Structural and functional studies on epigenetic regulators

Jan Kadlec

Structural and functional studies on epigenetic regulators

Habilitation à diriger des recherches

Spécialité: Biologie

Jan KADLEC

UNIVERSITE GRENOBLE ALPES

2017

Date de soutenance : 9 Mars 2017

ECOLE DOCTORALE DE CHIMIE ET SCIENCES DU VIVANT

UNIVERSITE GRENOBLE ALPES

HABILITATION A DIRIGER DES RECHERCHES

Jan KADLEC

Title: Structural and functional studies on epigenetic regulators

Jury:

Prof. Marc Jamin	Rapporteur
Dr. Christophe Romier	Rapporteur
Dr. Sébastien Fribourg	Rapporteur
Prof. Thomas Schalch	Examinateur
Dr. André Verdel	Examinateur

Institut de Biologie Structurale, Grenoble

Summary

This manuscript describes my research activities over the last fifteen years. I graduated in Molecular Biology from the Charles University in Prague in 2000. At that time, structural biology was very poorly developed in my home country. During my training at EMBL in Grenoble I discovered for the first time the world of protein structures and realized how important detailed structural information can be to understand protein function. I then joined the laboratory of Stephen Cusack at the EMBL and began my PhD jointly with the Université Joseph Fourier. I started using X-ray crystallography and studied proteins involved in targeted mRNA degradation pathway called the Nonsense Mediated mRNA Decay.

For my postdoc I decided to change the research filed and joined the group of David Stuart at the University of Oxford to work on viral fusion proteins. The preparation of these viral surface glycoproteins for crystallization required their expression in mammalian cells and an efficient deglycosylation. The scientific, instrumentation and technological environment at the department was truly exceptional so it was possible to succeed in these projects in a relatively short time. The structure of the baculovirus fusion protein GP64 we determined was perhaps the most revealing structure I have worked on so far and was a significant contribution to the viral fusion field.

While working in Oxford was a great experience I decided to return to France and was offered a staff scientist position back at EMBL Grenoble in my original PhD laboratory. I was very fortunate to be given a maximal scientific independence and support within the group. I established a new research line in the group focusing on epigenetic regulatory complexes. Chromatin/epigenetic control of DNA-based processes is an extremely important and sophisticated layer of regulation. Perhaps, due to the challenging nature of chromatin regulatory complexes for structural analysis, most of them still remain structurally uncharacterised. I started two very fruitful collaborations with expert cell biologist in the chromatin field, Asifa Akhtar (MPI Freiburg) and Bernard de Massy (IGH Montpellier). Together with them we can carry out an interdisciplinary research combining biochemistry, structural and cell biology and genetics. All my work on the histone acetyltransferase complexes (MSL and NSL complex) and PRDM9 at the EMBL included supervision of Master and PhD students.

In 2014, I obtained the ATIP/AVENIR start-up grant and moved to the Institut de Biologie Structurale (IBS) in Grenoble where I set up my research team. We continue our effort to structurally and functionally characterise the MSL and NSL complexes as well as the PRDM9 methyltransferase. Given the large size of these complexes, the method of choice for the analysis will be a combination of X-ray crystallography and cryo-electron microscopy.

Table of contents

I. Curriculum vitae 4

II.List of publications 7

III. Scientific supervision 9

IV. Previous research activities

Thesis work at EMBL Grenoble (2000-2005) 10

Postdoctoral work at the University of Oxford (2005-2007) 28

V. Current Research

EMBL Grenoble (2008-2015) and IBS Grenoble (since 2015) 38

- PRDM histone methyltransferases 39

- Mof acetyltransferase containing complexes 50

MSL complex 50

NSL complex 76

I. Curriculum Vitae

- **PERSONAL INFORMATION**

Name: KADLEC Jan
http://orcid.org/0000-0002-4853-6918
Date of birth: 12.05.1976
Nationality: Czech
Married, 2 children
Address: Institut de Biologie Structurale (IBS), 71 avenue des Martyrs, CS 10090, 38044 Grenoble Cedex 9, France
Tel: 0457428776
E-Mail: jan.kadlec@ibs.fr
Website: http://www.ibs.fr/research/research-groups/ebev/epigenetic-regulators

- **EDUCATION**

2001 - 2005 PhD in Structural Biology
Joint PhD degree at the European Molecular Biology Laboratory (EMBL)
and Université Joseph Fourier in Grenoble, France.

1998 - 2000 Master degree in Molecular Biology, Faculty of Science,
Charles University in Prague, Czech Republic.

1995 - 1998 Bachelor degree in Biology, Faculty of Science,
Charles University in Prague, Czech Republic.

- **CURRENT POSITION**

Since 2015 Team leader - Institut de Biologie Structurale (IBS), Grenoble, France
Epigenetic Regulators team
Since 2016 Permanent CR1 researcher – INSERM

- **PROFFESIONAL EXPERIENCE**

2009 - 2015 Staff Scientist at EMBL - Grenoble Outstation (with Stephen Cusack)
Structural analysis of epigenetic regulators.

2008 Postdoctoral fellow at EMBL - Grenoble Outstation, France (with Stephen Cusack).

2005 - 2007 Postdoctoral fellow - The Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, UK.
Structural analysis of viral fusion proteins (with David Stuart).

2001 - 2005 PhD thesis at EMBL - Grenoble Outstation, France. Structural analysis of Nonsense-Mediated mRNA Decay regulators (with Stephen Cusack).

2000 - 2001 Training at EMBL - Grenoble Outstation, France. Structural analysis of the SNARE complex. (with Winfried Weissenhorn).

1998 - 2000 Undergraduate training in the Laboratory of Physiology and Genetics of Actinomycetes at Institute of Microbiology in Prague. Biosynthetic pathways of lincosamid antibiotics (with Jiri Janata)

- **GRANTS AND FELLOWSHIPS**

2016 - Labex Gral - Individual projects (120k EUR)
2015 - Fondation ARC - Projet Fondation ARC (50k EUR)
2014 - ATIP/AVENIR start-up grant - CNRS/Plan Cancer (360k EUR+ salary)

2006 - Marie Curie Intra-European postdoctoral fellowship (awarded but declined)
2005 - EMBO long-term postdoctoral fellowship
2001 - Boehringer Ingelheim Fonds PhD scholarship (awarded but declined)
2000 - Fellowship of TEMPRA PECO program of Rhone-Alpes Region, France

- **TEACHING ACTIVITIES**

Since 2011 UJF Grenoble, Master 2 course on structural biology - lecture on experimental phasing.
2012 - 2013 EMBL Predoc course - lecture on "Readers and writers of the epigenetic landscape".

- **ORGANISATION OF SCIENTIFIC MEETINGS**

2013, 2015 Co-organiser of the "Signalling Through Chromatin" international symposium in Grenoble
Since 2012 Co-organiser of regular meetings of the Chromatin, Transcription and Epigenetics community of Grenoble.

- **INSTITUTIONAL RESPONSIBILITIES**

Since 2016 Co-responsible of the Molecular Medicine research axis at the IBS
2009 - 2015 Faculty member - EMBL
2009 - 2015 Chairing lab committee meetings of the EMBL Grenoble

- **OTHER SCIENTIFIC ACTIVITIES**

2011 - 2016 Editor of the online protocol database of the FP7 EU Network of Excellence 'Epigenesys'.
2009 - 2013 Provide support to beamline users of the European Synchrotron Radiation Facility (ESRF)

- **COMMISSIONS OF TRUST**

2016 External Grant Evaluator, European Research Council (ERC)
2014 External Grant Evaluator, National Science Center, Poland.
Since 2012 Reviewer for PLoS ONE, RNA Biology, Journal of Structural Biology.

- **MAJOR COLLABORATIONS**

Asifa Akhtar - Max Planck Institute of Immunobiologie and Epigenetics, Freiburg, Germany.
Functional analysis of MOF-containing complexes
Bernard de Massy - Institut de Génétique Humaine, Montpellier, France.
Functional analysis of PRDM9.

- **PRESENTATIONS**

- Central European Institute of Technology (CEITEC), Brno, Czech Republic, 2016, invited speaker
- Instruct Workshop on Molecular interactions, Grenoble, 2015, invited speaker
- Epigenesys Kick-Off meeting, Paris, France, 2010, invited speaker
- VIZIER/Spine2Complexes Joint Workshop on Structural Virology, Vienna, Austria, 2008, invited speaker
- 24th European Crystallographic meeting, Marrakech, Morocco, 2007, invited speaker
- Structure, Function and Dynamics of RNA-protein Complexes, Gottingen, Germany, 2003, selected speaker

- **SELECTED PUBLICATIONS**

[*] Corresponding author

Dias, J., Nguyen, N., Georgiev, P., Gaub, A., Brettschneider, J., Cusack, S., **Kadlec, J.**[*] and Akhtar, A.[*] Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex.
Genes Dev. 28: 929-942 (2014)

Wu, H., Mathioudakis, N., Diagouraga, B., Dong A., Dombrovski, L., Baudat. F., Cusack, S., de Massy, B.[*] and **Kadlec, J.**[*] Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. **Cell Rep.** 5:13-20 (2013)

Hallacli, E., Lipp, M., Georgiev, P., Spielman, C., Cusack, S., Akhtar, A.[*] and **Kadlec, J**.[*] MSL1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in Drosophila. **Mol. Cell** 48:587-600 (2012)

Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S[*]. and Akhtar, A.[*] The molecular basis for the recruitment of MOF and MSL3 into the dosage compensation complex by MSL1. **Nat. Struct. Mol. Biol.** 18:142-149 (2011)

Kadlec, J., Loureiro, S., Abrescia, N.G.A., Stuart, D.I. and Jones, I.M. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. **Nat. Struct. Mol. Biol.** 15: 1024-30 (2008)

Kadlec, J., Izaurralde, E. and Cusack, S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. **Nat. Struct. Mol. Biol.** 11, 330-337 (2004)

II. List of publications

* Corresponding author
Supervised students are underlined

20. El Omari, K., Iourin, O., **Kadlec, J**., Harlos, K., Sutton, G., Grimes, J.M., and Stuart, D.I. Unexpected structure for the N-terminal domain of Hepatitis C virus envelope glycoprotein E1. **Nat. Commun.** 5:4874 doi: 10.1038/ncomms5874 (2014)

19. El Omari, K., Iourin, O., **Kadlec,J**., Fearn, R., Hall, D.R., Harlos, K., Grimes, J.M., and Stuart, D.I. Pushing the limits of Sulfur-SAD phasing, de novo structure solution of the N-terminal domain of the ectodomain of HCV E1. **Acta Crystallogr. D Biol. Crystallogr.** 70:2197-2203 (2014)

18. Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Couté, Y., Conn, S., **Kadlec, J**., Sachidanandam, R., Kaksonen, M., Cusack, S., Ephrussi A., and Pillai, R.S. RNA clamping by Vasa assembles a piRNA Amplifier complex on transposon transcripts. **Cell**, 157:1698-711 (2014)

17. **Dias, J.,** Nguyen, N., Georgiev, P., Gaub, A., **Brettschneider, J.,** Cusack, S., **Kadlec, J.**[*] and Akhtar, A.[*] Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex.
Genes Dev. 28: 929-942 (2014)

16. Guilligay, D., **Kadlec, J**., Crepin, T., Lunardi,T., Bouvier, D., Kochs, G., Ruigrok, R. and Cusack, S., Comparative structural and functional analysis of orthomyxovirus polymerase cap-snatching domains. **PLOS One** 9:e84973 (2014)

15. Wu, H., Mathioudakis, N., Diagouraga, B., Dong A., Dombrovski, L., Baudat. F., Cusack, S., de Massy, B.[*] and **Kadlec, J.**[*] Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. **Cell Rep.** 5:13-20 (2013)

14. Najmanova, L., Kutejova, E., **Kadlec, J**., Polan, M., Olsovska, J., Benada, O., Novotna, J., Kamenik, Z., Halada, P., Bauer, J., and Janata., J Characterization of *N*-demethyllincosamide methyltransferases LmbJ and CcbJ. **ChemBioChem** 14:2259-62 (2013)

13. Iourin, O., Harlos, K., El Omari, K., Lu, W., **Kadlec, J**., Iqbal, M., Meier, C., Palmer A., Jones, I., Thomas, C., Brownlie, J., Grimes, J.M., and Stuart, D.I Expression, purification and crystallization of the ectodomain. **Crystallogr. Sect. F Struct. Biol. Cryst. Commun.** 69:35-38 (2013)

12. Hallacli, E.,[#] **Lipp, M.,**[#] Georgiev, P., Spielman, C., Cusack, S., Akhtar, A.[*] and **Kadlec, J**.[*] MSL1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in Drosophila. **Mol. Cell** 48:587-600 (2012)

11. O'Flynn, N.M.J., Patel, A., **Kadlec, J.** and Jones I.M. Improving promiscuous mammalian cell entry by the baculovirus AcMNPV. **Biosci. Rep.** 33:23-36 (2012)

10 Mathioudakis, N., Palencia, A., **Kadlec, J**., Round, A., Tripsianes, K., Sattler, M., Pillai, R.S. and Cusack, S. The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. **RNA** 18:2056-72 (2012)

9. **Kadlec, J**., Hallacli, E., **Lipp, M.,** Holz, H., Sanchez-Weatherby, J., Cusack, S.* and Akhtar, A.* Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. **Nat. Struct. Mol. Biol.** 18:142-149 (2011)

8. Raja, S.J., Charapitsa, I., Conrad, T., Vaquerizas, J.M., Gebhardt, P., Holz, H., **Kadlec, J**., Fraterman, S., Luscombe, N.M. and Akhtar, A. The nonspecific lethal complex is a transcriptional regulator in Drosophila. **Mol. Cell** 38:827-41 (2010)

7. Sanchez-Weatherby, J., Bowler, M.W., Huet, J., Gobbo, A., Felisaz, F., Lavault, B., Moya, R., **Kadlec, J**., Ravelli, R.B. and Cipriani, F. Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. **Acta Crystallogr. D Biol. Crystallogr.** 65:1237-46. (2009)

6. Clerici M., Mourão, A., Gutsche, I., Gehring, N.H, Hentze, M., Kulozik A., **Kadlec J.**, Sattler M., and Cusack S. Unusual bipartite mode of interaction between the nonsense mediated decay factors UPF1 and UPF2. **EMBO J.** 28:2293-2306 (2009).

5. **Kadlec, J.**, Loureiro, S., Abrescia, N.G.A., Stuart, D.I. and Jones, I.M. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. **Nat. Struct. Mol. Biol.** 15: 1024-30 (2008)

4. Walter, T.S. Mancini, E.J., **Kadlec, J**., Grahem, S.C., Assenberg, R., Ren, J., Sainsbury, S., Owens, R.J., Stuart, D.I., Grimes, J.M. and Harlos, K. Semi-automated microseeding of nanolitre crystallization experiments. **Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.** 64, 14-8 (2008)

3. **Kadlec, J.**, Guilligay, D., Ravelli, R.B. and Cusack, S. Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. **RNA** 12, 1817–1824 (2006)

2. **Kadlec, J.**, Izaurralde, E. and Cusack, S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. **Nat. Struct. Mol. Biol.** 11, 330-337 (2004)

1. Bracher, A., **Kadlec, J.**, Betz, H. and Weissenhorn, W. X-ray structure of a neuronal complexin-SNARE complex from squid. **J. Biol. Chem.** 277, 26517-26523 (2002)

III. Scientific supervision

BTS student (Bachelor degree):
2013 Laika Bastel - ETSL Paris (at EMBL)

Master students:
2016 Seif Eddine Cheghib – University of St. Etienne, France (at IBS)
2013 Janine Brettschneider - Mittweida University of Applied Sciences, Germany (at EMBL)
co-author of Dias *et al*, Genes Dev. 28: 929-942 (2014)
2011 Ganeshapandian Pitchai - Uppsala University, Sweden
2009 Michael Lipp - Ludwig - Maximilians University in Munich, Germany (at EMBL)
co-author of Kadlec *et al*, Nat. Struct. Mol. Biol. 18:142-149 (2011)

PhD students:
2012-2015 Jorge Dias - EMBL/University of Heidelberg
first author of Dias *et al*, Genes Dev. 28: 929-942 (2014)
2009-2010 Michael Lipp - EMBL/UJF Grenoble
first co-author of Hallacli *et al*., Mol. Cell 48:587-600 (2012)

Postdoctoral fellows:
Since 2016 Shasha Shi (at IBS)
2015-2016 Rong Chen (at IBS)
2015-2016 Jorge Dias (at IBS)

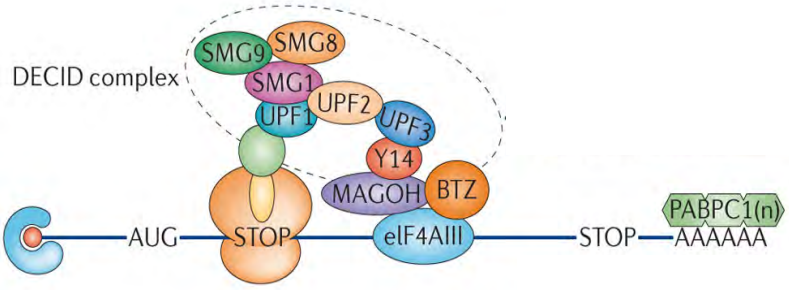
Engineer:
Since 2016 Anne-Emmanuelle Foucher (at IBS, ITA CNRS)

IV. Previous research activities

Thesis work at EMBL Grenoble (2000-2005)

Structural studies on proteins involved in Nonsense-Mediated mRNA Decay

I obtained my PhD in structural biology jointly from the European Molecular Biology Laboratory (EMBL) in Grenoble and Université Joseph Fourier. The aim of my PhD project in Stephen Cusack's group at EMBL was to obtain structural information on several proteins involved in nonsense-mediated mRNA decay (NMD). NMD is a surveillance mechanism by which eukaryotic cells detect and degrade transcripts containing premature termination codons that would otherwise be translated into truncated, potentially dangerous proteins (Kervestin and Jacobson, 2012) (Figure 1). NMD is involved in a large number of genetic disorders. When the project was initiated, the molecular mechanisms underlying this pathway were unknown. The main question was how the cell can distinguish between the genuine and premature stop codons and how it selectively eliminates defective transcripts. Three 'up-frameshift' proteins, UPF1, UPF2 and UPF3 had been shown to be essential for this process in yeast and human. The goal of my PhD work was to structurally characterize these proteins and their mutual interactions.



from Kervestin & Jacobson - Nat. Rev. Mol.Cell Biol. 2012

Figure 1. Premature translational termination leads to retention of the downstream UPF3-containing exon-exon junction complex (EJC) on the transcript, which facilitates interaction of UPF1 with UPF2, leading to the formation of the DECID (decay inducing) complex and subsequent mRNA degradation.

First, we were able to produce and purify a complex between an MIF4G (middle portion of eIF4G) domain of UPF2 and an RNP domain (ribonucleoprotein-type RNA-binding domain) of UPF3b and determine its crystal structure (Kadlec et al., 2004). The protein-protein interface is mediated by highly conserved charged residues in UPF2 and UPF3b and involves the beta-sheet surface of the UPF3b RNP domain. Interestingly, RNP domains were believed to only use corresponding surfaces to bind nucleic acids. This was one of the first examples showing that RNP domains can in fact interact with both nucleic acids and protein partners with their beta-sheet surface (Fribourg et al., 2003; Kadlec et al., 2004). In addition, in collaboration with the group of Elisa Izaurralde we showed that the UPF3b RNP does not bind RNA, whereas the UPF2 construct and the complex do. Using site directed mutagenesis we also identified several residues that are essential for the UPF2/UPF3 interaction as well as for the UPF2 binding of RNA.

