEPPGFLHTWWCVFDWLYCAAPERRDT-CDHSSEAKAFHDYGFVNSGYGVNGTIGPGG
Hs ITLGEPPGFLHSWWCVFDWLYCAAPERRDT-CEHSSEAKAFHDYAAAAAPSPVLGNI---
Mm ITLGEPPGFLHSWWCVFDWLYCAAPERRDT-CEHSSEAKAFHDYAAAAAPSPVLGNI---
XI ITLGEPPGFLHSWWCVFDWLYCAAPERRDT-CEHSSEAKAFHDYAAAAAPSPVMGNNM---
Dr ITLGEPPGFLHSWWCVFDWLYCAAPERRDT-CEHSSEAKAFHDYAAAAAPSPVLGNNM---
Ce STIKLSDKSFLLFWLLFWDLYSAAPERRDAGGDPFSAEAKYFHEAMIGMPPGNGHFAP-

Bm -HNAG-PAPP-----NDG--MGGGG-MPPGFFPNSPLRSPAPHPGSGQSPDH--GP
Dm PHNAGGPAPSPLGQMPPGDGGPMGPGGPGNFFPNSTMRSPPT-HASS-----P
Hs -----PP-----NDG--MPGGPIPPGFFQGGPGSGQSPDHAQPPPHNPSS-MMG
Mm -----PP-----NDG--MPGGPIPPGFFQ
XI -----PP-----NDG--IPGGPMPPGFFQGGPGSGQSPDHAQPPPHNPNNPMMG
Dr -----PP-----GDG--MPGGPMPPGFFQ
Ce -----PPMEMMMGGHPG--AFGGRFAPGRMPPGAMADGGMPDGAFFMFPD-----

Bm QPQPMPGVGRGAWSGGGAGAAFLNYSGGSPGAYGPPSGAVRMGMGNDFNGPPGQGMGN
Dm QPQPSPQMP-----QPP--FMGGPRYPGGPRPGVRMGMGNEFNGPPGQPMMPN
Hs PHSQPFMSPR-----YAGGPR--PD--IRMGNOPPGGVPGTQPLLPN
Mm ---PFMSPR-----YAGGPR--PD--IRMGNOPPGGVPGTQPLLPN
XI PHSQPFMSPR-----YPGGPR--PS--LRMPNOPPVGVPGTQPLLPN
Dr ---PFMSPR-----FGGPR--PD--IRMGNOPPGGVPAAGPMLPN
Ce -PRLQRMADN-----QGMRMPPPVGGQPFPGAVGMRPVGGAPMDMS

Bm SME-RGGSGA--AGLLGPRMTTPR-PG-MGPMSPGAYAAAMRGPPPPQGP--GMADMGMG-
Dm SMDPTRPGGG--MGPMNPRMNPGRPGGMGPMGYGG-PGGMRGPAPGP--GMPMGMG
Hs SMDPTRQGGHPNMGSGMORMNPPRGMPGPG-PONYGSGMRPPNSLG-PAMPGINMGP
Mm SMDPTRQGGHPNMGSGMORMNPPRGMPGPG-PONYGSGMRPPNSLG-PAMPGINMGP
XI SMDPTRQGGHPNMGSGMORMNPPRGMGAG--PONYGSGMRPPNSLGGPGMPG-NMGP
Dr -MDP-----LQGMORMNVPGRMGMPGPG-PQGFGGMRPPHNSMG-PGMPGVNMGP
Ce GQRFDFMGGPPPGGAQPFPGASGSGGMVPMNGAHPHMSLNSPSMGVPPADMPFFMGMP

Bm -----PPTP-IMPSPQD-----SSNS
Dm AGGRPPQWQPNASAPLNAYSSSPGNYGPGSNGPPGPTP-IMPSPQD-----NTQG
Hs GAGRPP--WPNPNSANSIPYSSSSPGTYVPPGGGGGPGTP-IMPSPAD-----STNS
Mm GAGRPP--WPNPNSANSIPYSSSSPGTYVPPGGGGGPGTP-IMPSPAD-----STNS
XI GGRGP--WPNPN-ANSIAYSSSPGNYVPPGGGGGPGTP-IMPSPQD-----STNS
Dr GNRGRP-WPNPN-ANNMPYSSPSGAYGGPGGGG-PGTPGIVPSPAD-----SNNS
Ce MPPTSSAMPFGMSSDHQPMASAGAAAAPGATTAGGPTGGMIGSVPGPGSVQVATTSV

