Supplementary Information for

Crystal structure of human LDB1 in complex with SSBP2

Hongyang Wang[1,2,3], Juhyun Kim[4], Zhizhi Wang[2], Xiao-Xue Yan[1,*], Ann Dean[4,*], Wenqing Xu[2,*]

1. National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

2. Department of Biological Structure, University of Washington School of Medicine, Seattle, WA 98195, USA

3. College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

4. Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

*Corresponding authors: Wenqing Xu: wxu@uw.edu, Phone: +1 (206)221-5609; current address: School of Life Science and Technology, ShanghaiTech University, 201210, Shanghai, China. email: xuwq2@shanghaitech.edu.cn; Ann Dean: ann.dean@nih.gov; Xiao-Xue Yan: snow@ibp.ac.cn

This PDF file includes:

Supplementary Materials and Methods;

Supplementary Figures S1 to S8;

Supplementary Tables S1 to S3;

Supplementary References.

www.pnas.org/cgi/doi/10.1073/pnas.1914181117

Supplementary Materials and Methods

Protein preparation

His-MBP-LDB1(56-285) and His-SSBP2(1-94) were co-expressed with pETDuet-1 vector including N-terminal Tobacco Etch virus (TEV) protease sites for removal of tags. The various LDB1 constructs were subcloned into pMAL vectors with an N-terminal 6×His tag. *Escherichia coli* strain BL21 (DE3) was used for protein overexpression. Overexpression of the above proteins was induced by 0.2 mM isopropyl β-D-thiogalactoside (IPTG) when cell density reached OD_{600} ~ 0.8. After induction at 18 °C for 16-18 h, the cells were collected and homogenized in lysis buffer containing 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 20 mM imidazole pH 8.0, 2 mM DTT.

For LDB1/SSBP2 purification (*SI Appendix*, Fig. S2 *B-D*), after sonicated, the cell lysate was cleared by centrifugation (26,000 g for 1 h). The supernatant was collected and loaded onto Ni-NTA affinity resin (Qiagen) and washed with lysis buffer. The complex was eluted with buffer containing 50 mM Tris-HCl pH 8.5, 100 mM NaCl, 300 mM imidazole pH 8.0, 2 mM DTT. After cleavage of His-MBP tag and His tag at 4 °C overnight, the complex was diluted by five folds with buffer containing 50 mM Tris-HCl pH 8.5, 2 mM DTT and then loaded onto a HiTrap Q column (GE Healthcare) equilibrated with buffer A (20 mM Tris-HCl pH 8.5, 20 mM NaCl, 2 mM DTT) and the complex was eluted with a 0–100% gradient of 20 mM Tris-HCl pH 8.0, 1 M NaCl, 2 mM DTT in 20 column volumes. The complex contained fractions were collected and concentrated and further purified by gel filtration (Superdex 200 10/300 GL; GE Healthcare) equilibrated with a buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM DTT. For obtaining good ratio LDB1/SSBP2 complex samples, the complex containing fractions of the first SEC run was collected and reloaded onto a Superdex 75 10/300 GL size exclusion column (GE Healthcare) to remove excess SSBP2 proteins. After concentrated to ~ 2 mg/mL, the LDB1/SSBP2 complex was stored in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM DTT, flash-frozen in liquid nitrogen, and stored at -80 °C until use.

For the purification of LDB1 constructs, after sonication the cell debris was removed by centrifugation for 1 h at 26000 g. The supernatant was loaded onto Ni-NTA affinity resin and washed with lysis buffer. The LDB1 proteins were eluted with lysis buffer containing 300 mM imidazole pH 8.0. After being concentrated to 15 mg/mL, LDB1 was further purified by a Superdex 200 10/300 GL size exclusion

column (GE Healthcare) equilibrated with a buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM DTT. The aggregation and dimer or monomer fractions of LDB1 were collected.

Selenomethionine substitute of LDB1/SSBP2 complex was expressed using autoinduction method (1) and purified the same as the wild-type LDB1/SSBP2 complex described above. The LDB1/SSBP2 complex mutants and LDB1 mutants were generated with a standard PCR-based strategy and were subcloned, overexpressed and purified in the same way as the wild-type proteins except that the steps of TEV cleavage and HiTrap Q were skipped.

Crystallization of LDB1/SSBP2 complex

The LDB1/SSBP2 complex crystals were grown at room temperature by hanging drop vapor diffusion by mix 1 μl of the protein solution and 1 μl of solution containing 100 mM lithium sulfate monohydrate, 100 mM sodium citrate tribasic dihydrate pH 5.6, 1% v/v PEG400, 10 mM DTT. Crystals were cryoprotected with 10% v/v PEG 400 and 15% v/v ethylene glycol by using a quick-soak/flash-freeze method.

Data collection and structure determination

A 2.8 Å data set of its selenomethionine-substituted LDB1/SSBP2 complex crystal was collected from the beamline 821 of the Advanced Light Source at 0.9792 Å. The data set was indexed, integrated and scaled using HKL2000 (2). The complex structure was determined by the single-wavelength anomalous dispersion (SAD) method. The Se sites and the initial phases were determined by PHENIX (3). The crystal belongs to $P6_522$ space group and there are one LDB1 molecule and two SSBP2 molecules in one asymmetric unit. The complex model was improved using iterative cycles of manual rebuilding with program Coot (4) and refinement using REFMAC (5) and PHENIX. All structural model figures were generated using PyMOL (6). The data collection and refinement statistics are summarized in *SI Appendix*, Table S1.