We also determined a structure of the highly conserved cysteine–histidine-rich domain of human UPF1. The structure revealed that this domain is a unique combination of three zinc-binding motifs arranged into two tandem modules related to the RING-box and U-box domains of ubiquitin ligases. This UPF1 domain interacts with UPF2, and we identified by mutational analysis residues in

two distinct conserved surface regions of UPF1 that mediate this interaction (Kadlec et al., 2006). We could prepare a complex between the interacting domains of UPF1 and UPF2 and obtained crystals diffracting to a resolution of 3.1 Å. This structure was solved together with a structure of a complex between the full-length UPF1 and the interaction domain of UPF2 in a follow-up project by Marcello Clerici and revealed an unusual bipartite mode of interaction between the two UPF proteins confirming our previous mutagenesis data (Clerici et al., 2009). Our work provided first structural information on the NMD key factors and advanced our understanding of the mechanism of this mRNA degradation pathway. Since then, many new factors essential for NMD were identified and most of their interactions have been analysed by structural approaches resulting in very detailed understanding of NMD (Figure 1).

nature structural & molecular biology

The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3

Jan Kadlec[1], Elisa Izaurralde[2] & Stephen Cusack[1,2]

Nonsense-mediated mRNA decay (NMD) is a surveillance mechanism by which eukaryotic cells detect and degrade transcripts containing premature termination codons. Three 'up-frameshift' proteins, UPF1, UPF2 and UPF3, are essential for this process in organisms ranging from yeast to human. We present a crystal structure at a resolution of 1.95 Å of the complex between the interacting domains of human UPF2 and UPF3b, which are, respectively, a MIF4G (middle portion of eIF4G) domain and an RNP domain (ribonucleoprotein-type RNA-binding domain). The protein-protein interface is mediated by highly conserved charged residues in UPF2 and UPF3b and involves the β-sheet surface of the UPF3b RNP domain, which is generally used by these domains to bind nucleic acids. We show that the UPF3b RNP does not bind RNA, whereas the UPF2 construct and the complex do. Our results advance understanding of the molecular mechanisms underlying the NMD quality control process.

NMD is the process by which cells recognize and eliminate mRNAs containing premature translation termination codons (PTCs) originating from, for example, aberrant transcription, errors in splicing and frameshift or nonsense mutations in DNA. Thus, NMD prevents the expression of potentially dangerous truncated proteins (reviewed in ref. 1). Three conserved eukaryotic proteins, UPF1, UPF2 and UPF3 (UPF, up-frameshift), have an essential role in NMD[1–4], although several other essential NMD proteins including Y14–Mago, Smg1 and Smg5–7 have been identified recently in different organisms[1,5–10].

According to the prevalent but still controversial model of NMD in mammalian cells, recognition of a PTC requires the splicing-dependent deposition of a multiprotein complex, the exon junction complex (EJC), 20–24 nucleotides upstream of a splice junction[11]. The EJC, which consists of at least six proteins, has been implicated in cytoplasmic mRNA localization as well as NMD (reviewed in ref. 1). During mRNA export into the cytoplasm, a perinuclear protein, UPF2, is recruited to the EJC by UPF3. In mammals, nonsense-mediated mRNA decay has been proposed to occur during the first round of translation[12]. With the first passage of a ribosome, the EJCs would normally be stripped from the mRNA. If, however, translation terminates at a PTC at least 50–55 nucleotides upstream of an EJC, UPF2 associated with the remaining EJC downstream can be bound by another NMD protein, UPF1, most likely in a complex with translation termination factors eRF1 and eRF3. This triggers mRNA degradation[12–15]. This does not happen in the case of normal stop codons because mammalian mRNAs rarely have introns in the 3′ untranslated region, and if they do occur, they are practically always <50 nucleotides downstream of the termination codon[16]. In *Saccharomyces cerevisiae*[17] and *Drosophila melanogaster*[7] PTCs are recognized by a different mechanism that does not involve the EJC. However in all organisms so far investigated, the three UPF proteins are essential for NMD[1–4,7].

Human UPF2 is a large perinuclear protein (1,272 residues) that interacts with both UPF1 and UPF3 (refs. 2–4,18). It is predicted to contain three phylogenetically conserved MIF4G domains (**Fig. 1a,b**)[19]. MIF4G domains comprise a single unit of five helical hairpins[20–22] and are found in several proteins involved in mRNA processing and maturation, including the translation initiation factor eIF4G[21] and the large subunit of the cap-binding complex CBP80 (ref. 22). The UPF3-binding site on UPF2 has been mapped to a region including the third MIF4G domain[4,23], whereas the UPF1 has been mapped to the N- and C-terminal regions of UPF2 (**Fig. 1a**)[4,24]. Additionally, UPF2 interacts with translation initiation factor eIF4AI and SuiI, a subunit of the translation initiation factor eIF3, again via a region including the third MIF4G domain[3].

Two homologous UPF3 proteins in human cells, UPF3a and UPF3b, are nucleocytoplasmic shuttling proteins with 452 and 470 residues, respectively[2,4]. They contain a conserved N-terminal region comprising a putative ribonucleoprotein domain (RNP, also known as RNA recognition motif, RRM) that lacks a clear consensus RNP2 motif[25] (**Fig. 1c**). RNP domains are commonly involved in RNA binding, but neither UPF3a nor UPF3b has been shown to bind directly to RNA. The UPF2-binding site on UPF3 has been mapped to the N-terminal region, that includes the RNP-like domain[4,23].

To understand the structural basis and functional importance of the UPF2-UPF3 interaction, we have solved the crystal structure at a resolution of 1.95 Å of the complex between their interacting domains, residues 761–1054 of hUPF2 and residues 42–143 of hUPF3b. According to the structure, the N-terminal region of UPF3b indeed folds like a canonical RNP domain, but, surprisingly, it binds to UPF2

[1]European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, BP 181, F-38042 Grenoble Cedex 9, France. [2]European Molecular Biology Laboratory, Gene Expression Programme, Meyerhofstrasse 1, 69117 Heidelberg, Germany. Correspondence should be addressed to S.C. (cusack@embl-grenoble.fr).

Published online 7 March 2004; doi:10.1038/nsmb741

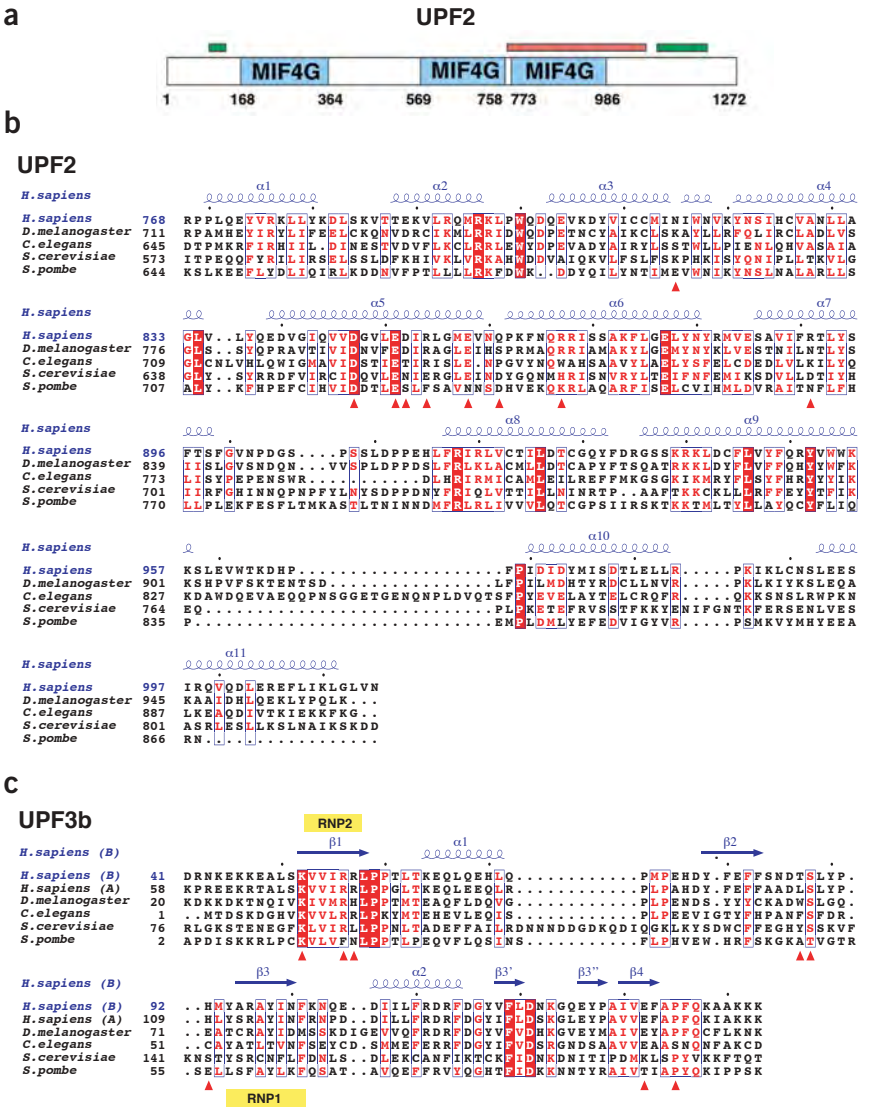


Figure 1 Human UPF2 domain organization and sequence alignment of UPF2 and UPF3b. (**a**) Schematic representation of human UPF2 domain structure and putative protein-binding sites. Three MIF4G domains identified by SMART[19] are blue. Red (residues 761–1072) and green (residues 94–133 and 1085–1194) bars represent putative UPF3- and UPF1-binding sites, respectively[4,23,24]. Translation initiation factors eIF4AI and Suil are proposed to interact with UPF2 within the region of residues 757–1272 (ref. 3). (**b**) Sequence alignment of UPF2 proteins. (**c**) Alignment of UPF3b proteins and human UPF3a. Residues that are 100% conserved are in red boxes. Similarity >70% is red. Red triangles indicate residues involved in polar interactions between UPF2 and UPF3b. The secondary structures of human UPF2 and UPF3b are blue (α, α-helix; β, β-strand). Figure was generated with CLUSTALX[44] and ESPript[45].

using its β-sheet surface, the major RNA-binding surface in many RNP domains. Consistent with this, we have found that the RNP domain of UPF3b does not bind RNA, but that, unexpectedly, the UPF2 construct and its complex with UPF3b do.

RESULTS

The third MIF4G domain of human UPF2 (residues 761–1054; **Fig. 1a**) and the N-terminal RNP-like domain of UPF3b (residues 42–143) were expressed separately in *Escherichia coli*. The protein complex formed from the purified UPF2 and UPF3b fragments was crystallized and the X-ray structure solved at a resolution of 1.95 Å using MAD (**Fig. 2**) with selenomethionine (SeMet)-substituted UPF2. The crystallographic asymmetric unit contains two UPF2–UPF3b heterodimers that are structurally very similar. Comparing all Cα positions between the two complexes, the r.m.s. deviation is 0.37 Å but one complex has systematically higher *B*-factors. The final model, refined to an R_{free} of 22.7% and an *R*-factor of 20.5%, includes residues 768–1015 of UPF2, residues 50–140 of UPF3b and 318 water molecules.

Structure of third MIF4G domain of UPF2

UPF2(761–1054) has a typical MIF4G fold composed of five antiparallel pairs of helices (**Fig. 2a**). An additional helix, α11, forms a C-terminal extension to the MIF4G domain and this interacts with noncrystallographic-symmetry (NCS)-related helix α11 in the crystal. No electron density was observed for the extreme C-terminal residues 1016–1054, suggesting that this region is mobile and/or unstructured in accordance with predictions from its highly hydrophilic and acidic sequence.

A database search of Dali[26] showed that the MIF4G domain of UPF2 is structurally very similar to those of human CBP80 (ref. 22) and eIF4G[21], the only two other such domains of known structure. The similarity is especially high for residues 26–243 of CBP80 (Dali *Z*-score 14.6; r.m.s. deviation 1.98 Å for 165 Cα; **Fig. 3a**) despite a very low sequence homology (11% identity). Structure-based comparisons of these domains do not reveal any strictly conserved residues important for the structure of these MIF4G domains. Rather, a particular pattern of hydrophobic residues defines the conserved interhelical packing. The loops between the helices show considerable variation in length and sequence. However most of the identified MIF4G domains[19] feature a well-conserved FIGEL motif (874-FLGEL-878 in the third MIF4G domain of UPF2). In UPF2, as in the other two known MIF4G structures, the hydrophobic residues of this motif are involved in internal interactions between helices, but the conserved acidic residue (Glu877 in UPF2) is solvent-exposed. In the first MIF4G domain of CBP80 (residues 26–243), where the FIGEL motif is not very well conserved, Asp133, corresponding to Glu877 of UPF2, is involved in binding a linker region that encircles the entire domain[22]. In UPF2, Glu877 constitutes part of a patch of conserved surface residues, which also includes residues Tyr879, Asn880, Tyr881, Tyr932, Lys941 and Tyr894 of helices α6–α9 (**Fig. 3b,c**). This UPF2 region, rich in exposed aromatic residues, might be involved in interactions with other proteins, nucleic acids or other parts of UPF2 (see Discussion). Apart from this and the UPF3-binding site, mapping of phylogenetically conserved,

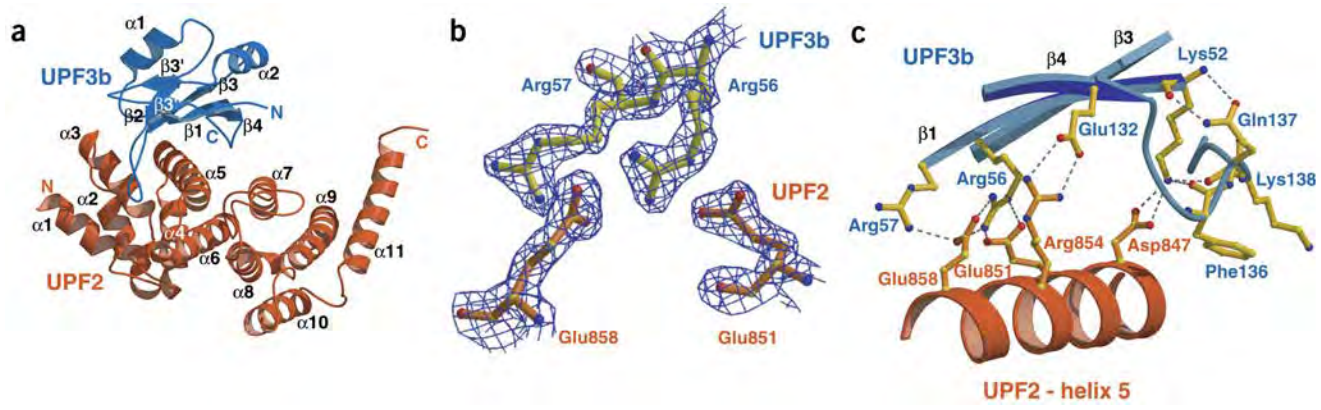


Figure 2 Crystal structure of the complex between UPF2 and UPF3b. (**a**) UPF3b(42–143) RNP domain (blue) interacts with helix 5 of UPF2(761–1054) (red) through its β-sheet. Secondary structures of both proteins are labeled. UPF2 MIF4G domain comprises helices 1–10. (**b**) A representative part of the $2F_0 - F_c$ electron density map contoured at 2.2 σ, corresponding to the principal UPF2-UPF3b interaction region (UPF2 residues, orange; UPF3b residues, yellow). (**c**) Schematic representation of principal interactions between UPF2 helix 5 (residues, orange; helix, red) and UPF3b β-sheet (residues, yellow; β-strands, blue) as well as Lys52 interactions with the UPF3b C-terminal loop (residues 134–140). These figures and ribbon diagrams in **Figure 3** were generated with MolScript[46] and BobScript[47].

solvent-accessible residues to the surface of UPF2(761–1054) reveals a basic patch formed by residues from helix α2 and the loop between α2 and α3, including Arg793, Arg796 and Lys797. To estimate the degree of residue conservation, the sequence of *Schizosaccharomyces pombe* Upf2 was excluded owing to its much lower similarity.

Structure of UPF3b N-terminal domain

The human UPF3b fragment we crystallized (residues 42–143) overlaps with the most conserved region of UPF3b, which approximately spans residues 40–230 (**Fig. 1c**). The fragment has a canonical RNP fold with a characteristic $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ arrangement (**Fig. 2a**), despite its lack of the consensus RNP2 motif, which made domain assignment originally uncertain. In addition, a loop following α_2 contains two short β-strands, β3′ and β3″. The UPF3b RNP domain is most similar to the first RNP of Sex lethal protein[27] (Dali[26] *Z*-score 9.2; r.m.s. deviation 1.83 Å for 77 superimposed Cα). In common with other such domains, the UPF3b RNP domain has two faces, one helical and one with an exposed four-stranded antiparallel β-sheet. The RNP domains shown to bind RNA often contain exposed aromatic residues within RNP2 and RNP1 motifs on β1 and β3 that are involved in stacking with RNA bases. Human UPF3b RNP domain contains one aromatic residue (Tyr98) in RNP1 (residues 94–101) and lacks any aromatic residues in RNP2 (residues 53–58), which is contained in the highly conserved 52-KVVIRRLPP-60 motif, a signature of UPF3 homologs (**Fig. 1c**).

The UPF2-UPF3b interaction

UPF2(761–1054) forms a stable complex with UPF3b(42–143) with numerous, mainly hydrophilic contacts between the two molecules (**Table 1**). The main interaction interface involves helix 5 (residues 839–859) of UPF2 and the β-sheet surface of UPF3b (**Figs. 2a** and **3d,e**). UPF3b β-strands β_1 (RNP2) and β_4 bind to UPF2 helix α_5 via a set of salt bridge and hydrogen bond interactions (**Fig. 2c**). The principal interacting residues of UPF3b are charged residues of the RNP2 motif, namely Lys52, Arg56 and Arg57, which form salt bridge interactions with Asp847, Glu851 and Glu858 of UPF2 α_5. Lys52 also has a key role in stabilizing the UPF3b C-terminal loop (residues 134–140) following β_4 by hydrogen bonding to the main chain carbonyls of Phe136 and Lys138 (**Fig. 2c**). UPF3b β_4 contributes to the UPF2-UPF3b interaction by a conserved salt bridge formed between Glu132 and Arg854 of UPF2. Residues of RNP1 motif on strand β_3 do not interact directly with UPF2. However, the conserved Tyr98 on RNP1 helps position the neighboring Arg96 to form a salt bridge with Asp86, contributing to stabilization of the UPF3b loop between β_2 and β_3 (residues 84–94). This loop, together with the previously mentioned loop of residues 134–140, wraps around helix α_5 of UPF2, forming additional polar and hydrophobic interactions with UPF2 helices α_3, α_5, α_6 and α_7.