Bm GLGGVDG-----MK-----SPGGVG-----
Dm GPVGGPQDSMYALMKPEFPMGGGPDGGGGGGPGGMGPMGGGPNSMGVNLGGGGPDGSG
Hs S-----DNIYTMINPVPPGGSRSNFPMPGPGSDGPMGGMGG-----
Mm S-----DNIYTMINPVPPGGSRSNFPMPGPGSDGPMGGMGG-----
XI S-----ENMYTMNPIGPGGSRSNFPMPGPGSDGPMGGMGA-----
Dr S-----ENLYTMIN--SGGGGRNFPITGPGSEGLGAMAG-----
Ce GSVGTPSSIGQQLHQPKQEITTNGEIIMKTEALTPTGGGGGSVPPPPP-----

Bm -----GGPGTPREDSGSGMGDYNLS-FGGPGGQSDQTESAAILKIKESMEEAK
Dm LDGMKNSPANGGPGTPREDSGSGMGDYNLGGFGGPG--ENDOTESAAILKIKESMEEAK
Hs --MEPHHMNGLSGSDIDLKPKNSPNNISGISNPPGTTPRDDGELGGNF-----LH
Mm --MEPHHMNGLSGSDIDLKPKNSPNNISGISNPPGTTPRDDGELGGNF-----LH
XI --MEPHHMNGLSGSDMDVLKPKNSPNNMAGMNNPPGTTPRDDGELGGNF-----LN
Dr --MDPMHMNGLSGSDIDLKPKNSPNNISGISNPPGTTPRDE-DVGGSY-----LH
Ce --ATAAVSMNGGPGSAPGSAHSVNNVNPCTPGSNPLSNPMNPPLESSGPPPPGSND