Size exclusion chromatography and SEC-MALS analyses

Size exclusion chromatography (SEC) was performed at room temperature on a Superdex 200 10/300 GL gel filtration column (GE Healthcare) in running buffer 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 15 mM beta-mercaptoethanol at 0.5 mL/min, and the protein elution was monitored by UV absorbance at 280 nm.

The molecular weights of wild-type or mutants of LDB1/SSBP2 complexes were measured using SEC-MALS. Briefly, 50 μg of protein complex was loaded onto Superdex 200 10/300 GL column (GE Healthcare) in a buffer of 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 15 mM beta-mercaptoethanol at a flow rate of 0.5 mL/min. The light scattering was monitored using DAWN HELEOS II system and the concentration was measured by Optilab rEX differential refractometer (Wyatt Technology). The data was processed with ASTRA 6.1 (Wyatt Technology) with a dn/dc value of 0.185 mL/g.

MBP pull-down assays

Human SSBP2 constructs were expressed and purified as descried (7). 20 μM of SSBP2 fragments, 5 μM of wild-type MBP-LDB1 fragments (or MBP-LDB1 mutants) and 20 μl dextrin sepharose high performance resin (GE Healthcare) were mixed in 100 μl of pull-down buffer containing 50 mM Tris-HCl pH 8.0, 100 mM NaCl, and 10 mM DTT. The mixed samples were incubated at room temperature for 2 h, followed by washing the resin with pull-down buffer three times. During each wash, 200 μl of pull-down buffer was added to each sample and incubated at room temperature for 5 min before centrifugation and removal of supernatant. After washing, the resin was boiled and analyzed by SDS-PAGE with Coomassie blue staining.

SPR assays

SPR experiments were performed using a BIAcore T100 machine (Biacore, GE Healthcare) in running buffer containing 20 mM HEPES pH 8.0, 100 mM NaCl, 2 mM DTT, 0.05% v/v Tween20 at 16 °C.The purified wild-type MBP-LDB1(56-287) was directly immobilized onto CM5 senor chips (Biacore, GE Healthcare). Subsequently, gradient concentrations (6.25, 12.5, 25, 50, 100, 200 nM) of various SSBP2 fragments were used to flow over the chip surface. In the same manner, the purified SSBP2(1-94) was directly immobilized onto CM5 senor chips (Biacore, GE Healthcare). Subsequently, gradient concentrations (6.25, 12.5, 25, 50, 100, 200 nM) of MBP-LDB1(56-287) or MBP-LDB1(56-287) mut5, or the gradient concentrations (100, 200, 400, 800, 1600, 3200 nM) of MBP-LDB1(56-287) mut3, MBP-LDB1(56-287) mut4, and His-MBP were used to flow over the chip surface. We also generate a blank control for each experimental group in consideration of nonspecific binding between flowing phase and CM5 senor chip. Gradient concentrations of MBP-LDB1 WT or mutants or SSBP2 fragments were used

to flow over the blank chip surface (blank group) or corresponding immobilized chip surface (experimental group). The 10 mM Glycine buffer (pH 2.0) was used to regenerate the chip. The final binding signal response is calculated by subtracting the signal of blank group from that of its corresponding experiment group. The binding affinities were analyzed using steady state affinity with the software BIAevalution Version 4.1.

Sedimentation velocity experiments

Sedimentation velocity assays were performed on a Beckman ProteomeLab XL-I analytical ultracentrifuge (AUC) at 20 °C. All protein samples were diluted with lysis buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 15 mM beta-mercaptoethanol) to 400 µl at A_{280nm} absorption of ~0.7. Data were collected at 45,000 rpm at a wavelength of 280 nm and analysed using the SEDFIT software program (www.analyticalultracentrifugation.com).

Cell culture

Mouse erythroid leukemia (MEL) cells were cultured in DMEM with 10% v/v fetal bovine serum in a humidified incubator at 5% CO2. MEL cell differentiation was induced with 1.5% v/v DMSO for 4 days.

Gene editing using CRISPR/Cas9

For CRISPR/Cas9 editing to delete LDB1, guide RNAs (gRNAs) with the lowest off-target hit scores were selected (http://crispr.mit.edu/) (*SI Appendix*, Table S3). gRNAs targeting exons 4 and 8 of Ldb1 were cloned into CRISPR-Cas9 and gRNA expression vector pSpCas9(BB)-2A-GFP (PX458) (gift from Feng Zhang, Addgene plasmid #48138) as described (8). MEL cells were electroporated with BTXPRESS electroporation solution (BTX) according to the manufacturer's instructions. Fluorescent cells were sorted by fluorescence-activated cell sorting (FACS) ARIA II (BD Biosciences), and individual clones were grown in 96-well plates. Deletions were validated by immunoblotting (*SI Appendix,* Fig. S8).

Western blotting

Cells were lysed with M-PER Mammalian Protein Extraction Reagent (Thermo Scientific) and protein concentration determined using BCA protein assay kit (Thermo Scientific). Samples were separated by

NuPAGE gel, transferred to PVDF membranes according to the manufacturer's instructions, and probed with antibodies LDB1 (AB96799; Abcam), GFP (AB183734; Abcam), and Tubulin (AB7291; Abcam), respectively. Blots were developed by ECL (Thermo Scientific).

Virus production and transduction

LDB1 and mutants followed by T2A GFP were constructed in a lentiviral vector and validated by Sanger sequencing. 293FT cells were transduced with lentiviral vector and packaging vectors (gift from Didier Trono, Addgene plasmid #12259 and #12260). Supernatants were harvested at 48 h and 72 h after transfection and concentrated with Amicon Ultra-15 (Merck Millipore). For stable clones, MEL cells were spinosulated with viral particles at 2,400 rpm in the presence of 8 mg/mL polybrene and fluorescent cells were sorted by fluorescence-activated cell sorting (FACS) ARIA II (BD Biosciences), and individual clones were grown in 96-well plates.