Most of the interacting residues of UPF2 and UPF3b are well conserved across species (**Fig. 1b,c**). The most highly conserved residues are UPF2 Asp847, Glu851, Glu858, Arg867 and UPF3b Lys52, Arg56, Arg57, Ser88 and Glu132. Notably, in *S. cerevisiae*, there is apparently a compensatory reversal from an Arg854-Glu132 to a glutamate-lysine salt bridge (**Fig. 1b,c**) that would preserve an important electrostatic interaction (**Fig. 2c**). However, the sequences of *S. pombe* Upf2 and Upf3 deviate substantially from the consensus sequence of all other known homologs (**Fig. 1b,c**). In particular, the respective equivalents of Arg56 in RNP2 and Arg96 in RNP1 are both phenylalanines, giving three rather than one aromatic residue on the β-sheet surface and making them more like canonical RNA-binding RNP domains. Furthermore, the Arg854-Glu132 salt bridge has equivalent residues of phenylalanine (adding a fourth aromatic residue to the interface) and threonine, respectively, and the equivalent residues of the Glu858-Arg57 salt bridge are both asparagines. Thus, in *S. pombe*, the interface

Table 1 Interactions between UPF2 and UPF3b

UPF3b	UPF2	UPF3b	UPF2
Salt bridges		**Hydrogen bonds**	
Lys52 (NZ)	Asp847 (OD1)	Thr87 (OG1)	Asp852 (OD1)
Lys52 (NZ)	Asp847 (OD2)	Thr87 (O)	Asn815 (ND2)
Arg56 (NE)	Glu851 (OE1)	Ser88 (OG)	Asp852 (OD2)
Arg56 (NH2)	Glu851 (OE2)	Ser88 (N)	Asp852 (OD2)
Arg56 (NH2)	Glu858 (OE1)	Ser88 (O)	Arg867 (NH1)
Arg56 (NH1)	Glu858 (OE2)	His92 (NE2)	Gln861 (NE2)
Glu132 (OE1)	Arg854 (NH1)	Pro135 (O)	Arg891 (NH1)
Glu132 (OE2)	Arg854 (NH2)		
Arg57 (NH1)	Glu858 (OE1)		

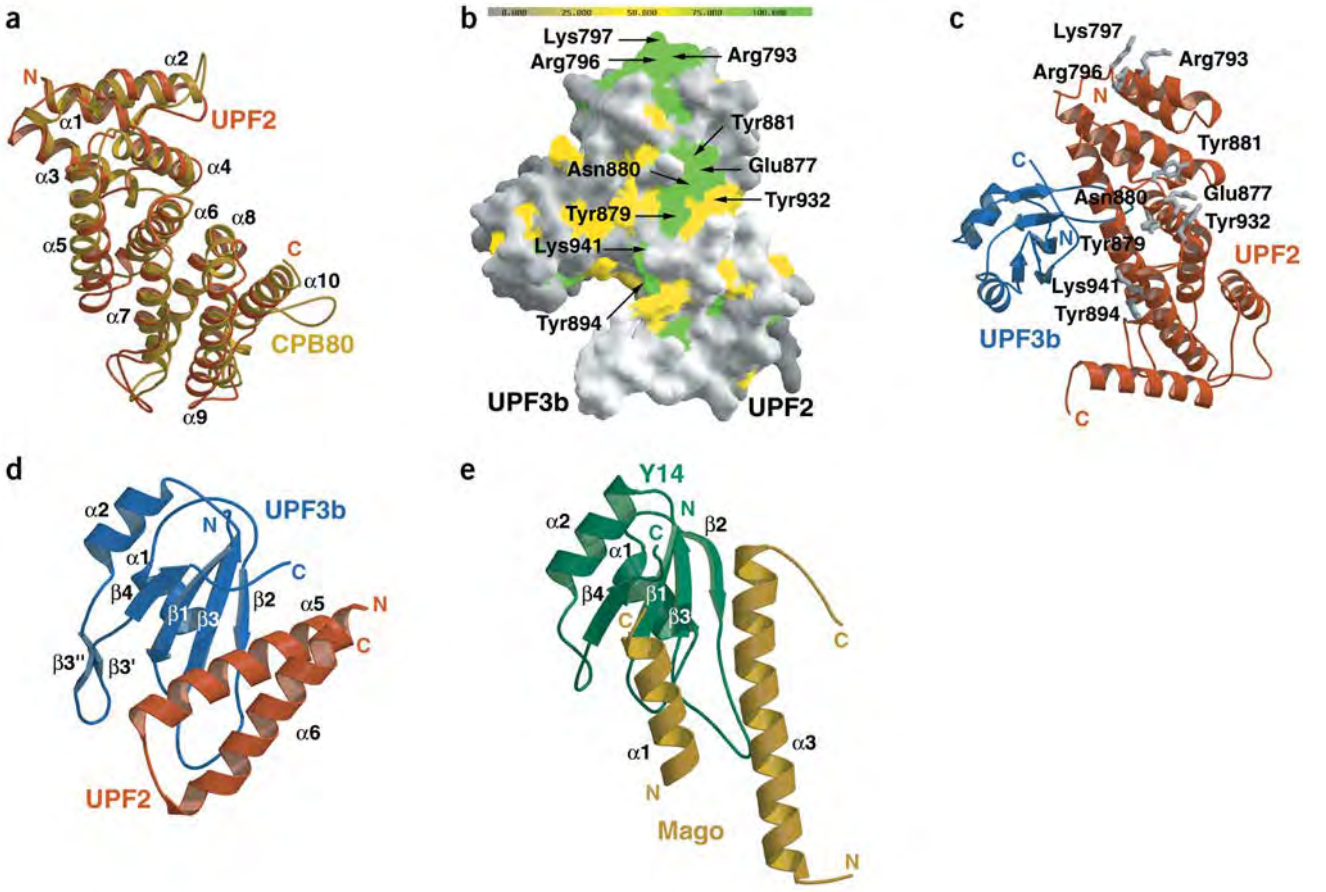


Figure 3 Structure and surface characteristics of the UPF2–UPF3b complex and a comparison with the Y14–Mago complex. (**a**) Superposition of human UPF2(770–983) (red) and CBP80(26–243) (yellow) MIF4G domains. Secondary structures are labeled according to UPF2. (**b**) Surface representation of the human UPF2(761–1054)–UPF3b(42–143) complex. Conserved surface residues (*S. pombe* sequence excluded) are represented from white to green according to the scale. All the labeled residues belong to UPF2. The complex is rotated –30° around the vertical axis relative to **a**. The figure was generated using GRASP[48]. (**c**) Ribbon representation of the UPF2(761–1054)–UPF3b(42–143) complex in the same orientation as in **b**, showing the conserved surface residues of UPF2 labeled in **b**. (**d**,**e**) Comparison of the UPF2–UPF3b and Y14–Mago complexes. (**d**) UPF3b β-sheet (blue) interacting with UPF2 helix 5 and 6 (red). Helix 5 is orientated across the β-sheet. (**e**) The Mago helices 1 and 3 (yellow) bind parallel to the flat Y14 β-sheet (green)[6]. The figure is based on the superposition of Y14 and UPF3b RNP domains with a Dali *Z*-score of 8.8.

between UPF2 and UPF3 would probably be much more hydrophobic, apart from the absolutely conserved Lys52-Asp847 interaction.

It is unlikely that the two other MIF4G domains in UPF2 could interact with UPF3b in a similar manner, as they have very low sequence homology with the third UPF2 MIF4G domain, and the residues involved in the interaction are not conserved (indeed the alignment is uncertain in this region). This is in agreement with the fact that the first two domains of UPF2 have never been reported to interact with UPF3b.

Mutational analysis of the UPF2-UPF3b interface

To test the importance of the principal interacting residues for the stability of the UPF2–UPF3b complex and to identify UPF2 and UPF3b mutants that prevent complex formation, we generated several constructs that we predicted would disrupt key electrostatic interactions. These included D847K, E858R, E851K D852R, and E851K D852R R854D in UPF2, and K52E and R56E in UPF3b. Using glutathione *S*-transferase (GST) pull-down assays we found that only the mutations UPF3b K52E and UPF2 E858R abolished completely the UPF2-UPF3b interaction (**Fig. 4a**, lane 8 and **Fig. 4b**, lane 16) whereas UPF3b R56E very strongly affected the interaction, but weak binding could still be observed (**Fig. 4a**, lane 9). Similar results were obtained with full-length UPF2 (data not shown). The correct folding of the UPF2 E858R mutant was confirmed by size-exclusion chromatography, which showed it to behave identically to the native protein (data not shown). Similarly, the K52E and R56E mutants were active in *in vivo* NMD assays (data not shown), suggesting that the observed phenotypes of these mutants originate from amino acid exchanges rather than from possible errors in folding. These results confirm the importance of these three highly conserved residues, but also illustrate the difficulty in predicting exactly what is most energetically significant in the interface. The residual binding of the UPF3b R56E mutation could possibly be explained by an interaction of the mutated glutamate residue with UPF2 Arg854. It is surprising that the double and triple mutants E851K D852R and E851K D852R R854D in UPF2 had no effect, but this may indicate that the long, flexible side chains of arginine and lysine find other conformations with favorable interactions. Given that the mutation UPF2 D847K did not affect the interaction, its salt bridge partner, the absolutely conserved UPF3b Lys52, may be more important in stabilizing the UPF3b loop of residues

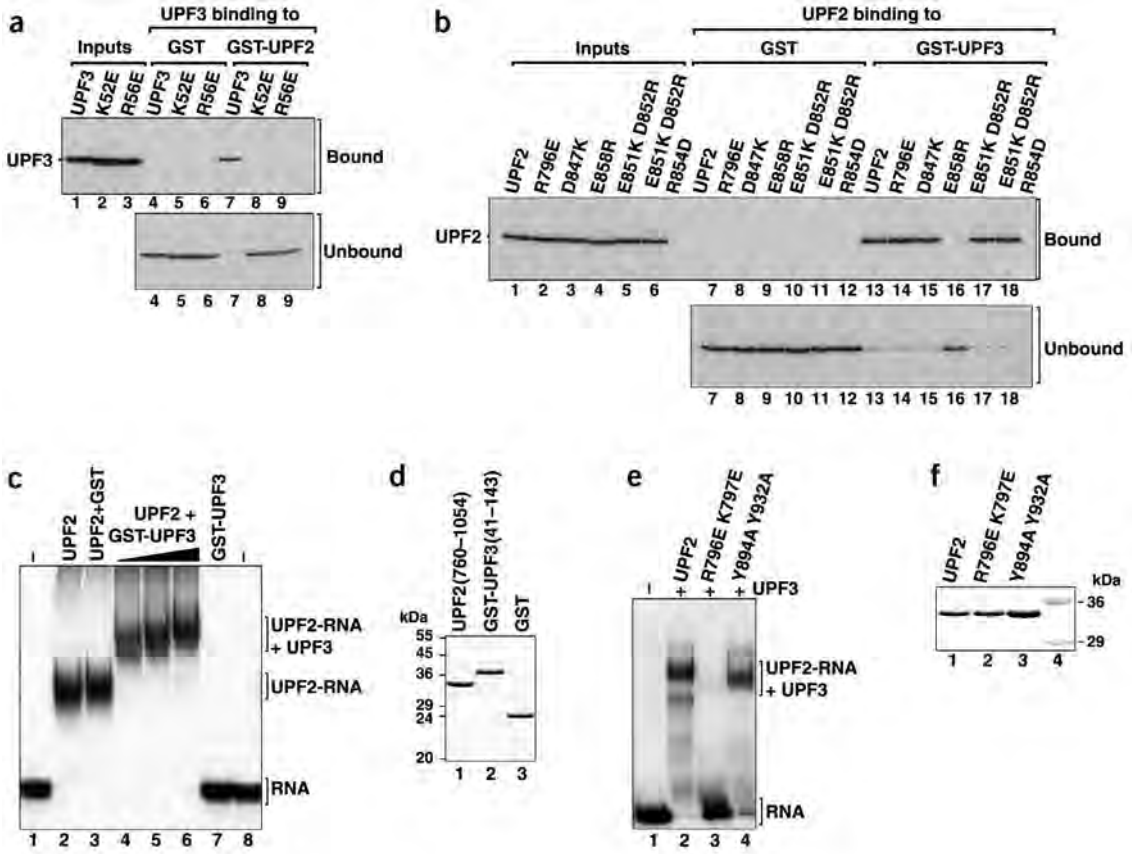


Figure 4 Identification of critical interacting residues and RNA-binding activity of the UPF2–UPF3B complex. (**a**) *In vitro*–synthesized [^{35}S]methionine-labeled UPF3 and the UPF3 mutants indicated above the lanes were incubated with glutathione agarose beads precoated with recombinant GST or a GST fusion of UPF2 (residues 761–1054). One-tenth of the input, one-third of the bound fractions and one-twentieth of the supernatants were analyzed by SDS-PAGE followed by fluorography. (**b**) [^{35}S]methionine-labeled UPF2 and the UPF2 mutants indicated above the lanes were incubated with glutathione agarose beads precoated with recombinant GST or a GST fusion of human UPF3B. Bound fractions were analyzed as described in **a**. (**c**–**f**) RNA-binding activity of UPF2–UPF3 complex. Gel-mobility assays (**c**,**e**) done with a labeled RNA probe and the purified recombinant proteins indicated above the lanes. UPF3 was N-terminally fused to GST. The position of the free RNA and of the RNA–protein complexes is shown at right. The UPF3 construct comprises residues 41–143 whereas the UPF2 fragment comprises residues 760–1054. Proteins used in the gel shift assays were analyzed by SDS-PAGE followed by Coomassie blue stain (**d**,**f**).

134–140 following β_4, which also has an important role in the interface (see above and **Fig. 2c**). In one of the two complexes in the asymmetric unit, UPF2 Arg854 is in a double conformation (the second one preserving only one hydrogen bond with Glu132; **Fig. 2c**), indicating a degree of mobility within the interface.

The third MIF4G domain of UPF2 binds RNA

The structure of the UPF2–UPF3b complex reveals that the RNP domain of UPF3b binds UPF2 using the β-sheet surface frequently used by other such domains to bind RNA. We therefore tested whether this domain of UPF3b could bind RNA using an electrophoretic gel-mobility retardation assay with purified proteins and a nonspecific single-stranded RNA probe (**Fig. 4c**,**d**). Even in the absence of UPF2, when the β-sheet surface of the RNP would be exposed to solvent, UPF3b did not bind RNA (**Fig. 4c**, lane 7). This observation is consistent with the occurrence of only one, rather than the more usual three, aromatic residue in the RNP1 and RNP2 motifs, although the presence of such residues is not necessarily a requirement for RNA binding (for example, in RNP domains 3 and 4 of polypyrimidine tract-binding protein[28]).

In contrast, the complex comprising the interacting domains of UPF2 and UPF3b showed nonspecific RNA-binding properties (**Fig. 4c**). This RNA-binding activity seems to be a property of the UPF2 construct, as UPF2 binds RNA even in the absence of UPF3b (**Fig. 4c**, lane 2) and a supershift occurs in the presence of increasing amounts of UPF3b (**Fig. 4c**, lanes 4–6), indicating that UPF2 can simultaneously interact with both RNA and UPF3.

As mentioned earlier, the third UPF2 MIF4G domain contains two surface patches of phylogenetically conserved residues (**Fig. 3b**,**c**). To test whether those could be involved in the RNA binding, we prepared two double mutants, R796E R797E (from the basic patch around Arg796) and Y894A Y932A (from the aromatic-rich region close to Glu877). Using the same electrophoretic gel-mobility retardation assay, we found that the R796E R797E double mutation strongly impairs RNA-binding activity of UPF2(761–1054), whereas the Y894A Y932A double mutation does not (**Fig. 4e**,**f**). The proper folding of the R796E R797E mutant UPF2 was confirmed by size-exclusion chromatography, which showed it to behave identically to the native protein (data not shown). These data suggest that the RNA-binding activity of UPF2(761–1054) involves the conserved basic region around Arg796 (**Fig. 3b**,**c**).

Table 2 Data collection and refinement statistics

	UPF2/3 native	UPF2/3 SeMet absorption peak	UPF2/3 SeMet inflection point	UPF2/3 SeMet remote point
Data collection				
Wavelength (Å)	0.976	0.9792	0.9795	0.9393
Space group	$P2_12_12$	$P2_12_12$		
Cell dimensions (Å)				
a	70.3	69.9		
b	100.2	99.6		
c	153.4	153.1		
Resolution range (Å)	30–1.95	30–2.3	30–2.3	30–2.4
Completeness (%)[a]	99.5 (98.6)	99.9 (99.9)	99.9 (99.9)	99.3 99.4)
R_{meas} (%)[a,b]	6.4 (30.5)	15.2 (33)	14.7 (39)	10.1 (26.9)
Total reflections	338,871	352,020	351,899	154,931
Unique reflections	79,316	91,641	91,583	80,499
Average redundancy	4.3	3.8	3.8	1.9
Figure of merit	0.23			
Z-score (SOLVE[38])	58.6			
Refinement statistics				
R-factor (%)[c]	20.5			
R_{free} (%)[c]	22.7			
R.m.s. deviations from ideal values				
Bond length (Å)	0.01			
Bond angle (°)	1.12			
Ramachandran plot (%)				
Favored	93.5			
Additional	6.5			
No. atoms				
Water	318			
β-ME[d] (S atoms only)	2			

[a]Values in parentheses are for the highest-resolution shell. [b]R_{meas} is as defined in ref. 49. [c]*R*-factor (R_{free}) = $[\Sigma||F_o| - k|F_c|| / \Sigma|F_o|] \times 100$, where 5% of randomly selected reflections were used for R_{free}. [d]β-mercaptoethanol.

DISCUSSION

UPF1, UPF2 and UPF3 are essential for nonsense-mediated mRNA decay in all eukaryotes. Many experiments in different organisms have shown that a direct interaction between UPF2 and UPF3 is essential for NMD in normal cells even though recent work has demonstrated that tethering of an N-terminally truncated UPF3 to a reporter RNA is sufficient to elicit NMD in the absence of the domain that interacts with UPF2 (ref. 5). We have solved the crystal structure of a complex between the interacting domains of human UPF2 and UPF3b. Contrary to expectations, UPF3b binds to UPF2 via the β-sheet surface of its RNP domain, which was originally speculated to be involved in RNA binding. Another recent example of an RNP domain using the β-sheet face for protein-protein interactions was revealed by the crystal structure[6,29,30] of the heterodimeric complex of Mago and Y14, components of the EJC and essential for NMD[6] as well as RNA localization[31]. The RNP domain of Y14 binds to Mago using its β-sheet surface. However, in contrast to the hydrophilic residues at the UPF2-UPF3b interface, mainly exposed hydrophobic residues, notably of the Y14 RNP1 and RNP2 motifs, mediate the Y14-Mago interaction. The Y14 β-sheet is rather flat, with the strands aligned with helix 1 and 3 of Mago, whereas the more concave UPF3b β-sheet wraps around helix 5 of UPF2, which is in the perpendicular orientation (**Fig. 3d,e**). Neither Y14 (ref. 6) nor the RNP domain of UPF3b binds to RNA. However, we have shown that the fragment of UPF2 used here, which includes the third MIF4G domain, does bind to RNA, and this binding is maintained in the complex with the RNP domain of UPF3b.

The RNP (RRM) domain is one of the most widespread modules in eukaryotic systems, readily identifiable in primary sequence analysis, and usually assumed to be an RNA-binding domain. The examples of UPF3 and Y14, together with that of U2AF65-RRM3 (whose C-terminal helical extension binds tightly to the β-sheet surface, also precluding RNA binding[32]) show that some RNP domains have evolved for specific protein-protein interactions rather than as nucleic acid modules. Thus neither the sequence nor the fold of RNP can serve as an indicator of RNA binding. On the other hand, some RNP domains carry out both functions, notably the small subunit CBP20 of the nuclear cap–binding complex, which binds capped RNA with its β-sheet surface and interacts tightly with CBP80 with its opposing α-helical face[33].