Bm RFEKDP-DHPDYFMQ
Dm RFEKDT-DHPDYFMP
Hs SFQNDN-YSPSMTMSV
Mm SFQNDN-YSPSMTMSV
XI PFOSES-YSPSMTMSV
Dr SFQENQYSPSMTMSV
Ce AGKDDNGEISKIREGLLDGFGA

```

D

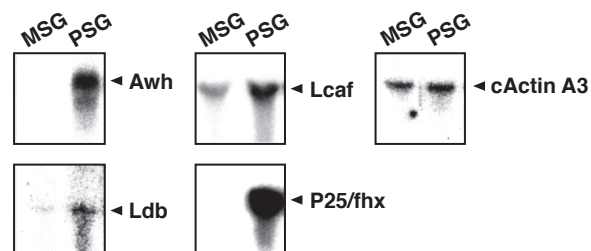


Figure 3, K.Ohno

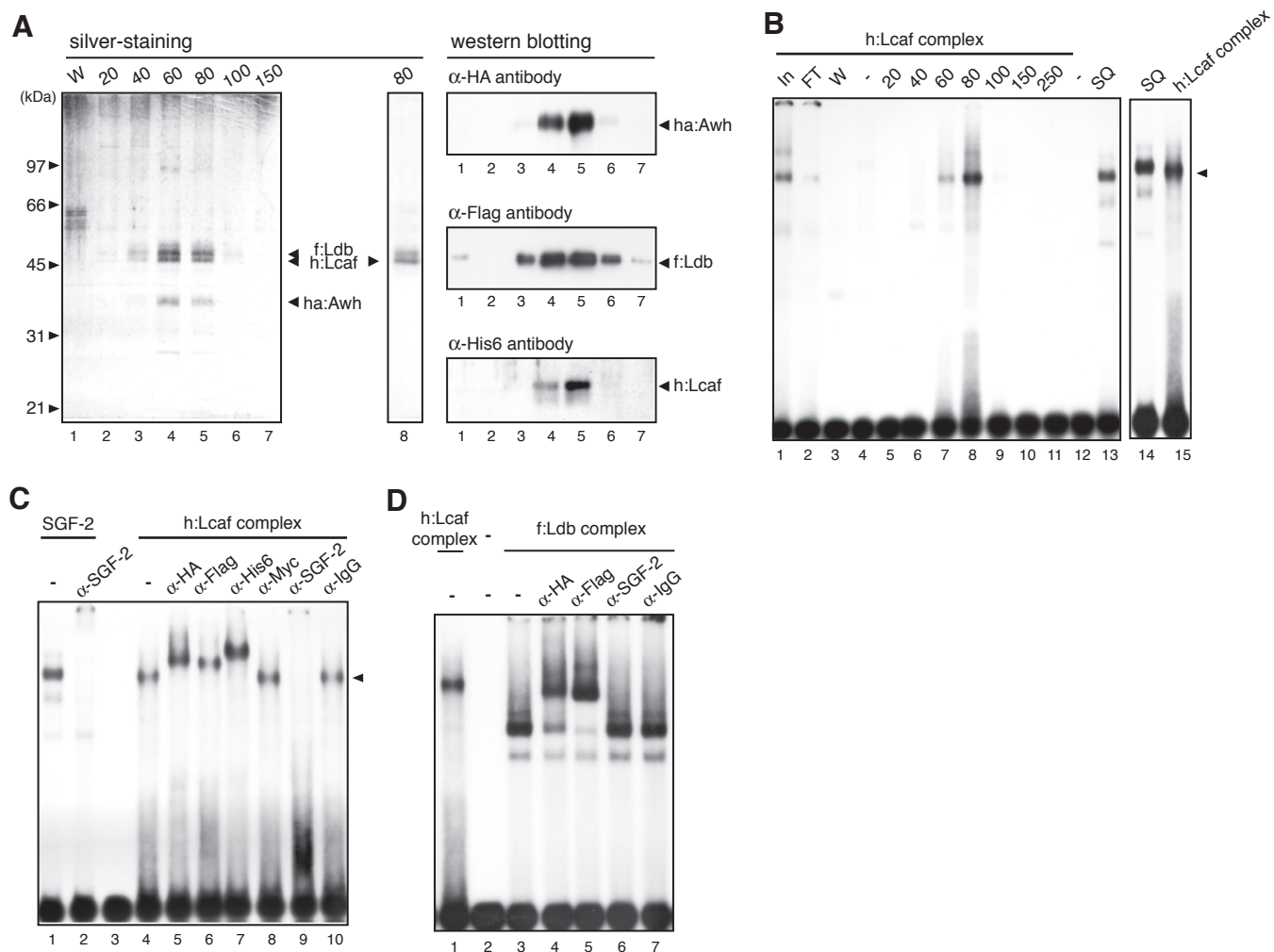