RT-qPCR

RNA was isolated from 1x106 MEL cells with the RNeasy Plus kit (Qiagen). RNA (1 μg) was reverse transcribed by using the High Capacity CDNA RT (Life Technology) according to the manufacturer's instructions. RT-qPCR was performed using the SYBR Green Supermix (Bio-Rad) with the ABI 7900HT (Applied Biosystems). Data was normalized to GAPDH. Primers are listed in *SI Appendix*, Table S3.

Chromosome conformation capture assay

3C was performed as described using BglII as the restriction enzyme (9). Relative crosslinking between the anchor fragment and fragments of interest was analyzed using SYBR green chemistry (SYBR Green Supermix) with the ABI 7900HT (Applied Biosystems). Interaction between 2 fragments within the a-tubulin gene was used as the internal normalization control. Primers are listed in *SI Appendix*, Table S3.

Supplementary Figures

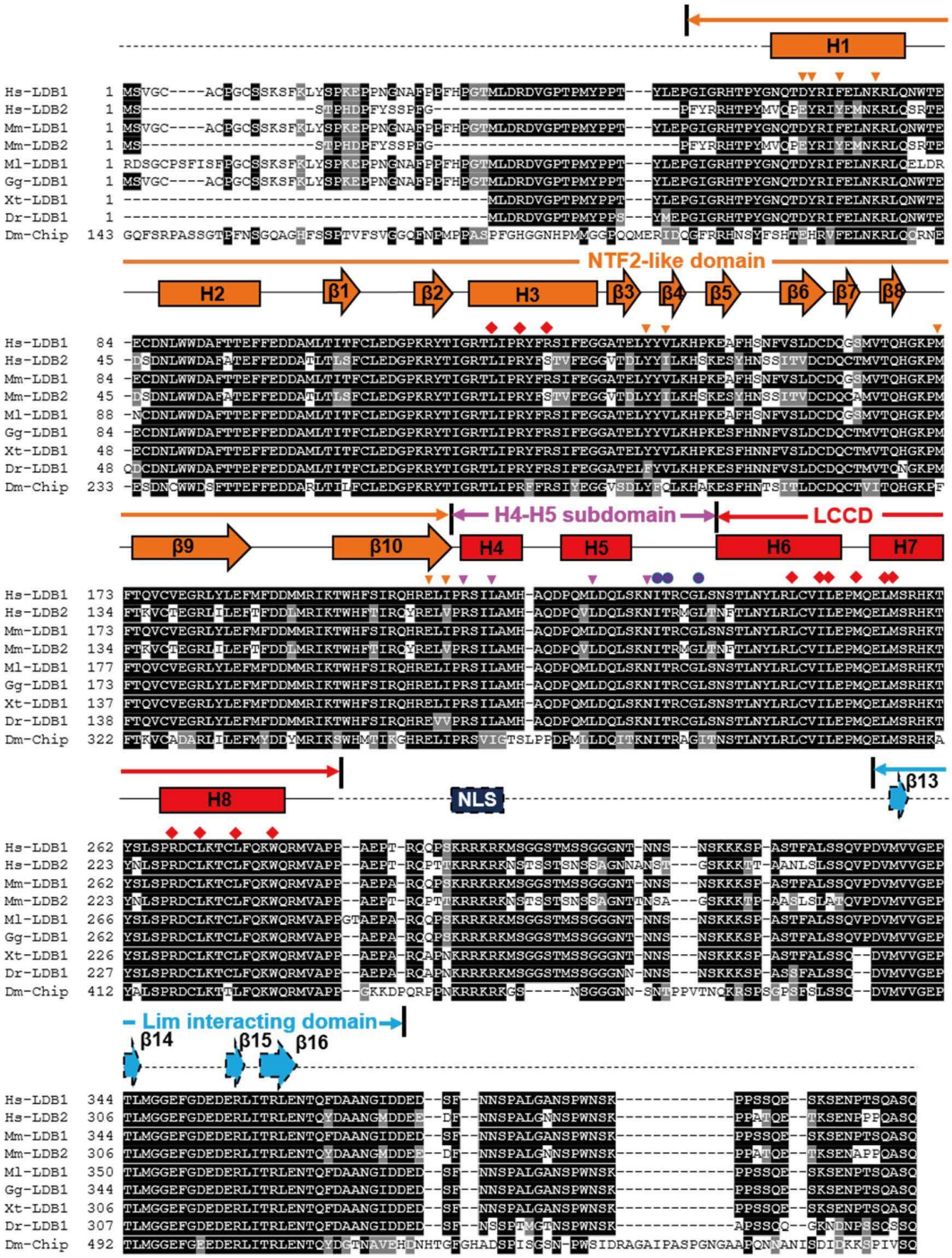

Fig. S1. Sequence alignment of LDB1 proteins from 7 different species using the T-Coffee server (10). Secondary structural elements and domain composition of LDB1 are indicated above the alignment. LDB1 residues that mediate the dimer formation are identified by inverted triangles, while residues involved in direct interaction with SSBP2 are indicated with diamonds. Residues that covered the potential ligand binding pocket in NTF2-like subdomain are indicated in dots. The regions determined in our structure are indicated in solid lines, other parts are showed in dash lines. Abbreviations: Hs, *Homo sapiens*; Mm, *Mus musculus*; Ml, *Myotis lucifugus*; Gg, *Gallus gallus*; Xt, *Xenopus tropicalis*; Dr, *Danio rerio*; Dm, *Drosophila melanogaster*.