UPF3b has been reported to contain a nuclear export sequence (NES, 53-VVIR-RLPPTL-62) that enables its shuttling between nucleus and cytoplasm[4,34,35]. In *S. cerevisiae*, double and triple mutations of residues corresponding to human UPF3b residues Val53, Ile55 and Leu58 caused impaired nuclear export of UPF3 and inhibited NMD even if nuclear export was restored by fusing an extraneous NES. At the same time, a loss of interaction between UPF2 and UPFB in yeast and human was observed for these mutants as assayed by yeast two-hybrid experiments and coimmunoprecipitation[4,35]. UPF3b has also been found to be exported from the nucleus in a Crm1p-independent manner[35]. Our structure shows that the putative NES sequence corresponds to the RNP2 motif on strand β1 of UPF3b. Val53, Ile55 and Leu58 (conserved as hydrophobic residues in most RNP domains) are buried in the hydrophobic core of the domain, whereas neighboring and exposed residues at positions 52, 56 and 57 are crucially involved in the interface with UPF2. Mutation of Val53, Ile55 and Leu58 to alanine would probably severely destabilize the protein fold and disrupt the interaction with UPF2, as indeed has been observed[4,35]. That these mutations also abolish NMD is consistent with the importance of the UPF2-UPF3 interaction for this process. The structural results also suggest that the putative NES is unlikely to be a true NES (as the relevant hydrophobic residues are buried) and that the effect these mutations have on UPF3 nuclear export in yeast is probably an indirect result of the disruption of the UPF2-UPF3 interaction.

Mapping of phylogenetically conserved residues to the UPF2 molecular surface revealed three main conserved areas: the UPF3b-binding site, a site rich in aromatic residues close to Glu877 and a basic patch around Arg796 (**Fig. 3b,c**). Joint mutation of the residues corresponding to Phe874 and Glu877 (part of the conserved FIGEL motif in many MIF4G domains) in *S. pombe*, Phe748 and Glu751, to alanine, causes only a small decrease in NMD efficiency[3]. The same effect is observed when the corresponding residues from the second MIF4G domain of *S. pombe* UPF2, Phe545 and Glu548, are mutated. However, a strain expressing a combination of these mutations in UPF2 shows a complete loss of NMD function, suggesting that these conserved MIF4G

motifs stabilize the protein and/or mediate interactions with additional components of the NMD machinery. Our preliminary results on mapping the RNA-binding site on the third MIF4G domain of UPF2 suggest that the basic patch around Arg796 is involved. However the biological significance of the activity remains to be established.

In conclusion, the structure of the UPF2–UPF3 complex provides a starting point for further biochemical and structural studies to help elucidate the network of protein-protein and protein-RNA interactions that ultimately lead to the degradation of mRNAs carrying premature stop codons.

METHODS

Protein expression, purification and crystallization. Human UPF2(761–1054) and UPF3b(42–143) were expressed separately in *E. coli* BL21Star(DE3) pLysS (Invitrogen) from pProEXHTb expression vector (Invitrogen) as His-tag fusion proteins. Both proteins were first purified by affinity chromatography using Ni^{2+} resin. After His-tag cleavage with TEV protease, the proteins were applied to a heparin column and a second Ni^{2+} column. Protein complex was formed on ice in an excess of UPF3b and purified by size-exclusion chromatography. Pure UPF2–UPF3b protein complex was concentrated to ~15 mg ml^{-1} in a buffer containing 20 mM Tris, pH 7.0, 100 mM NaCl and 10 mM β-mercaptoethanol for crystallization. The best-diffracting crystals grew within 2 d at 20 °C in 100 mM ammonium acetate, 15 mM magnesium acetate, 50 mM sodium cacodylate, pH 6.5 and 8% (v/v) isopropanol. For data collection at 100 K, crystals were snap-frozen in liquid nitrogen with a solution containing mother liquor and 30% (v/v) glycerol. SeMet-substituted UPF2 was produced using *E. coli* BL21Star(DE3) pLysS in a defined medium containing 60 mg l^{-1} of SeMet. Purification of SeMet UPF2, complex formation with UPF3b and crystallization were done as for the native protein.

Data collection and structure determination. The UPF2(761–1054)/UPF3b (42–143) complex crystallizes in space group $P2_12_12$ with two complexes per asymmetric unit with a solvent content of 57%. The crystals diffract to a resolution of 1.9 Å. Diffraction data were collected using a Q4R ADSC CCD detector on beamline ID14-EH4 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and processed using XDS[36]. The structure was solved by a SeMet MAD experiment. Datasets with a resolution of 2.3–2.4 Å were collected at wavelengths corresponding to the inflection point, peak and remote wavelengths of the Se K-edge. A native data set was collected with a resolution of 1.95 Å. Selenium sites (10 of 12, for the two molecules) were identified using SHELXD[37] despite the mediocre quality of the MAD data. These sites were refined and used for phasing in SOLVE[38]. After solvent flattening with RESOLVE[39], a clearly interpretable electron density map was obtained. Phases were then transferred to the nearly isomorphous native data set and most of the model was automatically built using ARP-wARP[40]. The model was completed in O[41] and refined to an R_{free} of 23.3% and R-factor of 21.2% using first CNS and then REFMAC5 (with TLS refinement)[42]. NCS restraints between the two complexes in the asymmetric unit were maintained until the final stages of refinement. Dual conformations are observed for UPF2 residues Arg854, Ser895 and Ser898 in one of the complexes in the asymmetric unit. Detailed data and refinement statistics are in **Table 2**.

GST pull-down assays. Plasmids allowing the expression of GST fusions of full-length UPF3b and UPF2(761–1054) in *E. coli* were generated by inserting the coding sequences into the *Xho*I and *Not*I sites of pGEX5X-1 (Pharmacia). GST fusions were expressed in *E. coli* BL21(DE3) pLysS. UPF3b, full-length UPF2 and UPF2(761–1054) were also cloned into the *Xho*I and *Not*I sites of pCIλN vector[5]. Mutations were generated in these vectors using an oligonucleotide-directed *in vitro* mutagenesis system from Stratagene (QuikChange site-directed mutagenesis) and confirmed by sequencing. These vectors served as templates for the synthesis of [^{35}S]Met-labeled proteins using the combined transcription-translation (TNT) kit from Promega. For GST pull-down assays, 5 μg of GST-tagged recombinant protein immobilized on 20 μl of packed glutathione agarose beads was used per binding reaction. After binding of the GST-tagged proteins, beads were washed 3× with 0.5 ml of binding buffer (2× PBS, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100). *In vitro*–synthesized proteins (10 μl) were added to the beads in a final volume of 200 μl binding buffer. Binding was for 1 h at 4 °C. Beads were washed 3× with 0.5 ml binding buffer and once with 0.5 ml binding buffer without Triton X-100. Bound proteins were eluted with SDS sample buffer and analyzed by SDS-PAGE followed by fluorography.

***In vitro* RNA-binding assay.** For the RNA-binding assay (**Fig. 4**), a 77-nucleotide RNA probe was used (U1SII^{-}; ref. 43). This probe derives from stem-loop II of human U1 snRNA, but all complementary residues were mutated so that no secondary structure is formed. Similar results were obtained with a probe transcribed with T3 RNA polymerase using pbluescribe linearized with *Bam*HI as a template, indicating that binding is independent of the RNA sequence. Binding reactions were done in 15 mM HEPES, pH 7.9, 100 μM KCl, 0.2 mM EDTA, 5 mM $MgCl_2$, 10% (v/v) glycerol, 1 μM DTT, 0.05 mg ml^{-1} BSA, 0.05% (v/v) NP40 and 0.1 mg ml^{-1} of herring sperm single-stranded DNA (as unlabeled nonspecific competitor). Final sample volumes were 10 μl. The concentration of the recombinant proteins in the binding reactions was 0.1 mg ml^{-1}. After 20 min incubation at room temperature, 1 μl of a solution containing 0.05% (w/v) bromophenol blue was added to the reaction mixtures. Samples were applied to 5% nondenaturing PAGE (19:1, acryl/bisacryl ratio). Electrophoresis was carried out at a constant voltage of 12 V cm^{-1} at 4 °C in 0.5× TBE buffer. Complexes were visualized by autoradiography.

Coordinates. The coordinates and structure factors have been deposited in the Protein Data Bank (accession code 1UW4).

ACKNOWLEDGMENTS

We thank members of EMBL-ESRF Joint Structural Biology Group, notably A. McCarthy and R. Ravelli, for assistance with data collection on ESRF beamlines and for help with the crystallographic analysis. We are grateful to C. Petosa and C. Mazza for their frequent advice throughout the project. We also thank L. Maquat (University of Rochester) for providing UPF2 and UPF3 cDNA.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 21 October 2003; accepted 5 February 2004
Published online at http://www.nature.com/natstructmolbiol/

1. Wagner, E. & Lykke-Andersen, J. mRNA surveillance: the perfect persist. *J. Cell Sci.* **115**, 3033–3038 (2002).
2. Lykke-Andersen, J., Shu, M.D. & Steitz, J.A. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103**, 1121–1131 (2000).
3. Mendell, J.T., Medghalchi, S.M., Lake, R.G., Noensie, E.N. & Dietz, H.C. Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. *Mol. Cell. Biol.* **20**, 8944–8957 (2000).
4. Serin, G., Gersappe, A., Black, J.D., Aronoff, R. & Maquat, L.E. Identification and characterization of human orthologues to *Saccharomyces cerevisiae* Upf2 protein and Upf3 protein (*Caenorhabditis elegans* SMG-4). *Mol. Cell. Biol.* **21**, 209–223 (2001).
5. Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W. & Kulozik, A.E. Y14 and hUpf3b form an NMD-activating complex. *Mol. Cell* **11**, 939–949 (2003).
6. Fribourg, S., Gatfield, D., Izaurralde, E. & Conti, E. A novel mode of RBD-protein recognition in the Y14–Mago complex. *Nat. Struct. Biol.* **10**, 433–439 (2003).
7. Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P. & Izaurralde, E. Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *EMBO J.* **22**, 3960–3970 (2003).
8. Denning, G., Jamieson, L., Maquat, L.E., Thompson, E.A. & Fields, A.P. Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. *J. Biol. Chem.* **276**, 22709–22714 (2001).
9. Chiu, S.Y., Serin, G., Ohara, O. & Maquat, L.E. Characterization of human Smg5/7a: a protein with similarities to *Caenorhabditis elegans* SMG5 and SMG7 that functions in the dephosphorylation of Upf1. *RNA* **9**, 77–87 (2003).
10. Anders, K.R., Grimson, A. & Anderson, P. SMG-5, required for *C. elegans* nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. *EMBO J.* **22**, 641–650 (2003).
11. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M.J. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–4997 (2001).
12. Ishigaki, Y., Li, X., Serin, G. & Maquat, L.E. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**, 607–617 (2001).
13. Czaplinski, K. *et al.* The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev.* **12**, 1665–1677 (1998).

14. Kim, V.N., Kataoka, N. & Dreyfuss, G. Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* **293**, 1832–1836 (2001).
15. Wang, W., Czaplinski, K., Rao, Y. & Peltz, S.W. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *EMBO J.* **20**, 880–890 (2001).
16. Nagy, E. & Maquat, L.E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**, 198–199 (1998).
17. Zhang, S., Ruiz-Echevarria, M.J., Quan, Y. & Peltz, S.W. Identification and characterization of a sequence motif involved in nonsense-mediated mRNA decay. *Mol. Cell. Biol.* **15**, 2231–2244 (1995).
18. Schell, T. *et al.* Complexes between the nonsense-mediated mRNA decay pathway factor human upf1 (up-frameshift protein 1) and essential nonsense-mediated mRNA decay factors in HeLa cells. *Biochem. J.* **373**, 775–783 (2003).
19. Letunic, I. *et al.* Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30**, 242–244 (2002).
20. Ponting, C.P. Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. *Trends Biochem. Sci.* **25**, 423–426 (2000).
21. Marcotrigiano, J. *et al.* A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol. Cell* **7**, 193–203 (2001).
22. Mazza, C., Ohno, M., Segref, A., Mattaj, I.W. & Cusack, S. Crystal structure of the human nuclear cap binding complex. *Mol. Cell* **8**, 383–396 (2001).
23. He, F., Brown, A.H. & Jacobson, A. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol. Cell. Biol.* **17**, 1580–1594 (1997).
24. He, F., Brown, A.H. & Jacobson, A. Interaction between Nmd2p and Upf1p is required for activity but not for dominant-negative inhibition of the nonsense-mediated mRNA decay pathway in yeast. *RNA* **2**, 153–170 (1996).
25 Hall, K.B. RNA-protein interactions. *Curr. Opin. Struct. Biol.* **12,** 283–288 (2002).
26. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
27. Handa, N. *et al.* Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585 (1999).
28. Conte, M.R. *et al.* Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* **19**, 3132–3141 (2000).
29. Lau, C.K., Diem, M.D., Dreyfuss, G. & Van Duyne, G.D. Structure of the y14-magoh core of the exon junction complex. *Curr. Biol.*, 933–941 (2003).
30. Shi, H. & Xu, R.M. Crystal structure of the *Drosophila* Mago nashi–Y14 complex. *Genes Dev.* **17**, 971–976 (2003).
31. Hachet, O. & Ephrussi, A. *Drosophila* Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr. Biol.* **11**, 1666–1674 (2001).
32. Selenko, P. *et al.* Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell.* **11**, 965–976 (2003).
33. Mazza, C., Segref, A., Mattaj, I.W. & Cusack, S. Large-scale induced fit recognition of an m(7)G*P*G cap analogue by the human nuclear cap-binding complex. *EMBO J.* **21,** 5548–5557 (2002).
34. Shirley, R.L., Lelivelt, M.J., Schenkman, L.R., Dahlseid, J.N. & Culbertson, M.R. A factor required for nonsense-mediated mRNA decay in yeast is exported from the nucleus to the cytoplasm by a nuclear export signal sequence. *J. Cell Sci.* **111**, 3129–3143 (1998).
35. Shirley, R.L., Ford, A.S., Richards, M.R., Albertini, M. & Culbertson, M.R. Nuclear import of Upf3p is mediated by importin-α/-β and export to the cytoplasm is required for a functional nonsense-mediated mRNA decay pathway in yeast. *Genetics* **161**, 1465–1482 (2002).
36. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
37. Uson, I. & Sheldrick, G.M. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643–648 (1999).
38. Terwilliger, T.C., Kim, S.H. & Eisenberg, D. Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr. A* **43**, 1–5 (1987).
39. Terwilliger, T.C. Reciprocal-space solvent flattening. *Acta Crystallogr. D* **55**, 1863–1871 (1999).
40. Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
41. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
42. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
43. Scherly, D. *et al.* Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**, 4163–4170 (1989).
44. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
45. Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
46. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
47. Esnouf, R.M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
48. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Gen.* **11**, 281–296 (1991).
49. Diederichs, K. & Karplus, P.A. Improved *R*-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269–275 (1997).

Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1

JAN KADLEC, DELPHINE GUILLIGAY, RAIMOND B. RAVELLI, and STEPHEN CUSACK
European Molecular Biology Laboratory, Grenoble Outstation, BP 181, 38042 Grenoble Cedex 9, France

ABSTRACT

UPF1 is an essential eukaryotic RNA helicase that plays a key role in various mRNA degradation pathways, notably nonsense-mediated mRNA decay (NMD). In combination with UPF2 and UPF3, it forms part of the surveillance complex that detects mRNAs containing premature stop codons and triggers their degradation in all organisms studied from yeast to human. We describe the 3 Å resolution crystal structure of the highly conserved cysteine–histidine-rich domain of human UPF1 and show that it is a unique combination of three zinc-binding motifs arranged into two tandem modules related to the RING-box and U-box domains of ubiquitin ligases. This UPF1 domain interacts with UPF2, and we identified by mutational analysis residues in two distinct conserved surface regions of UPF1 that mediate this interaction. UPF1 residues we identify as important for the interaction with UPF2 are not conserved in UPF1 homologs from certain unicellular parasites that also appear to lack UPF2 in their genomes.

Keywords: nonsense-mediated mRNA decay; NMD; surveillance complex; UPF1; X-ray crystallography

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is an mRNA degradation pathway that detects and eliminates aberrant coding transcripts containing premature termination codons (PTC) originating from nonsense or frameshift mutations. The PTC-containing transcripts would otherwise be translated into truncated proteins that might have a deleterious effect on the cell (Maquat 2004).

According to the current consensus model of NMD in mammalian cells, the recognition of a PTC requires a splicing-dependent deposition of a multiprotein complex, the exon junction complex (EJC), 20–24 nucleotides (nt) upstream of a splice junction (Le Hir et al. 2001). During mRNA export to the cytoplasm, a perinuclear protein, UPF2, is recruited to the EJC via UPF3 (Lykke-Andersen et al. 2000). In mammals nonsense-mediated mRNA decay has been proposed to occur during the first "pioneer" round of translation (Ishigaki et al. 2001). With the first passage of a ribosome, the EJCs would normally be stripped from the mRNA. If, however, translation terminates at a PTC upstream of an EJC, UPF2 associated with a downstream EJC can be bound by UPF1 that is recruited to the terminating ribosome within the so-called SURF complex, which also includes the translation release factors eRF1 and eRF3 and the UPF1 kinase, Smg1 (Kashima et al. 2006). The association of SURF and UPF2-EJC (the DECID complex) is followed by phosphorylation of UPF1 by Smg1 (Conti and Izaurralde 2005; Kashima et al. 2006). The phosphorylated form of UPF1 is then recognized by Smg7, which targets the aberrant transcript for decay in mRNA degradation foci (Unterholzner and Izaurralde 2004; Fukuhara et al. 2005).

Human UPF1 (hUPF1, also known as RENT1) consists of 1118 amino acid residues. In its N-terminal region, UPF1 has a conserved cysteine–histidine-rich (CH-rich) region (residues 123–213), while centrally it possesses the seven conserved motifs characteristic of eukaryotic group I RNA helicases (Applequist et al. 1997). UPF1 displays nucleic acid-dependent ATPase activity and ATP-dependent 5′–3′ helicase activity, which is required for its NMD function (Bhattacharya et al. 2000).

UPF1 has been assigned an increasing number of functions and consequently interacts with numerous proteins. Yeast and/or human UPF1 interacts with several NMD factors including UPF2, Smg1, Smg5, Smg6, Smg7, and HrpI (Gonzalez et al. 2000; Conti and Izaurralde 2005).

Reprint requests to: Stephen Cusack, European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France; e-mail: cusack@embl-grenoble.fr; fax: (33) 476207199.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.177606.