Figure 4, K. Ohno

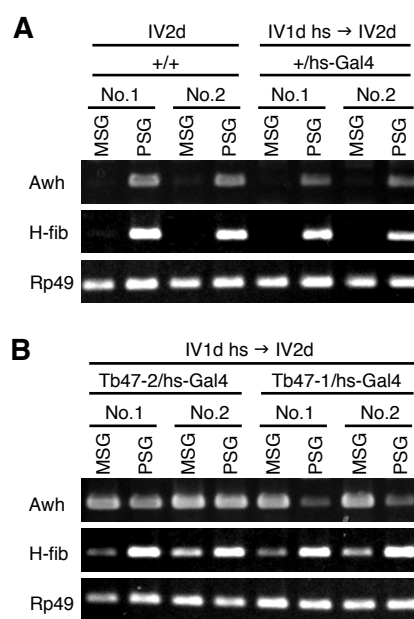


Figure 5, K.Ohno

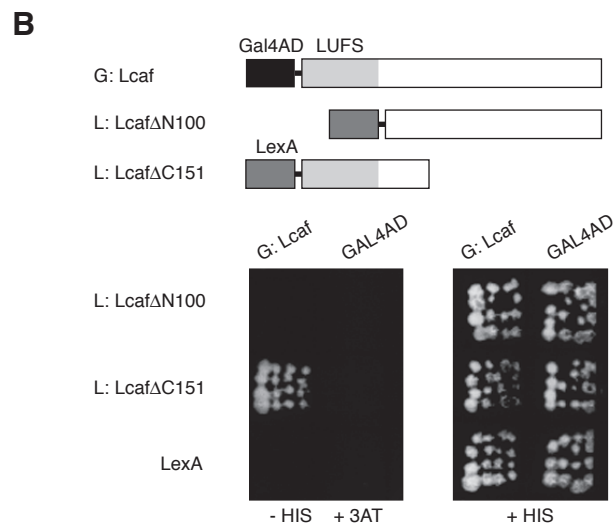
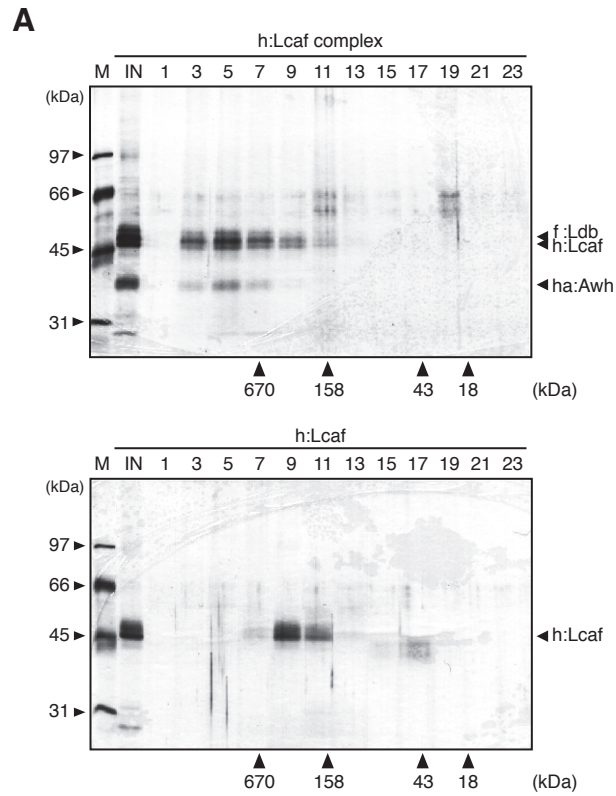


Figure 6, K.Ohno

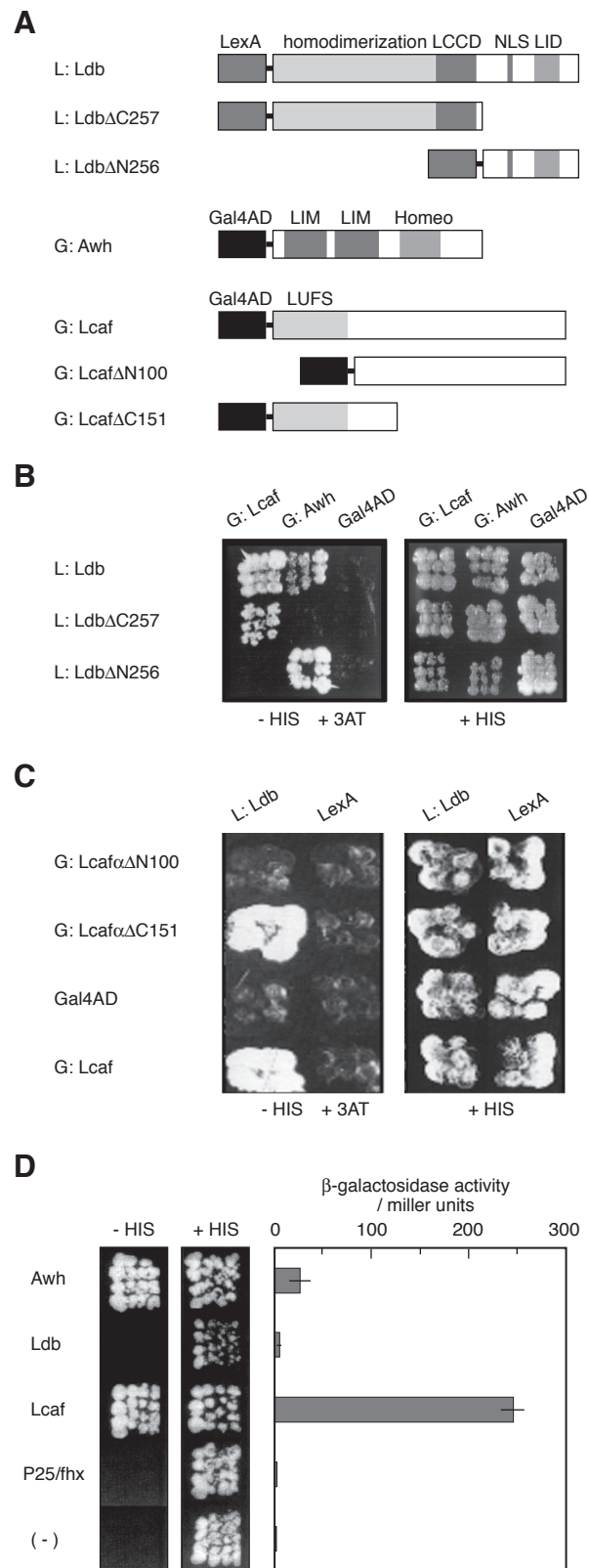


Figure 7, K.Ohno