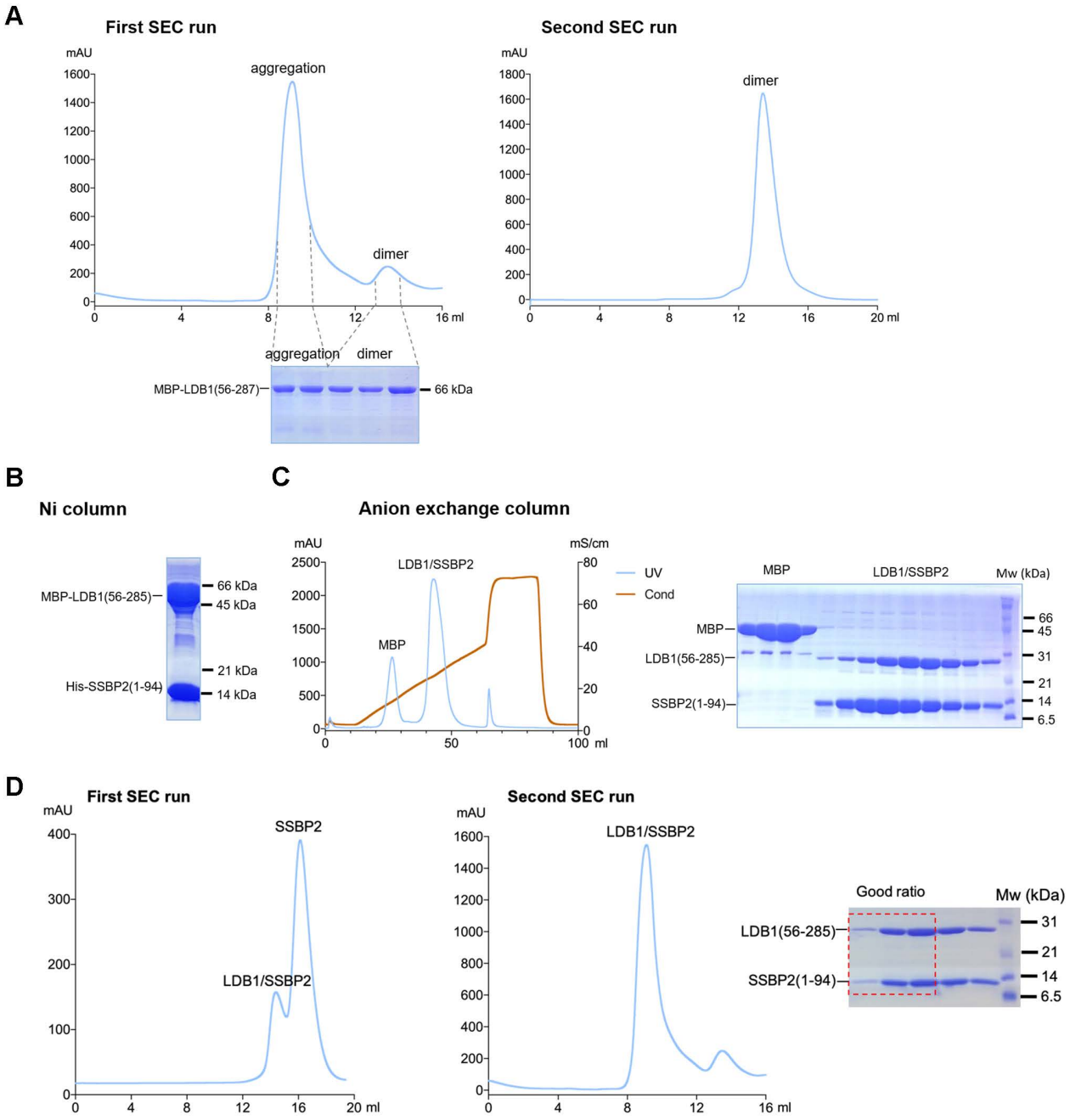

Fig. S2. Purification of LDB1 and the LDB1/SSBP2 complex. (**A**) Purification of MBP tagged human LDB1(56-287) using a Superdex 200 10/300 GL size exclusion column. The dimer fractions were concentrated and used for the second SEC run. (**B**) Co-expression of His-MBP-tev-LDB1(56-285) and His-tev-SSBP2(1-94) using the pETDuet-1 vector. (**C, D**) Further purification of the human LDB1(56-285)/SSBP2(1-94) complex by anion exchange chromatography and SEC. The LDB1/SSBP2 complex fractions were concentrated and used for the second SEC run.