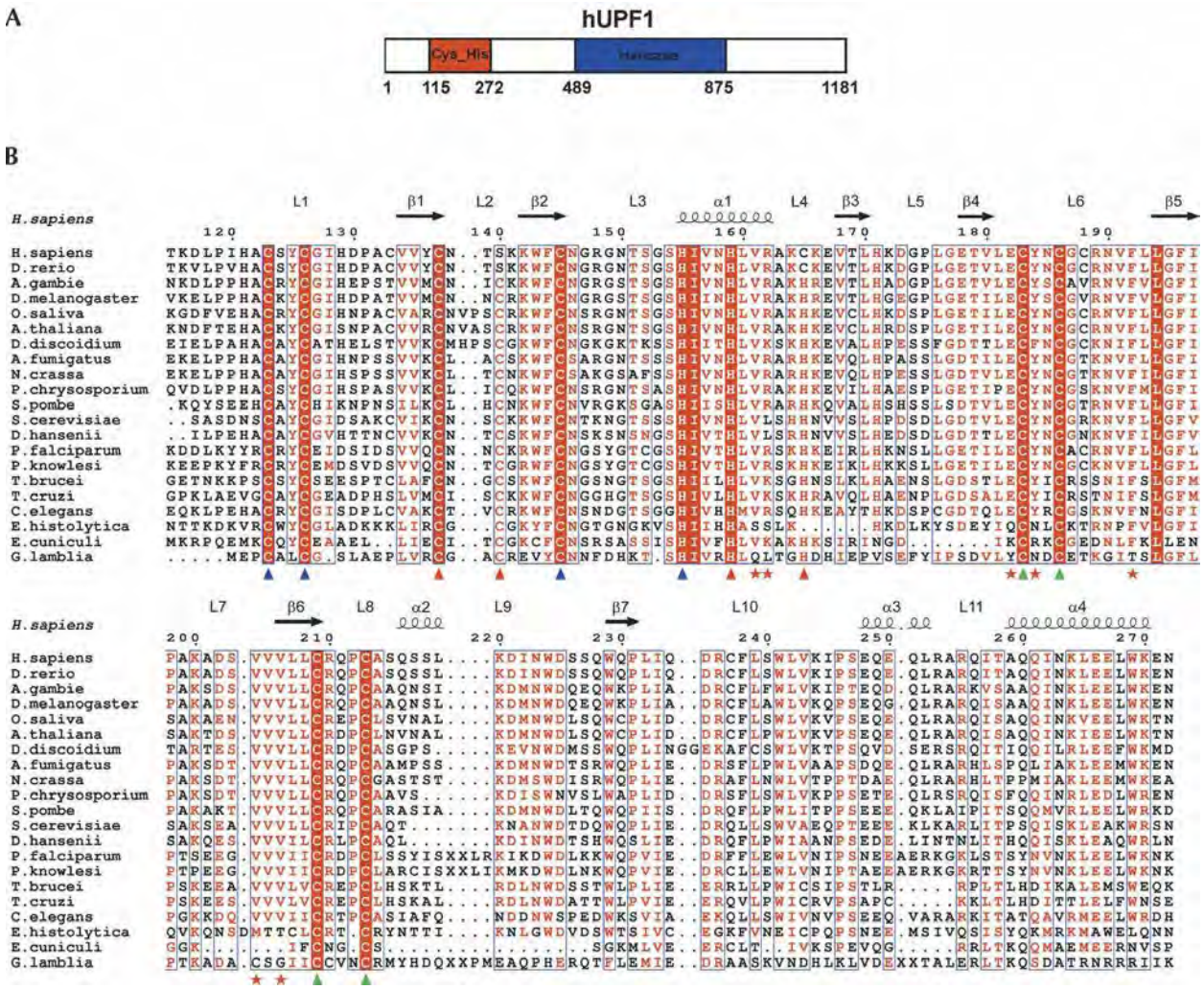


FIGURE 1. Human UPF1 organization and sequence alignment of the CH-rich domain. (*A*) Schematic representation of human UPF1 domain structure. The RNA helicase domain is shown in blue (Applequist et al. 1997). The CH-rich domain as defined in this work is in red. (*B*) Sequence alignment of UPF1 proteins. Residues that are 100% conserved are in solid red boxes. Those with similarity >70% are labeled in red. The secondary structures of hUPF1 are blue; (α) α-helix; (β) β-strand; (L) L-loop. Blue, red, and green triangles indicate residues involved in the coordination of Zn1, Zn2, and Zn3, respectively. Red stars indicate residues that abolish the interaction with UPF2. In the sequences of *E. histolytica* and *G. lamblia*, XX indicates where long insertions have been omitted for clarity. The figure was generated with CLUSTALX (Thompson et al. 1997) and ESPript (http://espript.ibcp.fr/ESPript/ESPript/).

The UPF2 binding region was mapped to the N terminus of UPF1, including the CH-rich region (He et al. 1997). UPF1 also interacts with decapping enzymes Dcp1 and Dcp2 and exosome-associated protein Ski7p, suggesting an active role in the recruitment of degradation enzymes to nonsense transcripts (Lykke-Andersen 2002; Takahashi et al. 2003). Yeast UPF1 increases translation termination efficiency by preventing nonsense codon read-through, and both yeast and human UPF1 interact with translation termination factors eRF1 and eRF3 (Czaplinski et al. 1998; Kashima et al. 2006). Yeast UPF1 also interacts with nucleoporins Nup100 and Nup116 (Nazarenus et al. 2005). UPF1 is required for nonsense-mediated altered splicing (Mendell et al. 2002) and in *Caenorhabditis elegans*, together with Smg5 and Smg7, is important for persistence of mRNA silencing by RNA interference (Domeier et al. 2000). Recently, UPF1 was shown to interact with StauI, an interaction essential for a newly identified mRNA degradation pathway called Stau1-mediated decay (Kim et al. 2005) and with SLBP for the specific degradation of histone mRNA (Kaygun and Marzluff 2005). Probably as a result of its multiple important functions, UPF1 is essential for mammalian embryonic viability (Medghalchi et al. 2001).

Here we present the crystal structure of the CH-rich domain of human UPF1 at a resolution of 3 Å. Overall the domain has a unique fold coordinating three zinc atoms.

However, two substructures coordinating one and two zinc atoms, respectively, have remarkable and unexpected structural similarities to the RING domain commonly found in ubiquitin ligases. Structure-based mutagenesis of conserved residues revealed two distinct UPF1 surface regions likely to be involved in the interaction with UPF2.

RESULTS AND DISCUSSION

The crystal structure of CH-rich domain of UPF1

Expression of the putative UPF2-interacting region of hUPF1 (residues 115–245) in *Escherichia coli* yielded insoluble protein. However, a C-terminally extended fragment, spanning residues 115–272 (Fig. 1), could be expressed in a soluble form, and crystals grown. Previous measurements using the proton-induced X-ray emission (microPIXE) (Garman and Grime 2005) technique had shown that the domain contains three zinc atoms (data not shown). This enabled the atomic structure to be solved at 3 Å resolution using the anomalous scattering of the native zinc atoms (Table 1).

Residues 118–272 of hUPF1 are visible in the electron density maps, and these form a single domain with a compact fold containing three structural zinc atoms (Fig. 2). The main feature of the structure is a central, pseudo-twofold symmetric, four-stranded antiparallel β-sheet (strands β1–β4) with symmetric flanking loops L1 and L6 (Fig. 2A). While L1 is involved in coordinating Zn1, L6 is stabilized by Zn3. The β-sheet packs against helix α1 that follows β2. L6 connects this β1β2α1β3β4 arrangement to a three-stranded antiparallel β-sheet (strands β5–β7) that is in a perpendicular orientation to the first β-sheet. β6 is followed by L8, which participates in coordination of Zn3 and a one-turn helix, α2. The C terminus of the domain is formed by long meandering loops L10 and L11 and two helices, α3 and α4, that wrap around the N-terminal part of the structure (Fig. 2B). L10 packs against strands β3 and β4, helix α1, and loops L5 and L3. Helices α3 and α4 and the intervening linker (L11) bind between loop L3, helix α1, and loop L1 (Fig. 2). Using the DALI server (Holm and Sander 1993), no similar structures were found, indicating that overall this domain represents a previously unseen fold.

TABLE 1. Data collection and refinement statistics

	UPF1
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	64.6, 73.0, 73.2
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	20–2.95 (3.0–2.95)[a]
R_{sym} or R_{merge}	0.124 (0.424)
$I/\sigma I$	10.2 (2.9)
Completeness (%)	90.9 (96.5)
Redundancy	4.9 (3.4)
Refinement	
Resolution (Å)	20–2.95
No. reflections	6685 (work), 315 (test)
R_{work}/R_{free}	0.221/0.259
No. atoms[b]	
Protein	2458
Ligands/ions	6 zinc
Average *B*-factor (Å^2)	
Protein	34.5
RMS deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.23
Ramachandran plot	
Favorable regions	89.1%
Additionally allowed	10.9%

[a]The highest resolution shell is shown in parentheses.
[b]Total atom count for two molecules in the asymmetric unit.

Zinc coordination within the UPF1 CH-rich domain

The three zinc atoms occur in different variants of zinc fingers. Zn1 is coordinated by three cysteines and one histidine within a treble clef zinc-finger-like motif (Fig. 2D; Krishna et al. 2003). This motif consists of a zinc knuckle (a unique turn with the consensus sequence CPXCG) providing two zinc ligands, the two others coming from a β-hairpin and C-terminal part of an α-helix. For Zn1, ligands Cys123 and Cys126 are provided by a Zn-knuckle within loop L1, Cys145 comes from the C terminus of β2 of the following β-hairpin (β1–β2), and His155 from the N terminus of downstream helix α1. Zn2, with two ligands (Cys137 and Ser140) being provided by a Zn-knuckle in a β-hairpin (loop L2), and His159 and Cys165 coming, respectively, from helix α1, which packs against the β-hairpin and the following loop L4 (Fig. 2D). This motif is, in fact, a classical two-cysteine, two-histidine-like (C2H2) zinc finger and as shown in Figure 1B, most UPF1 homologs, indeed, have the conventional C2H2 ligands. Only in vertebrates has C2H2 evolved to become CSHC, the structure suggesting that Ser140 is actually a zinc ligand, although this cannot be certain at the current resolution. Finally, Zn3 is coordinated within a second treble-clef zinc-finger motif, with Cys183 and Cys186 provided by a Zn-knuckle (L6), Cys209 coming from the C terminus of β6 of a following β-hairpin (β5–β6) and Cys213 coming from L8 (Fig. 2E). A feature of this zinc site is that it includes a *cis*-proline, Pro212. Both treble-clef motifs of UPF1 can be very well superposed on L24E, a ribosomal protein interacting with 23S RNA via the corresponding helix (Klein et al. 2004). All the zinc coordinating residues are absolutely conserved among known UPF1 homologs (Fig. 1B), suggesting that the zinc atoms are crucial for UPF1 structure and hence function.

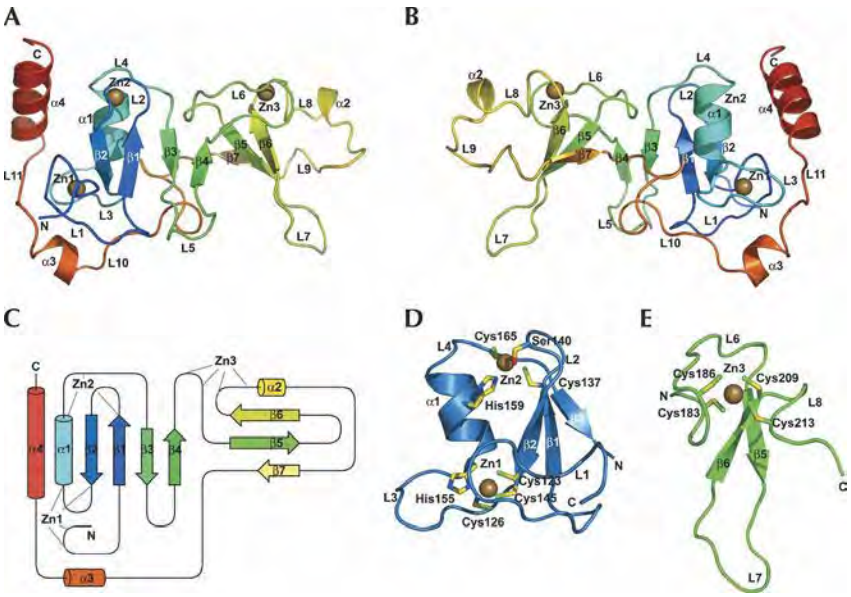


FIGURE 2. Crystal structure of the CH-rich domain of human UPF1. (*A*) Ribbon diagram of the structure of hUPF1(*115–272*). The polypeptide chain is colored from the N terminus (blue) to the C terminus (red). The coordinated zinc atoms are in brown. This and other ribbon diagrams were generated with Pymol (http://www.pymol.org). (*B*) The same ribbon model as in *A* rotated 180° around vertical axes. (*C*) Schematic diagram of the UPF1 CH-rich domain topology. The figure was generated with TopDraw (Ohi et al. 2003). (*D*) A view of the N-terminal part of the CH-rich domain (residues *118–172*) coordinating Zn1 and Zn2 within a treble-clef and C2H2 zinc-finger motifs, respectively. Ser140 is shown in a position where other UPF1 homologs have a fourth cysteine. (*E*) Coordination of Zn3 within the second treble clef zinc-finger motif (residues *181–216*) of the UPF1 CH-rich domain.

Several reports aiming to characterize UPF1 function relied on mutagenesis of cysteine and histidine residues within the CH-rich region (Weng et al. 1996; de Pinto et al. 2004; Kashima et al. 2006). Most of the results can be rationalized on the basis of whether or not the mutations were of zinc ligands and hence whether they disrupted the structural integrity of the domain. For example, in yeast, individual mutations C72S, H110R (Cys133, His171 in hUPF1) had no effect on UPF1 function (Weng et al. 1996), consistent with these residues not being involved in zinc binding, whereas mutations H94R, H98R, C122S, and C125S (His155, His159, Cys183, and Cys186 in hUPF1), which are involved, respectively, in Zn1, Zn2, Zn3, and Zn3 binding, result in complete loss of UPF1 activities in NMD as well as in preventing nonsense codon read-through (Weng et al. 1996). Our structural results enable design of more rational mutations that selectively perturb putative interaction surfaces without affecting the overall domain structure and stability (see below).

The UPF1 CH-rich domain contains two modules similar to ubiquitin ligase RING or U-box domains

A search of the PDB with the protein structure comparison service SSM at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm) revealed that the CH-rich domain of UPF1 contains two similar structural modules that both resemble the RING-box or U-box found most often in E3 ubiquitin ligases (Fig. 3; Ohi et al. 2003; Petroski and Deshaies 2005). Neither of these modules could be identified by sequence analysis, and their juxtaposition in UPF1 is unique. The RING-like module 1 of hUPF1 spans residues 121–172 (loops L1–L4, strands β1–β3, and helix α1), while the RING-like module 2 includes residues 180–233 (loops L6–L9, strands β5–β7, and helix α2) (Fig. 3A,B). RING-box domains usually coordinate two structural zinc atoms, while U-box domains are zinc free, the fold being instead maintained by a hydrogen-bond network (Fig. 3C,D; Zheng et al. 2000; Ohi et al. 2003; Andersen et al. 2004). UPF1 module 1 has the two zinc atoms, but compared to canonical RING-boxes, contains a lengthened loop L3 between β2 and α1, and a shortened loop L4 connecting α1 and β3 (Fig. 3E, cf. also A with C and D). Unusually, UPF1 module 2 coordinates only a single zinc atom and also has a minimal α-helix. Additionally, both RING/U-box domains, as well as both UPF1 RING-like modules, are similar in structure to the ZZ domain of CREB-binding protein (CBP), another two-zinc binding module found in diverse proteins (Fig. 3F; Legge et al. 2004). The ZZ domain has been proposed to function in protein–protein interactions but is not well characterized.

Although there is no evidence that ubiquitination plays a role in NMD or any other function of UPF1, structural homology with RING/U-box domains can perhaps be used to define potential protein–protein interacting surfaces. The structure of the RING domain from c-Cbl complexed with the E2 ubiquitin conjugating enzyme UbcH7 serves a model for interactions between RING and U-box E3s with E2 enzymes (Zheng et al. 2000). In this structure, Ile383, Trp408, Pro417, and Phe418 of c-Cbl make a largely hydrophobic interacting surface contacting loops 1 and 2 of UbcH7 (Fig. 3C). Similar residues are exposed in the U-box domain of splicing factor Prp19 (Fig. 3D; Ohi et al. 2003). In UPF1 RING-like module 1, a corresponding surface formed by Tyr125, His159, Lys142, and Phe144 is partially occluded by interaction with residues from the C-terminal helix α4 (Fig. 3A). On the other hand, in UPF1 module 2, it is formed by conserved Tyr184, Trp224, Val204, Val206, and

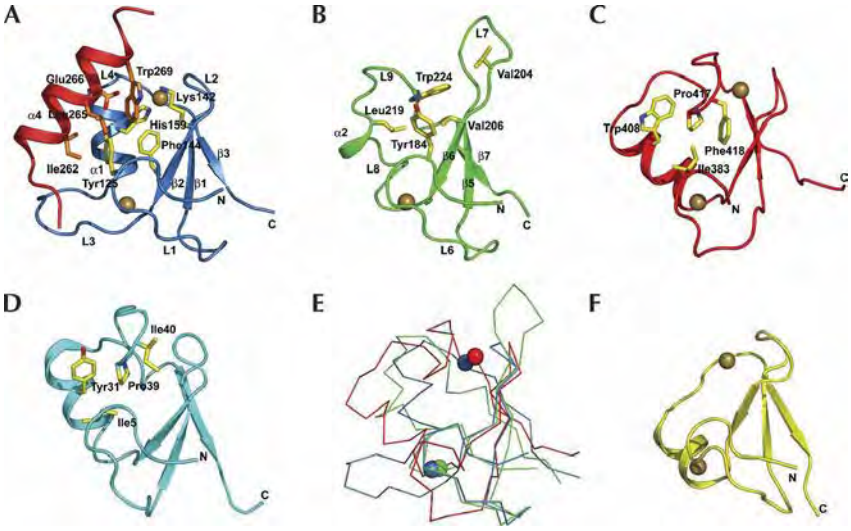


FIGURE 3. UPF1(*115–272*) RING-like modules are similar to RING and U-box domains of E3 ubiquitin ligases. (*A*) RING-like module 1 (*121–172*) of the UPF1 CH-rich domain. Conserved Tyr125 and His159 together with conserved Lys142 and Phe144 are shown that interact with residues of α4 and form a putative protein-binding surface. Tyr125 and His159 are in positions corresponding to Ile383 and Trp408 of c-Cbl (*C*) and Ile5 and Tyr31 of Prp19 in *D*. (*B*) RING-like module 2 (*180–233*) of the UPF1 CH-rich domain. A hydrophobic surface is shown formed by conserved residues Tyr184, Trp224, Val204, Val206, and Leu219. Superposing the two modules of UPF1 gives a root mean square (RMS) deviation of 1.57 Å for 38 C_α atoms (DALI Z-score of 4.1). (*C*) The RING domain (residues *381–431*) of c-Cbl showing residues involved in the binding of E2 enzyme UbcH7 (Zheng et al. 2000) (PDB entry 1FBV). The figure is based on the superposition with the UPF1 RING-like module 1 with an RMS deviation of 1.95 Å for 32 C_α atoms (DALI *Z*-score of 1.5). (*D*) U-box domain (residues *1–56*) of splicing factor Prp19 (PDB entry 1N87) and the residues of a putative E2 enzyme interaction interface (Ohi et al. 2003). The UPF1 RING-like module 1 and the Prp19 domain superpose with an RMS deviation of 1.89 Å for 32 C_α atoms (DALI *Z*-score of 3.4). (*E*) A superposition of C_α trace of UPF1 RING-like module 1 (*115–172*) in blue, UPF1 module 2 (*180–233*) in green, and the RING-box domain of c-Cbl in red, showing also the zinc atoms (the first zinc atom of c-Cbl overlaps with Zn1 of UPF1). (*F*) ZZ domain of CREB-binding protein (Legge et al. 2004) (PDB entry 1TOT). The UPF1 RING-like module 1 and the ZZ domain superpose with an RMS deviation of 2.03 Å for 32 C_α atoms (DALI *Z*-score of 3.6).