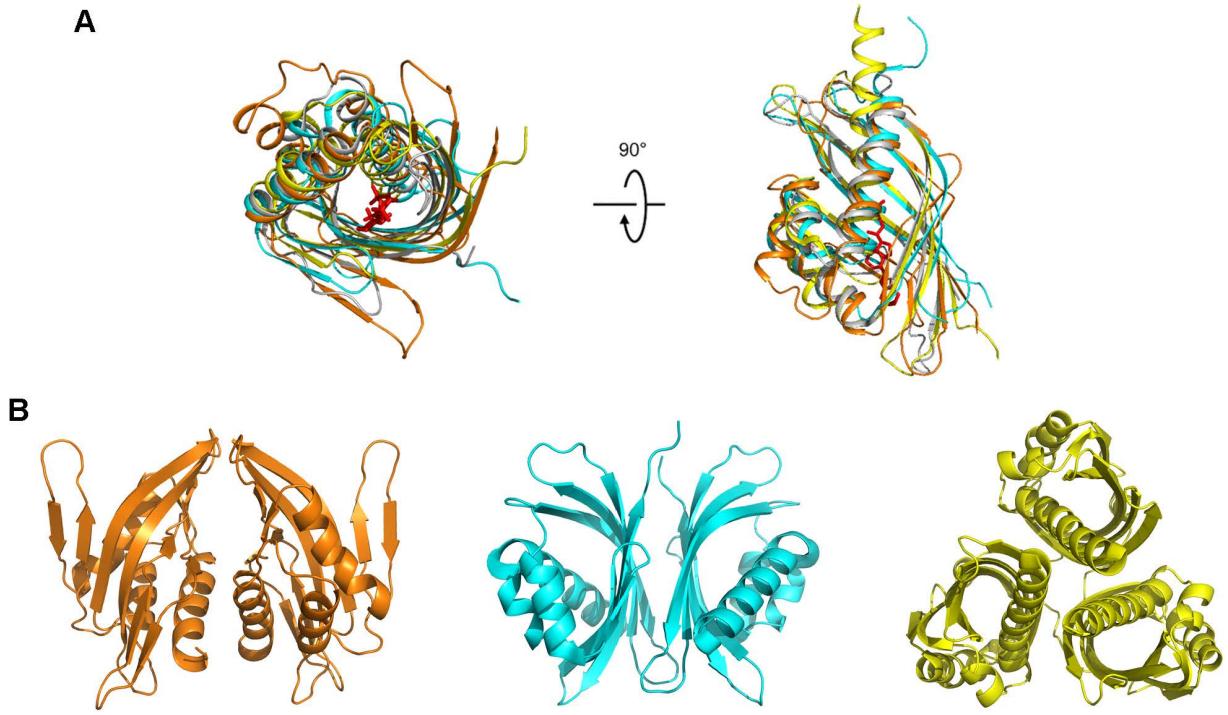

Fig. S3. Comparison of the LDB1 NTF2-like subdomain with other NTF2 family proteins, and oligomerization modes. (**A**) Two orthogonal views of structural superposition of different NTF2-like domains: these of LDB1 (in orange), LinA (in yellow, PDB code: 3A76), CDL2.2, a computationally designed Vitamin-D3 binder (in gray, PDB code: 5IEN), and NTF2 (in cyan, PDB code: 1U5O). The ligand from 5IEN is colored in red. (**B)** The oligomerization modes of different NTF2-like domains: LDB1 dimer (in orange), NTF2 dimer (in cyan) and LinA trimer (in yellow).

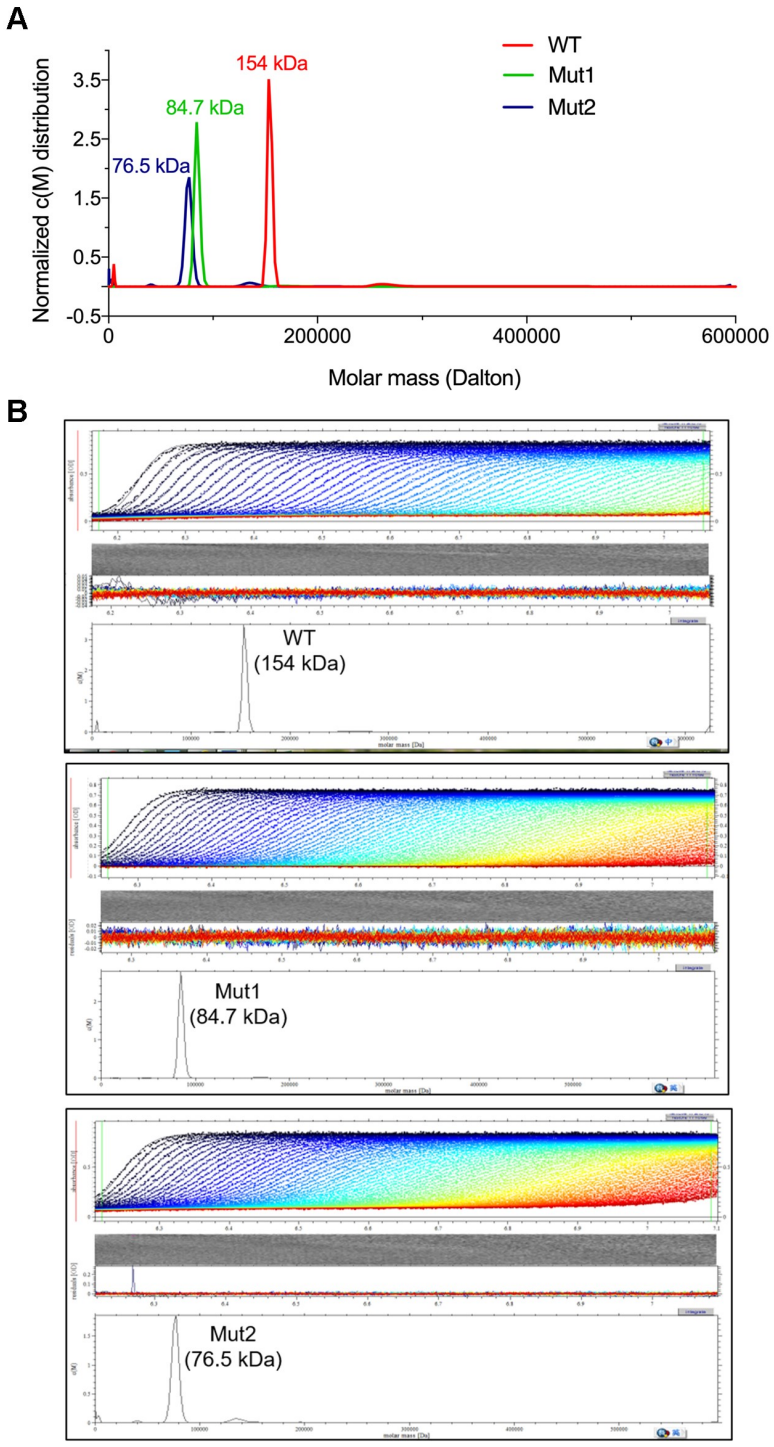

Fig. S4. (**A** and **B**) Mutagenesis analysis of the LDB1 dimerization interface using analytical ultracentrifugation (AUC). Sedimentation velocity experiments indicated that the wild-type LDB1/SSBP2 complex consists of two LDB1 and four SSBP2 molecules, while LDB1 mut1 and mut2 monomers form complexes with two SSBP2 molecules, respectively. These AUC results are consistent with SEC-MALS analysis results.

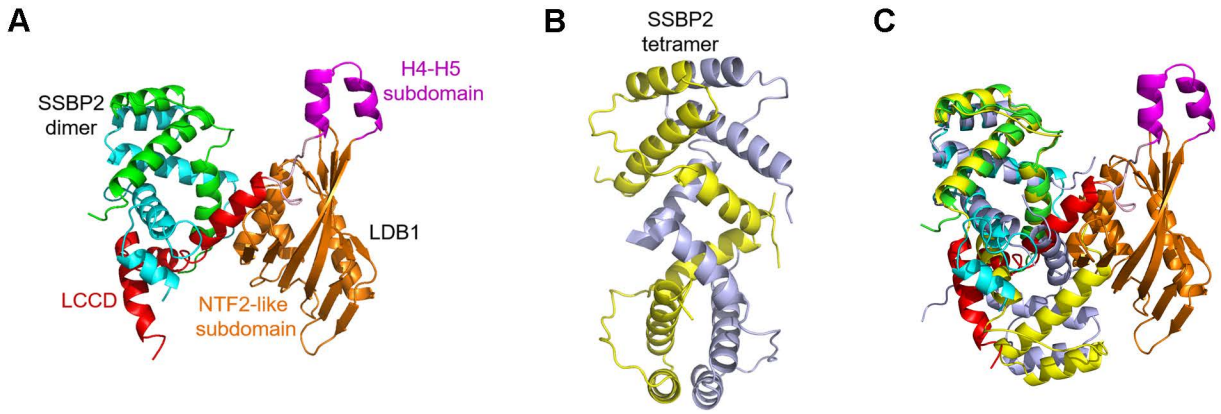

Fig. S5. Comparison of SSBP2 dimer structures in the SSBP2 homotetramer and the LDB1/SSBP2 complex. (**A**) Structure of half of the LDB1/SSBP2 complex. The two SSBP2 molecules in the SSBP2 dimer are colored in green and cyan, respectively. (**B**) Structure of the SSBP2 homotetramer (PDB code: 6IWV), in which the two SSBP2 molecules in both dimers are colored in yellow and light blue, respectively. (**C**) Structural superposition of SSBP2 dimers in the SSBP2 homotetramer and the LDB1/SSBP2 complex.