Leu219 and being solvent-exposed, could be available for protein–protein interactions (Fig. 3B).

Mutational analysis of UPF1–UPF2 interaction

Recently, it has been confirmed that the interaction between hUPF1 and hUPF2 is, indeed, essential for NMD. It promotes the contact between the SURF complex (Smg1, UPF1, eRF3, and eRF1) at the premature termination codon and a downstream hUPF2–EJC complex, thus triggering UPF1 phosphorylation (Kashima et al. 2006). We tested whether the hUPF1 CH-rich domain can bind hUPF2 in vitro. Indeed, when His-tagged hUPF1(115–272) is coexpressed with hUPF2(761–1207) in *E. coli*, a binary complex can be purified using a Ni^{2+} resin (Fig. 4E, lane 1). The complex is stable enough to be purified by gel filtration. Complex formation did not require the presence of CBP80, which has been recently described as enhancing the interaction in mammalian cells (Hosoda et al. 2005). After adding to this binary complex purified hUPF3b(42–143), which binds tightly to the third MIF4G domain of UPF2 (residues 761–1054) (Kadlec et al. 2004), a stable ternary UPF complex could be formed (data not shown).

Mapping of phylogenetically conserved residues to the molecular surface of the CH-rich domain of UPF1 revealed two major concentrations, both of which are at the interface of the two β-sheets (Fig. 4). The first one overlaps with the hydrophobic surface of RING-box-like module 2 (see above) and is formed by residues of β4, β6, and loops L6, L9, and L10 (Fig. 4A,C). The other conserved surface is formed by residues from loops L6, L9, L10, and helix α1 (Fig. 4B,D).

We prepared several point mutants of solvent-exposed residues in these two conserved surfaces as well as in the putative interaction surfaces identified by similarity to c-Cbl and tested them for binding to UPF2(761–1207) in His-tag pull-down assays (Fig. 4E–G). The interaction with UPF2 was substantially reduced when Val204 was mutated to aspartate (V204D). Similar effects were observed for the single mutant V206E and double mutant E182R + Y184D (Fig. 4F, lanes 3–5). All these residues are in one conserved surface (Fig. 4A,C). Interestingly, the UPF2 binding was also affected by mutations in the other conserved patch (e.g., F192E and V161E + R162E) (Fig. 4E, lanes 4,5). Additionally, a triple mutation, encompassing both regions, V206E + V161E + R162E essentially abolished the interaction with UPF2 (Fig. 4G, lane 4). These data suggest that the full UPF2-binding activity of UPF1 involves both of these two regions, and this correlates with the previous two-hybrid analysis of the UPF1-binding site in UPF2 that was mapped to two adjacent regions in the UPF2 C terminus, separated by a flexible linker (He et al. 1996). The mutations in the putative interaction surfaces identified by similarity to c-Cbl did not affect UPF2 binding (Fig. 4E). The correct folding of the UPF1 mutants that affected UPF2 binding (V204D, V206E, E182R + Y184D, F192E, V161E + R162E, V206E + V161E + R162E, Y125E) was confirmed by size-exclusion chromatography, which showed them to behave identically to the native protein (data not shown). Moreover these experiments

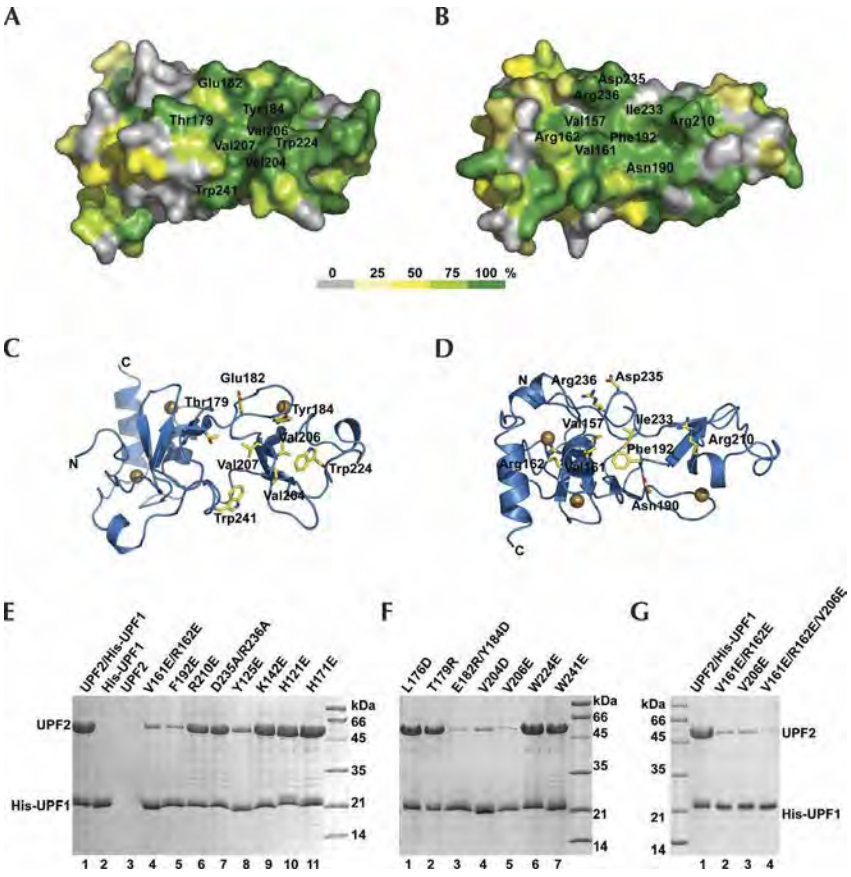


FIGURE 4. Identification of UPF2 binding surfaces. (*A*,*B*) Surface representation of the domain, highlighting areas of conserved surface residues. The conservation of the surface residues, based on the first 12 sequences in the alignment shown in Figure 1B, is represented from gray to green (green is 100% conserved) according to the color scale bar. (*C*,*D*) Ribbon representation of the UPF1 CH-rich domain in the same orientation as in *A* and *B*, showing the conserved surface residues of UPF1 labeled in *A* and *B*. (*E–G*) His-tagged UPF1 mutants indicated above the lanes were coexpressed with UPF2(*761–1207*) and purified using Ni^{2+} resin. The resin was extensively washed with 20 mM and 100 mM imidazole. Bound proteins were eluted with 250 mM imidazole and analyzed by 15% SDS-PAGE.

showed that whereas by pull-down assays these UPF1 mutants show some residual binding to UPF2, the complex is not stable to gel filtration, unlike the wild-type complex.

In the vast majority of eukaryotic genomes, including metazoans, yeasts, and the parasites *Trypanosoma brucei*, *Trypanosoma cruzi* and *Plasmodium falciparum*, the residues E182, Y184, and 204-VVV-206 are absolutely conserved in the UPF1 homologs (Fig. 1), and furthermore, a corresponding putative UPF2 homolog can be identified in each case (data not shown). The importance of these residues for interacting with UPF2 is also supported by the following phylogenetic data. In the intracellular parasite *Encephalitozoon cuniculi*, there are small deletions and substitutions in the CH-rich domain of the putative UPF1 (accession code Q8SR02) that would completely alter the putative UPF2-interacting surface. In particular, a part of loop L7 and β6 containing 204-VVV-206 is missing, and E182 and Y184 are mutated to, respectively, lysine and serine (Fig. 1). Moreover, the minimal genome of this parasite does not apparently code for UPF2. The absence of UPF2 and conserved UPF2-interacting residues in *E. cuniculi* UPF1 suggests first that if NMD occurs in this organism, it involves a different, UPF2-independent, mechanism than that normally invoked. Second, the conservation of the CH-rich domain of UPF1, despite the absence of UPF2, suggests that this domain has additional roles, perhaps related to conserved non-NMD functions of UPF1. A similar situation occurs in the putative UPF1 homologs of *Entamoeba histolytica* (XP_657569 and EAL50744) and *Giardia lamblia* (Q7QU62), where, again, UPF2-interacting residues E182, Y184, and 204-VVV are mutated (Fig. 1), correlating with the apparent absence of UPF2 in the available genome sequence. On the other hand, UPF2 homologs are apparent in the *T. brucei*, *T. cruzi*, and *P. falciparum* genomes.

In conclusion, the structure of the CH-rich domain of UPF1 is the first structural information on this key NMD factor and provides a starting point for further functional studies to help elucidate its exact functions in gene expression including NMD.

MATERIALS AND METHODS

Expression, purification, and crystallization of UPF1(115–272)

Human UPF1(115–272) was expressed in *E. coli* BL21Star(DE3) from a pProEXHTb expression vector as a His-tag fusion protein and was purified by affinity chromatography using Ni^{2+} resin. After His-tag cleavage with TEV protease, the protein was further purified by a second Ni^{2+} column and size-exclusion chromatography. Pure UPF1(115–272) was concentrated to ~6.5 mg/mL in a buffer containing 20 mM Tris (pH 7.0), 100 mM NaCl, and 10 mM β-mercaptoethanol for crystallization. Diffracting crystals grew within a week at 4°C in a condition containing 30% pentaerythritol ethoxylate (15/4EO/OH) (v/v), 50 mM ammonium sulfate, and 50 mM Bis-Tris (pH 6.5). For data collection at 100 K, crystals were snap-frozen in liquid nitrogen with a solution containing mother liquor and 30% (v/v) glycerol.

Data collection and structure determination

The UPF1(115–272) crystals belong to space group $P2_12_12_1$ with unit cell dimensions a = 64.6 Å, b = 73.0 Å, c = 73.2 Å and a solvent content of 48.2%. The crystals diffract to a resolution of 3.0 Å. Diffraction data were collected at 100 K using a Q4R ADSC CCD detector on beamline ID14-EH4 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and processed using XDS (Kabsch 1993). The structure was solved by a zinc single anomalous dispersion (SAD) experiment. Data were collected at the Zn K-absorption edge (X-ray wavelength 1.2821 Å) using the inverse beam data collection method to increase the accuracy of anomalous difference measurements. However, owing to ice rings and the small size of the crystal, data between 4 and 3 Å resolution are of lower quality. Using 4 Å resolution data, SHELXD (Uson and Sheldrick 1999) readily found six sites, corresponding to two molecules of UPF1 per asymmetric unit each containing three zinc atoms. These six sites were refined and used for phasing in SHARP (Bricogne et al. 2003). Phases were then extended to 3 Å with DM (Cowtan 1994). In the initial electron density map calculations, three helices and three β-strands per molecule were identified with the help of FFFEAR. Together with the positions of the zinc sites, this partial model served to establish the NCS operators and to define a mask that was then used for NCS averaging in DM to improve the electron density maps. By iterating this process, a complete model could eventually be built using O (Jones et al. 1991). During refinement, the two UPF1 molecules were constrained by tight noncrystallographic symmetry except for flexible regions 199–204 (L7) and 216–228 (α2 and L9), which differ significantly in conformation. Three TLS groups per subunit (residues 118–183, 184–240, 241–272) were defined with the help of the TLSMD server (http://skuld.bmsc.washington.edu/~tlsmd/). TLS parameters accounted for much of the individual *B*-factor variation and use of this procedure reduced the R_{free} by 2.3%. The final model, obtained with REFMAC5 (Murshudov et al. 1997), has an R_{free} of 25.9% and *R*-factor of 22.1% with good geometry and all residues in the favored (89.1%) or additionally allowed (10.9%) regions of the Ramachandran plot, as analyzed by PROCHECK (Laskowski et al. 1993). The final structure was also validated by calculating with CNS a systematic simulated annealing omit map covering the whole asymmetric unit. Detailed data and refinement statistics are in Table 1.

His-tag pull-down assays

UPF2(761–1207) and UPF1(115–272) were, respectively, cloned into a pProEXHTb expression vector in order to produce His-tag fusion protein and into pRSFDuet-1 (Novagen) to produce native protein without any tag. Mutations were generated using a QuickChange site-directed mutagenesis kit and confirmed by sequencing. His-tagged UPF1 and UPF2 were coexpressed in *E. coli* BL21Star(DE3), and the resultant complexes were purified using Ni^{2+} resin. The resin was extensively washed with 20 mM and 100 mM imidazole. Bound proteins were eluted with 250 mM imidazole and analyzed by 15% SDS-PAGE.

ACKNOWLEDGMENTS

We thank members of the EMBL-ESRF Joint Structural Biology Group, notably, Andrew McCarthy, for assistance with data collection on ESRF beamlines. We are grateful to Carlo Petosa, Luc Bousset, and Thibaut Crepin for their help in the crystallographic analysis. We also thank Lynne Maquat (University of Rochester) for providing hUPF1 cDNA. We thank Elisa Izaurralde for critical reading of the manuscript. Coordinates and structure factors have been deposited in the Protein Data Bank with entry code 2IYK.

Received June 6, 2006; accepted July 14, 2006.

REFERENCES

Andersen, P., Kragelund, B.B., Olsen, A.N., Larsen, F.H., Chua, N.H., Poulsen, F.M., and Skriver, K. 2004. Structure and biochemical function of a prototypical *Arabidopsis* U-box domain. *J. Biol. Chem.* **279:** 40053–40061.

Applequist, S.E., Selg, M., Raman, C., and Jack, H.M. 1997. Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein. *Nucleic Acids Res.* **25:** 814–821.

Bhattacharya, A., Czaplinski, K., Trifillis, P., He, F., Jacobson, A., and Peltz, S.W. 2000. Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA* **6:** 1226–1235.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* **59:** 2023–2030.

Conti, E. and Izaurralde, E. 2005. Nonsense-mediated mRNA decay: Molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* **17:** 316–325.

Cowtan, K. 1994. Joint CCP4 and ESF-EACBM. *Newsletter Protein Crystallogr.* **31:** 34–38.

Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., and Peltz, S.W. 1998. The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes & Dev.* **12:** 1665–1677.

de Pinto, B., Lippolis, R., Castaldo, R., and Altamura, N. 2004. Overexpression of Upf1p compensates for mitochondrial splicing deficiency independently of its role in mRNA surveillance. *Mol. Microbiol.* **51:** 1129–1142.

Domeier, M.E., Morse, D.P., Knight, S.W., Portereiko, M., Bass, B.L., and Mango, S.E. 2000. A link between RNA interference and nonsense-mediated decay in *Caenorhabditis elegans*. *Science* **289:** 1928–1931.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. 2005. SMG7 is a 14–3–3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol. Cell* **17:** 537–547.

Garman, E.F. and Grime, G.W. 2005. Elemental analysis of proteins by microPIXE. *Prog. Biophys. Mol. Biol.* **89:** 173–205.

Gonzalez, C.I., Ruiz-Echevarria, M.J., Vasudevan, S., Henry, M.F., and Peltz, S.W. 2000. The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. *Mol. Cell* **5:** 489–499.

He, F., Brown, A.H., and Jacobson, A. 1996. Interaction between Nmd2p and Upf1p is required for activity but not for dominant-negative inhibition of the nonsense-mediated mRNA decay pathway in yeast. *RNA* **2:** 153–170.

———. 1997. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol. Cell. Biol.* **17:** 1580–1594.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233:** 123–138.

Hosoda, N., Kim, Y.K., Lejeune, F., and Maquat, L.E. 2005. CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.* **12:** 893–901.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L.E. 2001. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106:** 607–617.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47:** 110–119.

Kabsch, W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26:** 795–800.

Kadlec, J., Izaurralde, E., and Cusack, S. 2004. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat. Struct. Mol. Biol.* **11:** 330–337.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. 2006. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes & Dev.* **20:** 355–367.

Kaygun, H. and Marzluff, W.F. 2005. Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nat. Struct. Mol. Biol.* **12:** 794–800.

Kim, Y.K., Furic, L., Desgroseillers, L., and Maquat, L.E. 2005. Mammalian Staufen1 recruits Upf1 to specific mRNA 3′UTRs so as to elicit mRNA decay. *Cell* **120:** 195–208.

Klein, D.J., Moore, P.B., and Steitz, T.A. 2004. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **340:** 141–177.

Krishna, S.S., Majumdar, I., and Grishin, N.V. 2003. Structural classification of zinc fingers: Survey and summary. *Nucleic Acids Res.* **31:** 532–550.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26:** 283–291.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. 2001. The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20:** 4987–4997.

Legge, G.B., Martinez-Yamout, M.A., Hambly, D.M., Trinh, T., Lee, B.M., Dyson, H.J., and Wright, P.E. 2004. ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. *J. Mol. Biol.* **343:** 1081–1093.

Lykke-Andersen, J. 2002. Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Mol. Cell. Biol.* **22:** 8114–8121.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. 2000. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103:** 1121–1131.

Maquat, L.E. 2004. Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* **5:** 89–99.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. 2001. Rent1, a *trans*-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* **10:** 99–105.

Mendell, J.T., ap Rhys, C.M., and Dietz, H.C. 2002. Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. *Science* **298:** 419–422.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53:** 240–255.

Nazarenus, T., Cedarberg, R., Bell, R., Cheatle, J., Forch, A., Haifley, A., Hou, A., Wanja Kebaara, B., Shields, C., et al. 2005. Upf1p, a highly conserved protein required for nonsense-mediated mRNA decay, interacts with the nuclear pore proteins Nup100p and Nup116p. *Gene* **345:** 199–212.

Ohi, M.D., Vander Kooi, C.W., Rosenberg, J.A., Chazin, W.J., and Gould, K.L. 2003. Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nat. Struct. Biol.* **10:** 250–255.

Petroski, M.D. and Deshaies, R.J. 2005. Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **6:** 9–20.

Takahashi, S., Araki, Y., Sakuno, T., and Katada, T. 2003. Interaction between Ski7p and Upf1p is required for nonsense-mediated 3′-to-5′ mRNA decay in yeast. *EMBO J.* **22:** 3951–3959.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25:** 4876–4882.

Unterholzner, L. and Izaurralde, E. 2004. SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol. Cell* **16:** 587–596.

Uson, I. and Sheldrick, G.M. 1999. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9:** 643–648.

Weng, Y., Czaplinski, K., and Peltz, S.W. 1996. Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Mol. Cell. Biol.* **16:** 5491–5506.

Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. 2000. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102:** 533–539.

Postdoctoral work at the University of Oxford (2005-2007)

Structural analysis of viral fusion proteins

I did my postdoctoral work at the Structural Biology Division (STRUBI) at the University of Oxford. I decided to change the research topic and moved from studying eukaryotic RNA degradation to structural analysis of viral proteins. The department and the laboratory of Prof. David Stuart were at that time at the forefront of the structural virology research and structural genomics. While the OPPF (Oxford Protein Production Facility) provided cutting edge high-throughput approaches, STRUBI had just implemented a very efficient pipeline of mammalian cell production of glycoproteins.

The main focus of my work was on structural characterization of several viral fusion glycoproteins. Viral fusion proteins are essential for the cell entry of enveloped viruses as they mediate fusion of the host and viral membranes preceding the delivery of the viral genetic material into the host cell. Only several structures of these proteins were known at that time, due to crystallization difficulties associated with their large size and glycosylation. Main interest of the field was to understand how do viral fusion proteins mediate membrane fusion through undergoing important conformational changes. Another important question was whether different types of fusion proteins evolved independently or they diverged from a common ancestor protein.