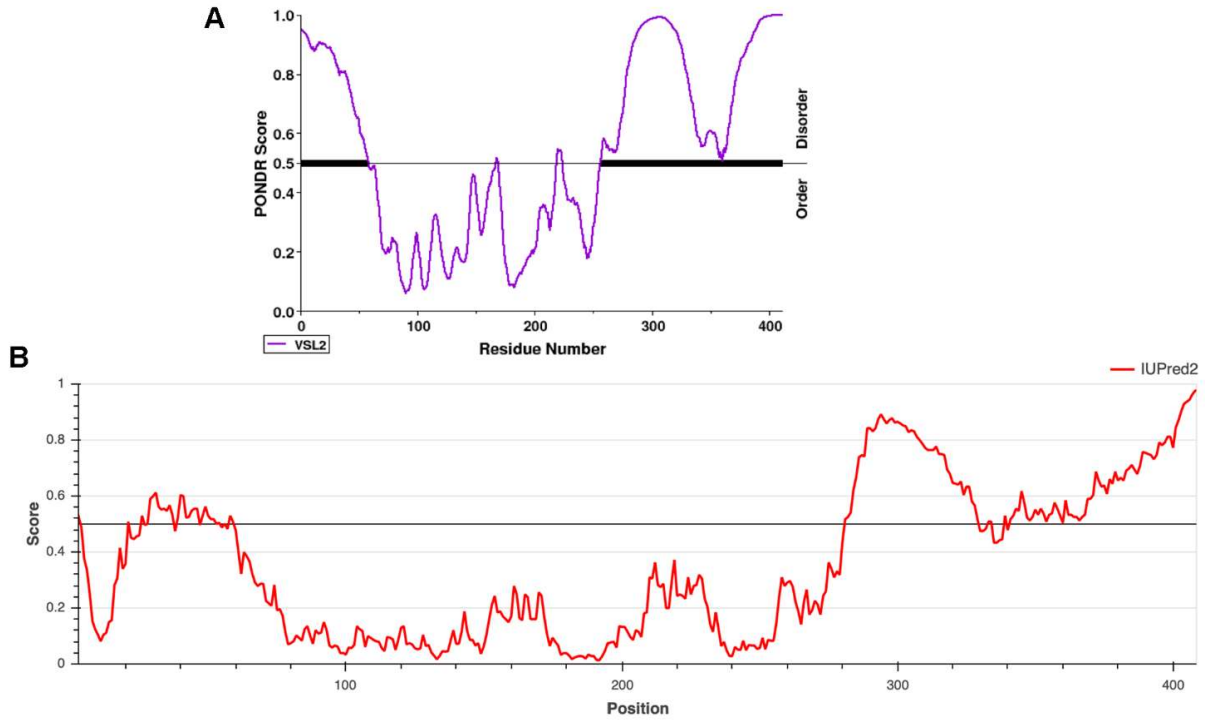

Fig. S6. The protein disorder tendency prediction of the full-length human LDB1, using: (**A**) the VSL2 server (http://www.dabi.temple.edu/disprot/predictorVSL2.php) and (**B**) the IUPred2 server (https://iupred2a.elte.hu). The X-axis represents residues 1-411 of the full-length human LDB1 protein, whereas the Y-axis is the disorder tendency score for each residue in the context of the LDB1 sequence. The regions with disorder tendency higher than 0.5 have a high tendency to be structurally disordered. The LDB1 DD-LCCD region is the only folded region predicted by both servers.

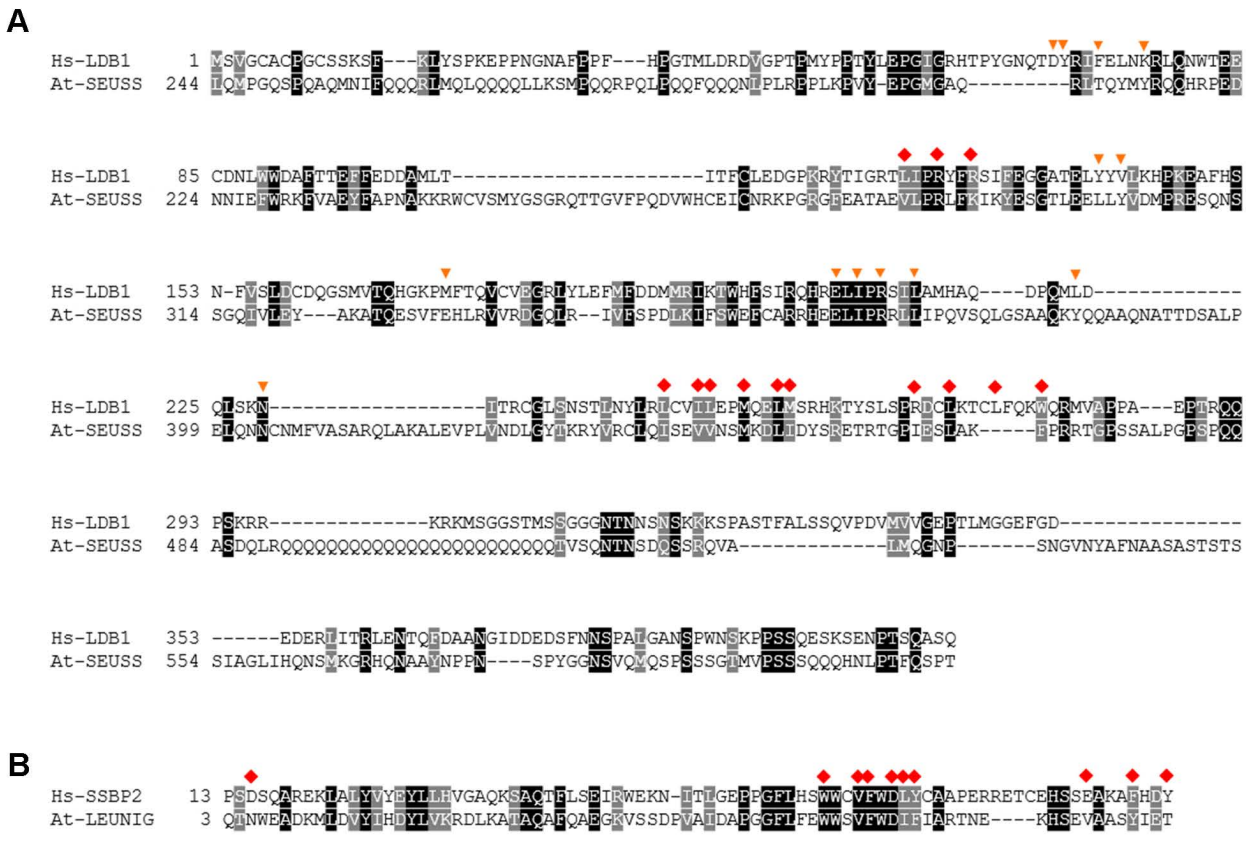

Fig. S7. Key residues in the LDB1 dimerization interface and LDB1/SSBP2 interface are conserved in SEUSS and LEUNIG. (**A**) Sequence alignment of human LDB1 and *Arabidopsis* SEUSS. (**B**) Sequence alignment of human SSBP2 and *Arabidopsis* LEUNIG. Residues that mediate the dimer formation are identified by inverted triangles, and LDB1 residues involved in direct interaction with SSBP2 and *vice versa* are indicated with diamonds. Abbreviations: Hs, *Homo sapiens*; At, *Arabidopsis thaliana*.

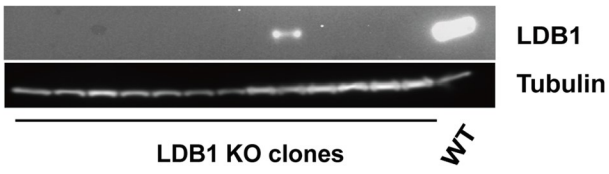

Fig. S8. Western blots with LDB1 antibodies of LDB1 knock out MEL cells and wild-type MEL cells. Deletion of Ldb1 was carried out using CRISPR/Cas9.

Supplementary Tables

Table S1. Statistics for X-ray crystallographic data collection and refinement

	LDB1/SSBP2 complex
Data collection	
Space group	$P6_522$
Cell dimensions	
a, b, c (Å)	104.3, 104.3, 250.4
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.9792
Resolution (Å)	50.0–2.80 (2.90–2.80)
R_{sym}	0.080 (1.180)
R_{meas}	0.086 (1.313)
R_{pim}	0.033 (0.555)
I / σI	38.7 (1.3)
$CC_{1/2}$	0.991 (0.513)
Completeness (%)	99.8 (99.4)
Redundancy	6.7 (5.3)
Refinement	
Resolution (Å)	43.80–2.80
No. reflections	20532
R_{work}/R_{free}[a,b] (%)	24.0/28.1
No. atoms	3079
Average B factor (Å^2)	
Protein	114.60
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.848
Ramachandran plot (%)	
Most Favorable	94.52
Allowed	5.48
Outliers	0

Values in parentheses are for highest-resolution shell.

[a] $R_{work} = \Sigma_{hkl} |F_o(hkl) - F_c(hkl)|/\Sigma_{hkl} Fo(hkl)$.

[b] R_{free} was calculated for a test set of reflections (5%) omitted from the refinement.

Table S2. Structural comparisons of LDB1 NTF2-Like domain with structurally similar proteins using DALI Search

No.	Protein	PDB ID	Z-score	RMSD	Sequence identity (%)	Ligand	Oligomeric state
1	Putative scytalone dehydratase	3EF8	13.9	2.3	7	--	Trimer
2	CDL2.2	5IEN	13.9	2.9	8	VDY	Dimer
3	A putative aromatic ring hydroxylase	3B8L	13.3	2.8	10	Glycerol	Trimer
4	Uncharacterized conserved protein	3ROB	13.2	2.5	5	Glycerol	Dimer
5	A NTF2-like protein	1TP6	12.8	2.8	11	--	Dimer
6	NTF2-like superfamily protein	3SOY	12.7	2.7	11	Imidazole	Dimer
7	Steroid Delta-isomerase	3T8N	12.5	2.7	11	--	Dimer
8	Gamma-hexachlorocyclohexane dehydrochlorinase LinA	3A76	12.4	3.0	9	Spermidine	Trimer
9	A putative dehydratase from the NTF2-like family	3CNX	12.4	2.4	10	--	Dimer
10	Protein SgcJ	4OVM	12.4	2.5	12	PG4	Dimer
11	Nuclear transport factor 2	1U5O	12.2	2.3	9	--	Dimer
12	A NTF2-like protein	3B7C	12.1	2.4	10	EDO	Dimer
13	Ca/calmodulin-dependent kinase ii association domain	3KSP	12.1	3.0	5	NHE, EDO	Dimer
14	OHP9_1c	5IF6	12.1	2.9	7	3QZ	Dimer
15	Ras GTPase-activating protein-binding protein 1	5FW5	12.1	2.6	10	--	Dimer

VDY: 25-hydroxyvitamin D3; PG4: Tetraethylene glycol; EDO: 1,2-Ethanediol; NHE, N-cyclohexyltaurine; 3QZ: (9beta)-17-hydroxypregn-4-ene-3,20-dione.

Table S3. gRNAs and Primer information

Name	Sequence
Target sequence used for guide RNA expression	
LDB1 sg 1 F	CACC TGGAATTGGACAGCCCACAC
LDB1 sg 1 R	AAAC GTGTGGGCTGTCCAATTCCA
LDB1 sg 4 F	CACC ACGGCTACAGAACTGGACAG
LDB1 sg 4 R	AAAC CTGTCCAGTTCTGTAGCCGT
Primers used for qPCR analysis	
B maj F	ACTAAACCCCCTTTCCTGCTC
B maj R	TTGCAGTGAACATAAATGCTT
GAPDH F	AAGGGCTCATGACCACAGTC
GAPDH R	CAGGGATGATGTTCTGGGCA
Primers used for chromatin conformation capture analysis (3C)	
HS2_anchor	CAGGCTTTTAGTTGGATATAGAGTGAA
b-maj	GGCTGGAACATCACTGGAATAAAT
rbH1	CCCATGTTACACCCCATTACAAG
ercc F	GCAGCCACCGACTTGGAT
ercc R	GCAGTGAAAACACAACACAGTTAATATG

Supplementary References

1. H. K. Sreenath *et al.*, Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr. Purif.* **40**, 256-267 (2005).
2. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326 (1997).
3. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
4. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
5. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242 (2011).
6. W. L. DeLano, A. T. Brunger, Helix packing in proteins: prediction and energetic analysis of dimeric, trimeric, and tetrameric GCN4 coiled coil structures. *Proteins* **20**, 105-123 (1994).
7. H. Wang, Z. Wang, Q. Tang, X. X. Yan, W. Xu, Crystal structure of the LUFS domain of human single-stranded DNA binding Protein 2 (SSBP2). *Protein Sci.* **28**, 788-793 (2019).
8. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281-2308 (2013).
9. I. Krivega, R. K. Dale, A. Dean, Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes Dev.* **28**, 1278-1290 (2014).
10. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217 (2000).