The main outcome of my postdoctoral work was a crystal structure the postfusion form of the baculovirus glycoprotein GP64, which we determined in collaboration with Prof. Ian Jones of the University of Reading (Kadlec et al., 2008). GP64 is unusual in promoting viral entry into both insect and mammalian cells and is distinct from the previously established class I (such as influenza hemagglutinin or paramyxovirus F protein) and class II fusion proteins (fusion proteins of the alphaviruses and flaviviruses).

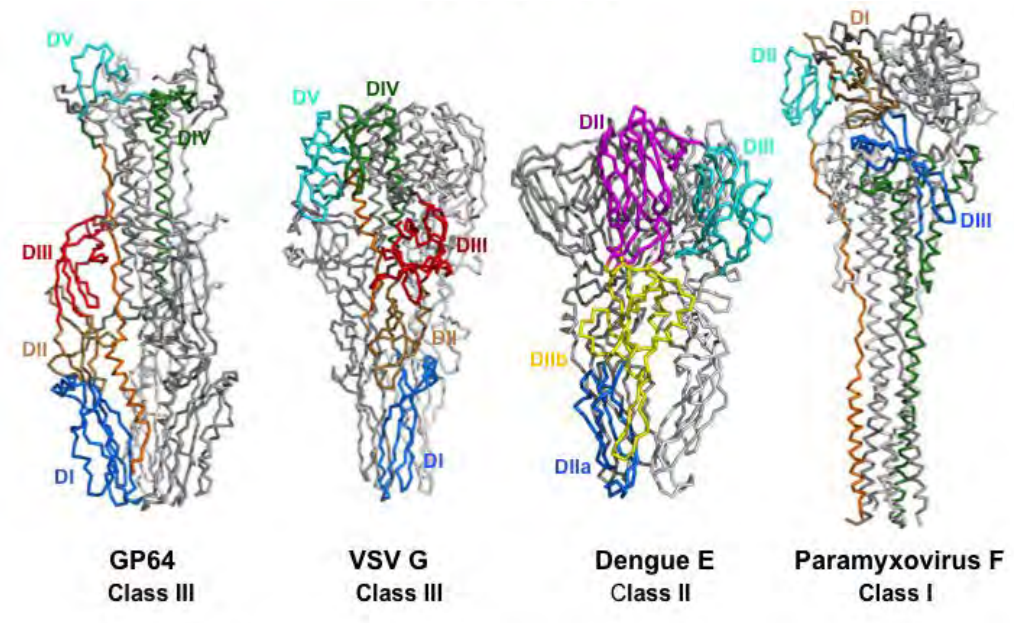


Figure 2. Comparison of GP64 postfusion conformation structure with known structures of the three classes of viral fusion. Related domains are shown in the same colour.

The structure explained several of GP64's biological properties, including its cellular promiscuity and pH-dependent conformational change triggered by differential protonation of key histidine residues. GP64 turned out to be the third representative of a newly established class (III) of fusion proteins as it exhibited unexpected structural homology with the vesicular stomatitis virus G and herpes simplex virus type 1 gB proteins described around that time. We identified fusion peptides for

GP64 by mutagenesis and cellular assays. Using the structure for a systematic analysis of known viral fusion proteins we established domain-level similarities that cut across defined categories of fusion proteins suggesting a possible common origin of viral fusion machines (Kadlec et al., 2008) (Figure 2). During crystallization trials on GP64 I also participated in developing an automated microseeding protocol using a crystallization robot (Walter et al., 2007).

Using mammalian cells, I produced and crystallized ectodomains of Hepatitits C virus E1 and Bovine viral diarrhea virus (BVDV) E2 glycoproteins, whose structures were later determined by Kamel El Omari (El Omari et al., 2014a, 2014b; Iourin et al., 2013).

nature structural & molecular biology

The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines

Jan Kadlec[1,3], Silvia Loureiro[2], Nicola G A Abrescia[1], David I Stuart[1] & Ian M Jones[2]

Viral fusion proteins mediate the merger of host and viral membranes during cell entry for all enveloped viruses. Baculovirus glycoprotein gp64 (gp64) is unusual in promoting entry into both insect and mammalian cells and is distinct from established class I and class II fusion proteins. We report the crystal structure of its postfusion form, which explains a number of gp64's biological properties including its cellular promiscuity, identifies the fusion peptides and shows it to be the third representative of a new class (III) of fusion proteins with unexpected structural homology with vesicular stomatitis virus G and herpes simplex virus type 1 gB proteins. We show that domains of class III proteins have counterparts in both class I and II proteins, suggesting that all these viral fusion machines are structurally more related than previously thought.

Cell entry is key to the life cycle of all viruses, requiring, for enveloped viruses, the fusion of viral and host membranes. Viral fusion proteins drive this reaction via a major conformational change triggered by low pH in the endosome or interaction with cellular receptors[1]. Class I fusion proteins, including influenza virus hemagglutinin and paramyxovirus F protein, form trimers in both conformations[2,3] and undergo proteolytic cleavage before formation of a metastable state, primed for conformational change. Class II fusion proteins, including flavivirus E and alphavirus E1 proteins, exist as dimers in the prefusion state and trimers in the postfusion conformation[4,5]. They are primed for fusion by cleavage of a second protein with which they fold. Vesicular stomatitis virus (VSV) G and herpes simplex virus type 1 (HSV-1) gB glycoproteins, which are trimeric in both the pre- and postfusion states and resemble each other, form class III fusion proteins, distinct from class I and II proteins[6–8]. In VSV G, the conformational change is reversible, removing the need for proteolytic cleavage to avoid preactivation[8]. All classes of fusion proteins attach to the viral membrane via C-terminal transmembrane helices and to the host membrane via fusion peptides, at the N terminus of class I and in a loop or loops in class II and III. The postfusion form of class I and III proteins is characterized by a central three-stranded coiled coil, often sheathed by antiparallel C-terminal helices.

Baculoviruses are large double-stranded DNA viruses that infect a wide variety of invertebrates, a type member being *Autographa californica* multiple nucleopolyhedrosis virus (AcMNPV)[9]. The major envelope protein of AcMNPV, glycoprotein 64 (gp64), is necessary and sufficient for cell entry, fulfilling receptor binding and low pH–triggered membrane fusion roles and facilitating baculovirus entry into many mammalian cell types, including adult mesenchymal and embryonic stem cells[10–13], suggesting baculoviruses as useful gene-delivery vectors[14]. Heterologous sequences added to the N terminus of gp64 are incorporated into budding virions[15] and are highly immunogenic[14,15]. The properties of gp64 are distinct from class I and II fusion proteins: it does not require proteolytic activation (the pH-dependent conformational switch is reversible)[10,16], and it forms discrete peplomers ~10-nm tall on one or both poles of the virus[17], similar to class I but not class II proteins[18], but forms fusion pores distinct from those of class I glycoproteins[19]. gp64 consists of 512 amino acid residues and comprises a large glycosylated N-terminal ectodomain, a transmembrane helix (residues 483–504) and short cytoplasmic tail. gp64 has sequence similarity with gp75, the envelope protein of thogotoviruses (insect-transmitted orthomyxoviruses), but has no other sequence homologs. Here we present the structure of the ectodomain of gp64 in its postfusion conformation, showing that it resembles the class III fusion proteins. The structure suggests the molecular basis of fusion in many cell types and explains the ability of gp64 to incorporate foreign polypeptides. We confirm the identification of unexpected fusion peptides for gp64 by mutagenesis and cellular assays. Finally, using this structure for a systematic analysis of known viral fusion proteins, we establish domain-level similarities that cut across currently defined categories and suggest a possible underlying unity in the origins of currently known viral fusion machines.

RESULTS

Structure determination and overall structure

The structure of the deglycosylated ectodomain (residues 1–499) of AcMNPV gp64, expressed in insect cells, was solved by X-ray crystallography at 2.95-Å resolution, phased by MAD analysis of a platinum derivative and refined to an R_{free} of 26.9% and R-factor 22.4% with all residues lying in allowed regions of the Ramachandran plot

[1]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [2]School of Biological Sciences (AMS), University of Reading, Reading RG6 6AJ, UK. [3]Present address: EMBL Grenoble, 6 rue Jules Horowitz, 38042 Grenoble, France. Correspondence should be addressed to D.I.S. (dave@strubi.ox.ac.uk).

Received 29 February; accepted 1 August; published online 7 September 2008; doi:10.1038/nsmb.1484

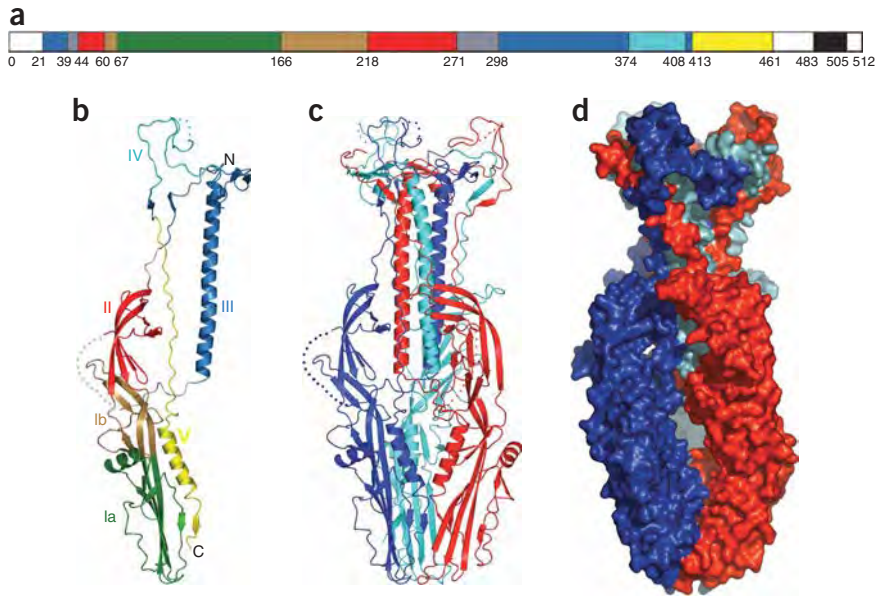


Figure 1 Overall structure of gp64. (**a**) Schematic representation of gp64 domain structure. Domains are colored according to the protomer structure shown in **b**. The signal sequence (residues 1–20) and membrane proximal C-terminal region (residues 461–482) not observed in the structure, as well as the cytoplasmic tail, are in white. The predicted transmembrane region is shown in black. (**b**) Ribbon diagram of a gp64 protomer colored by domains. Within domain I the two subdomains Ia and Ib are shown. The disordered segment connecting domain Ib and II and the one within domain IV are shown as dots. The domain nomenclature corresponds to that of the HSV-1 gB structure[6]. (**c**) Ribbon diagram of the gp64 trimer. The protomer shown in blue corresponds to that in **b**. (**d**) Surface representation of the gp64 trimer in the same orientation as in **b**.

(Methods). The gp64 structure is an elongated (150 × 55 Å) trimer, with each subunit being composed of five domains (I–V; **Fig. 1** and **Supplementary Fig. 1** online). Electron density is missing for the C-terminal residues 461–499, the domain II–III linker (residues 272–287) and part of domain IV (residues 394–398). The structure is rigidified by seven disulfide bridges, one intermolecular, and is composed mostly of β-sheets with a prominent central triple-stranded coiled coil that is involved in trimerization (**Fig. 1** and **Supplementary Figs. 1** and **2** online).

Baculovirus gp64 is a class III fusion protein

Although there are differences in the overall molecular shape and relative domain orientations, and the gp64 central helix is swapped between subunits, the individual domains are similar to those of the postfusion forms of VSV G and HSV-1 gB glycoproteins[6,7], suggesting (together with the fact that the crystals were grown at pH 5) that this represents the postfusion conformation of gp64 (**Supplementary Fig. 3** online) and that although the three proteins share no sequence homology and come from unrelated viral families, they belong to the same class. The extent of the similarity (discussed in more detail below) suggests that the proteins in this class share a common ancestor. In addition, the 26% sequence identity between gp64 and gp75 of Thogoto virus (which maps most strongly to domains Ia, III, IV and V) suggests that fusion proteins of this genus also have similar structures (**Supplementary Fig. 2**). Class III fusion proteins thus seem to be widely used among viral families, presumably having been disseminated by horizontal gene transfer.

Structure of gp64 domains

gp64 comprises five domains or regions (**Fig. 1b** and **Supplementary Fig. 1**). Domain I (residues 60–217), located at one end of the molecule, consists of subdomains Ia and Ib, which share two β-strands. Ia consists of a long four-stranded antiparallel β-sheet packing against helix A, and Ib is a six-stranded β-barrel. Within the trimer, the Ia domains make contact only through loops L1 and L2 (Asn78 and Asn151 form hydrogen bonds with the main chain carbonyls of Ser81 and Pro84). Compared to VSV G, both gp64 and HSV-1 gB have a long insertion in Ia downstream of helix A. In gp64, this insertion (residues 111–138) forms three short strands and contributes the fourth strand to the β-sheet. Conversely, Ib is more similar to its counterpart of VSV G than to the corresponding pleckstrin homology (PH) domain of HSV-1 gB, which harbors a long insertion.

Domain II (residues 44–59 and 218-271) is formed by two four-stranded antiparallel β-sheets, each stabilized by a disulfide bridge. Compared to the corresponding PH domains of VSV G and HSV-1 gB, a deletion in gp64 between residues 257–261 removes a β-hairpin and long helical segment. A disordered linker region after Lys271 (residues 272–287) forms part of the epitope of neutralizing antibody AcV1 in the prefusion form of gp64 (ref. 16). Domains Ib and II are similar to each other (see below).

Domain III (residues 22–39, 298–373 and 408–412) comprises helices B and C and a four-stranded β-sheet. Helix B (299–241) forms a triple-stranded α-helical coiled coil at the heart of the trimer. Packed in layers at the core of the coiled coil are 11 hydrophobic residues and a single polar residue, generally with a heptad (3-4) periodicity (**Supplementary Fig. 2**), although a stutter (3-4-4-3) occurs between residues 309–323. Stutters produce local underwinding of the supercoil and may terminate coiled-coil regions[20]; thus, the N-terminal part of helix B may partly unfold in the prefusion state. The β-sheet of domain III combines two strands from protomer A with a C-terminal strand of protomer B and an N-terminal strand of protomer C (nomenclature defined in **Supplementary Figure 1**), and is stabilized by an intermolecular disulfide bridge between Cys24 and Cys372 (consistent with biochemical data[21]). The N terminus is located on the top of this sheet and is exposed, whereas in other class III structures, the chain continues and participates in domain IV, explaining the ability of gp64 to display foreign proteins at this location[15]. The N-terminal residues 36–39 stretch along helix B of the neighboring protomer, forming charged contacts between Asp37, Leu38 and Asn39 and His335, Asn332 and Asn228.

Domain IV (residues 374–407) is much smaller than in VSV G and HSV-1 gB and is poorly defined, with no electron density for residues 394–398 and only weak density for the remainder (which are presumably somewhat stabilized by a disulfide bridge between Cys382 and Cys402). Domain V (residues 413–460) connects domain IV with domain I and contributes to trimerization. It consists of a long segment that packs between the B helices of the other subunits and the C-terminal helix D, with extensive interactions with domain I of the same protomer. The C-terminal strand forms a small sheet with a strand of domain I.

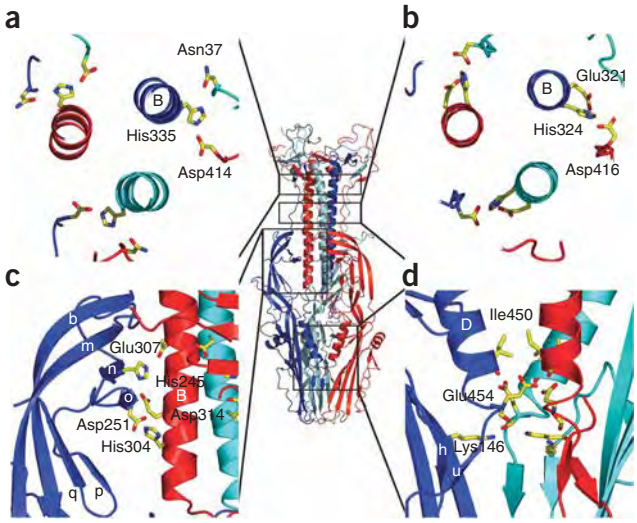


Figure 2 pH-sensitive interfaces of gp64. Four pH-sensitive trimerization areas of gp64, shown in **a**–**d**, are localized on the gp64 trimer by the black boxes in the central panel. (**a**) Protonated His335 of heptad 6 forms a salt-bridge interaction with Asp414 and a hydrogen bond with Asn37. (**b**) Salt-bridge interactions of protonated His324 of heptad 4 with Asp416 and Glu321. For clarity, the hydrogen-bond interaction of His326 with the carbonyl of Asp415 is not shown. (**c**) pH-sensitive interaction interface between the central stalk and a β-sheet (b,m,n,o) of domain II, including salt-bridge interactions between protonated His245 with Glu307 and Asp314, and between His304 and Asp251. (**d**) Trimerization interface formed by the C-terminal D helices, including hydrogen bonds between protonated Glu454 residues and main chain carbonyls of Ile450. Additionally, the hydrogen-bond interaction of the Glu454 main chain carbonyl with Lys146 is shown.

pH-dependent conformational change

There are several trimerization interfaces within gp64, including the central helical stalk, packing of domain II against the helix B, contacts between domain Ia and the C-terminal D helices and the mixed β-sheet of domain III. By analogy with VSV G, we expect massive, reversible, structural changes between the pre- and postfusion states of gp64, requiring disruption of some of the trimer contacts, during which, however, the molecule is constrained to remain trimeric by the disulfide bridge linking adjacent subunits via domain III. This suggests that other class III molecules may remain trimeric during the transformation, explaining the cooperative kinetics observed for this process[8].

Reversibility may be linked to the necessity of recovering a high-pH structure after passing through acidic Golgi compartments[7]. In VSV G, this low-pH to high-pH structure switch is largely controlled by deprotonation of several acidic residues[7], whereas the protonation of specific histidine residues has been proposed as a trigger in the high-pH to low-pH structure switch of several systems[8,22]. The structure reported here provides compelling evidence for such a mechanism, with the low-pH to high-pH conformational change of gp64 controlled by the deprotonation of numerous histidine residues, explaining how an apparently robust subunit interface (5,030 Å^2 surface buried between two subunits) can be rendered pH sensitive. Details of these switches are given in **Figure 2**. In brief, a cluster of nine histidine residues (three from each protomer, with calculated pK_as of 4, 6.6 and 12; **Fig. 2a**,**b**) dominates the upper portion of the central stalk, whereas the β-sheet of domain II buttresses the central region by burying two protonated histidine residues (with pK_as of 8.8 and 9; **Fig. 2c**), and at the bottom, His430 (domain V, calculated pK_a 6.7) forms a salt bridge with Asp301 (helix B). The low-pH to high-pH conformation switch in gp64 is probably achieved by the cooperative deprotonation of these residues, triggered by conformational changes of those with the lowest pK_a and leading to the disassembly of the lateral domains and the N- and C-terminal portions of the coiled coil. Finally, the trimeric base of gp64 would be destabilized by deprotonation of Glu454 (hydrogen-bonded to the carbonyl of Ile450, calculated pK_a 7; **Fig. 2d**).

Identification of fusion loops

Structural arguments suggested that the fusion peptides of VSV G and HSV-1 gB are located on the two loops at the tip of domain I[6,7]. VSV G possesses some hydrophobic residues at these positions, and mutation of one (A117K) abolishes fusion[23]. In HSV-1 gB, the putative fusion loops bear mostly charged residues, although mutation of buried hydrophobic residues in this region also affects fusion, suggesting local conformational changes during fusion[24]. Hydrophobic residues in corresponding fusion loops are crucial for fusion also in the Epstein Barr virus gB protein[25]. In gp64, double mutations of Leu229 and Leu230 to alanine or aspartic acid abolish membrane fusion, as measured by syncytium formation[26]. These residues are

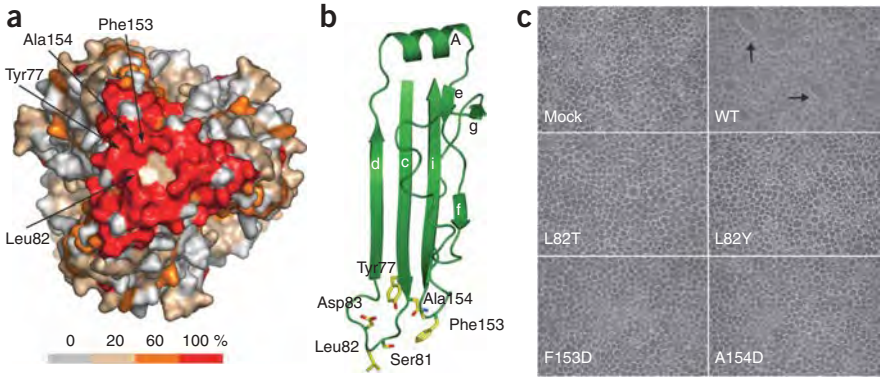


Figure 3 gp64 fusion loops. (**a**) Surface representation of the gp64 trimer, viewed from the membrane along the three-fold axis, highlighting areas of conserved surface residues. The conservation of the surface residues, based on the sequences alignment shown in **Supplementary Fig. 2**, is represented from gray to red (red is 100% conserved) according to the color scale bar. Tyr77 of strand c, also labeled, might be involved in membrane fusion in addition to the highlighted hydrophobic residues of the putative fusion loops. (**b**) The fusion domain of gp64 (domain Ia). The hydrophobic residues of loops L1 and L2 are shown in yellow. Additionally, charged residues Asp83 and Ser81 are shown. (**c**) Syncytium-formation assay. Syncytia formed following the decrease of pH to 5 for wild-type gp64 (WT) are shown with arrows. When *Sf*9 cells were transfected with any of the mutants indicated in the figure or an empty vector (mock), no syncitia were formed after pH was lowered to 5. No syncytium formation was observed for L82Y, F153D, S81L and D83F and for all samples at neutral pH (data not shown).

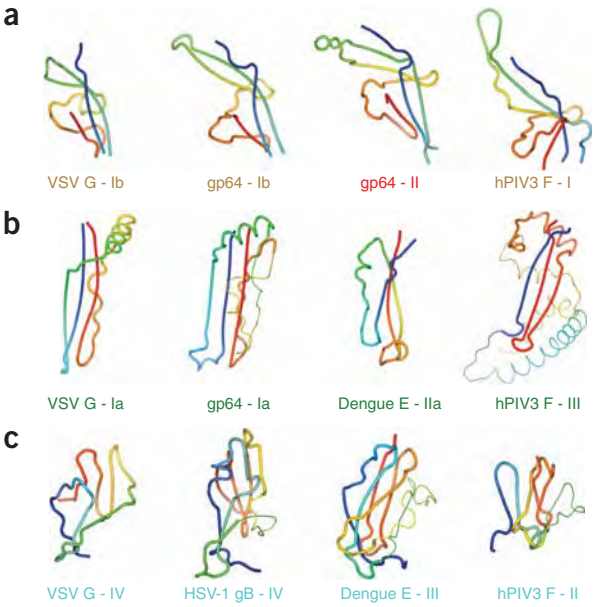


Figure 4 Structural similarities among the class I, II and III fusion proteins domains. The color of the labels corresponds to the color coding of the trimer molecules shown in **Supplementary Figure 5**. The domain nomenclature of VSV G was adjusted to that of gp64 and HSV-1 gB (**Supplementary Fig. 5**). Domains are colored in a gradient from blue (N terminus) to red (C terminus). For comparison purposes, core domains were used that can partially overlap within the full-length proteins. (**a**) Comparison of gp64 domains Ib (56–68, 165–214) and II (44–59, 218–271) with domain Ib of VSV G (50–60, 128–170) and paramyxovirus (human parainfluenza virus 3 (hPIV3)) F protein domain I (29–40, 286–341). (**b**) Comparison of gp64 (67–166) and VSV G (58–130) domain Ia with domain IIa of dengue virus E protein (64–120) and paramyxovirus F domain III (41–285). For clarity, the insertion in gp64 (111–138) and paramyxovirus F protein (56–239) are shown as thin tubes. (**c**) Comparison of VSV G (331–384) and HSV-1 gB (574–661) domains IV with dengue virus E protein domain III (299–394) and paramyxovirus domain II (372–429). The insertion in HSV-1 gB, dengue E and paramyxovirus proteins are shown as thin tubes.

located on a β-strand of domain II (**Fig. 1** and **Supplementary Fig. 2**), a position that is inconsistent with a direct role in fusion and is more likely to affect the conformational switch between pre- and postfusion states. However, mapping residues conserved across baculovirus gp64s and thogotovirus gp75s onto the surface of gp64 reveals a major concentration in the loops at the tip of domain I, with Leu82 of L1 and Phe153 and Ala154 of L2 forming an exposed hydrophobic patch, consistent with these residues having fusion activity (**Fig. 3a**,**b**).

We mutated these residues independently into charged residues and tested the ability of the gp64 mutants to fuse cell membranes in a syncytium-formation assay using insect *Sf*9 cells. Whereas cells transfected with wild-type gp64 showed clear syncytia after the pH was transiently decreased to 5, we observed no syncytia using cells transfected with an empty vector or any of the mutants A154D, L82T, L82Y, F153D or F153T (the last mutant introduced a new N-glycosylation site) (**Fig. 3c**). Fusion activity was also abolished when two polar residues in the fusion loop were mutated to hydrophobic residues (S81L, D83F), whereas a H155F mutant showed similar fusion activity to the wild type. Western blot analysis using an anti-gp64 monoclonal antibody showed that all the mutants were expressed (**Supplementary Fig. 4a** online), and cell-surface staining revealed that all mutant gp64 protein reached the cell surface, albeit with different efficiencies (**Supplementary Fig. 4b**). In addition, when the ectodomains of each mutant were expressed in HEK 293T cells, all were secreted to the same extent as the wild type, confirming that the mutated proteins were folded (**Supplementary Fig. 4c**).

We suggest that the class II and III fusion proteins share a common positioning for the fusion peptides, at the tip of the β-domains in the trimeric postfusion state. Furthermore, in gp64, these residues are involved in cell attachment because virus binding to insect cells in the presence of antisera generated to overlapping gp64 peptides indicates that residues 121–160 contain a putative receptor binding function, with Phe153 being a key residue[27]. To accommodate both observations, we suggest that the hydrophobic patch at the tip of the postfusion trimer acts both as receptor binding site and fusion peptide. Adherence to the plasma membrane at neutral pH could be effected by a small subset of gp64 trimers with a transiently extended domain I. Embedding of the extended hydrophobic side chains into the plasma membrane would anchor the virus long enough to trigger endocytosis, and at low endosomal pH, all molecules would be driven through this intermediate toward the postfusion state, drawing the viral and plasma membranes together. This is consistent with the inhibition of baculovirus cell entry by a number of lipids[28] and with the high multiplicities of virus required for entry into some mammalian cells[14], whose plasma membrane composition varies substantially from insect cells.

Class III proteins show similarities to classes I and II

Despite the fact that the membrane fusion mechanism proposed for all currently known viral fusion proteins is similar, differences in biochemical properties (for example, oligomeric state and proteolytic cleavage) and structural differences led to the assumption that class I and class II viral fusion machines are structurally and evolutionarily distinct. However, it is now clear that there are few reliable markers of class. For instance, within the class I fusion proteins, the conformational change in paramyxovirus F is not pH dependent and does not require proteolytic cleavage (unlike the influenza virus hemagglutinin, which, apart from a common central coiled coil, is also structurally very different)[2,29], and, although the class II proteins are more similar, flavivirus fusion proteins form homodimers in the prefusion form, whereas alphaviruses form E1–E2 heterodimers[1].

We used the structure of gp64 to define the core domains of the class III proteins, detect regions where differences occur and compare these with the other classes (**Fig. 4** and **Supplementary Fig. 5** online). Some of the similarities are too weak to be reliably detected by the commonly used servers (that is, the Dali and SSM servers), which rely on distance matrices or secondary-structural assignments, and we have therefore used a method, based on ideas put forward previously by Rossmann and Argos (ref. 30), that takes into account connectivity, residue positions and local conformation together implemented via dynamic programming with a simple gap penalty function[31]. In cases of distant similarity, there is no agreed measure of significance and, although further work is justified to put such comparisons on a robust statistical footing, we believe that, at present, morph videos (generated with the aid of an unpublished program by J. Diprose and available as **Supplementary Videos 1–4** online) provide the best assessment.

Domains Ib and II of gp64 are similar to each other (r.m.s. deviation 3.2 Å for 59 Cα atoms; **Fig. 4a**) and can be superimposed

on the generally more elaborate corresponding PH domains of VSV G and HSV-1 gB (**Fig. 4a**). gp64 Ib seems to represent the minimal core of these domains, which are found as structural repeats in all class III molecules. A similar domain in paramyxovirus F protein (domain I) can be superimposed on gp64 domain Ib (r.m.s. deviation 4.3 Å for 51 Cα atoms; although the orientation of the two C-terminal strands differs, the four-stranded β-sheet is preserved; **Fig. 4a** and **Supplementary Video 2**). Similarly to gp64 domains Ib and II, this paramyxovirus domain consists of two segments with the same topology.

gp64 and HSV-1 gB have a large insertion downstream of the helix in the fusion domain Ia compared to the VSV G structure (**Fig. 4b**). The apparent similarity between fusion domains of class III and class II proteins was suggested to be a striking case of convergence, on the basis of a different topology of the β-sheet[7,32]. However, VSV G fusion domain Ia can be superimposed onto the similar structure in dengue virus E domain IIa with correct β-sheet topology (4.4 Å r.m.s. deviation for 36 Cα atoms; **Fig. 4b** and **Supplementary Video 3**). In dengue virus E, the helix packs on the opposite side of the long β-sheet, but in every case, the fusion peptides are located on the loop(s) at the tip of the β-sheet. Additionally, a related structure also occurs within domain III of the paramyxovirus F protein, which is similar to dengue virus E fusion domain IIa, including the helix packing (r.m.s. deviation 4.3 Å for 39 Cα; **Fig. 4b** and **Supplementary Video 4**). Instead of one strand and a loop corresponding to the fusion loop of gp64, the central helical segment is inserted. The remaining β-sheet is complemented by a strand at the C terminus. The paramyxovirus domain III is inserted into domain I in the same way as gp64 domain Ia into domain Ib, further strengthening this similarity (**Fig. 4a**,**b**).

gp64 domain IV is greatly reduced compared to VSV G and HSV-1 gB (whose domains are similar and can be superimposed**,** with r.m.s. deviation 3.6 Å for 47 Cα atoms (**Fig. 4c**); the major difference is an insertion of a short β-hairpin (residues 616–627) in HSV-1 gB). Both dengue virus E domain III and paramyxovirus F domain II have an immunoglobulin-like fold that can be superimposed on HSV-1 gB domain IV (for example, r.m.s. deviation 4.6 Å for 57 Cα atoms for dengue virus E2; **Fig. 4c** and **Supplementary Video 1**) but possess an elongated version of the HSV β-hairpin insert and lack a C-terminal strand. All these fusion proteins also share similar positioning of the C-terminal chain downstream of this domain.

Finally, domain III of class III proteins is markedly similar to the postfusion form of influenza hemagglutinin (**Supplementary Fig. 6** online). This domain (almost the entire hemagglutinin HA2 fragment)[2] contains one long and one short helix forming a six-helical bundle within the trimer, followed by a β-sheet consisting of a β-hairpin and another strand that in gp64 comes from the adjacent protomer, whereas in influenza hemagglutinin, it has an inverted orientation and comes from HA1.

These data suggest that essentially all domains of class III fusion proteins have counterparts in class I proteins (paramyxovirus F or influenza hemagglutinin). In gp64 domain, Ib is duplicated (domain Ib and II) and the central helix is located between domain II and III, whereas in paramyxovirus F, the corresponding domain I is present as a single copy and the central helix is inserted in a loop of domain III. The flexibility in domain organization of these proteins is exemplified by gp64, where domain Ia is inserted in Ib and the whole of domain I then inserted in domain II. Furthermore, both class I and III proteins share similarities with the fusion domain and domain III of class II proteins, suggesting extensive relationships between all three classes of fusion proteins.

DISCUSSION

The structure of gp64 reveals another member of the class III viral fusion proteins, suggesting that these are widely used viral fusion machines. The details of the route for the conformational transition in any fusion system remains elusive; however, the postfusion structure of gp64 provides compelling evidence that protonation of specific histidine residues acts as a pH-dependent trigger, with the low-pH to high-pH conformational change being controlled by the deprotonation of numerous histidine residues, rendering an apparently extremely robust subunit interface pH sensitive. In line with other class III molecules, we have identified the fusion loops as being at the tip of gp64, which, however, seems to be also responsible for cell attachment, suggesting that the virus has dispensed with a separate attachment mechanism. Lack of a high-affinity receptor is consistent with the lifestyle of the virus[9], where initial virus delivery occurs by targeted release from polyhedra in the insect gut, with gp64 simply facilitating subsequent cell-to-cell spread. This suggests that further mutation of residues postulated to be responsible for fusion might be a straightforward route to providing a virus better suited for gene therapy and stem-cell infection applications. Finally, by performing careful domain-by-domain structural comparisons, we find that all three classes of fusion molecules are structurally more related than has been previously suggested. It seems that class III, rather than being another independent class, might in fact link class I and II together. Further structural and functional work may clarify the origins of these pervasive molecules.

METHODS

Expression, purification and crystallization of gp64 (1–499). Expression of gp64 (1–499) in insect cells, its purification and deglycosylation with peptide N-glycosidase F (PNGase F) has been described previously[33]. The pure protein was concentrated to ~2 mg ml^{-1} in a buffer containing 10 mM Tris, pH 7.4, 10 mM NaCl and 0.1 mM EDTA for crystallization. At first, very small crystals were obtained within 1 week at 20 °C in conditions of 10% (v/v) PEG 6000 and 100 mM citric acid, pH 5.0. Using a semiautomated microseeding technique into additive crystal screens, larger crystals were grown in 10% (v/v) PEG 6000 and 100 mM citric acid, pH 5.0, a buffer that could be used for diffraction experiments[34]. A crystal grown with the addition of 200 mM potassium sodium tartrate was soaked for 3 h in a solution containing mother liquor and ethyl mercury phosphate (EMP) before the data collection. As no anomalous signal was observed in the diffraction data, this data set was used as a native one. A crystal grown in presence of 6% (v/v) 1,5-diaminopentane was soaked for 1 h in a solution of mother liquor and K_2PtCl_4 and used for MAD experiments. For data collection at 100 K, crystals were snap frozen in liquid nitrogen with a solution containing mother liquor and 30% (v/v) glycerol.

Data collection and structure determination. The gp64 (1–499) crystals belong to space group *H*32, with unit cell dimensions a,b = 87.5 Å and c = 431.6 Å (hexagonal settings), a solvent content of 56% and one molecule in the asymmetric unit. The crystals diffract to a resolution of 2.95 Å. Diffraction data were collected using a MARCCD detector on UK MAD beamline BM14 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and processed using XDS[35] and HKL2000 (ref. 36). The structure was solved by a platinum MAD experiment. A 'native' data set with a resolution of 2.95 Å was collected at the wavelength of the Hg LI-edge on the crystal soaked in EMP (1.0039 Å). Data sets with a resolution of 3.3–3.75 Å were collected at wavelengths corresponding to the peak wavelength of the Pt LI-edge and peak and inflection point wavelengths of the Pt LIII-edge on the crystal soaked in K_2PtCl_4 (0.8856 Å, 1.0714 Å and 1.0711 Å, respectively).

Positions of seven platinum sites were identified, refined and used for phasing in autoSHARP[37]. After solvent modification with SOLOMON[38], we obtained electron density with clear secondary-structure features covering most of the molecule. Because the native and MAD data sets were not isomorphous, the phases could not be directly transferred to the higher-resolution data set.

Table 1 Data collection and refinement statistics

	gp64 native	gp64 Pt LI peak	gp64 Pt LIII inflection point	gp64 Pt LIII peak
Data collection				
Space group	*H*32	*H*32	*H*32	*H*32
Cell dimensions				
a, *b*, *c* (Å)	87.5, 87.5, 431.6	86.9, 86.9, 425.6	87.1, 87.1, 426.2	87.2, 87.2, 426.4
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	50–2.95, (3.05–2.95)[a]	25–3.3, (3.42–3.3)	30–3.3, (3.42–3.3)	30–3.75, (3.88–3.75)
R_{merge}	11 (55)	15 (–)	14.3 (–)	17 (81)
I / σI	16.29 (3.29)	46.4 (5.8)	38.5 (3.3)	37.1 (9.4)
Completeness (%)	97.3 (99.9)	91.3 (100)	90.9 (100)	96.3 (100)
Redundancy	6.8 (6.5)	10.6 (11.1)	9.4 (9.8)	10.6 (11)
Refinement				
Resolution (Å)	50–2.95			
No. reflections	12,936			
R_{work} /R_{free}	22.4/26.9			
No. atoms				
Protein	3,360			
Hg ions	5			
B-factors				
Protein	50.9			
Ligand/ion	64.9			
R.m.s. deviations				
Bond lengths (Å)	0.008			
Bond angles (°)	1.087			

[a]Values in parentheses are for highest-resolution shell.

A partial model was built at a 3.3-Å resolution with COOT[39]. The experimental electron density calculated with MAD phases covering this starting model was cut out in GAP (D.I. Stuart & J.M. Grimes, unpublished program) and used for molecular replacement into the native data set with PHASER[40]. After solvent flattening with DM[41], an improved model could be built at a resolution of 2.95 Å. Repeating this procedure with the model, a substantially better electron density map was obtained and side chains could be assigned. The model was completed with COOT and refined to R_{free} 26.9% and *R*-factor 22.4% using BUSTER[42] and REFMAC5[43] with good geometry and all residues in the allowed (96.12% in favored) regions of the Ramachandran plot, as analyzed by MolProbity[44]. The model also includes five Hg atoms with partial occupancy originating from the soaking in EMP. Detailed data and refinement statistics are given in **Table 1**.

Structure analysis. The superimpositions of the individual domain were performed using SHP[31]. The Supplementary Videos are based on an interpolation method (J. Diprose, unpublished program) using structural superposition of the particular domains and a list of equivalencies generated with SHP. The ribbon diagrams and actual videos were made with PyMol (http://pymol.sourceforge.net). The pK_a values of charged residues of gp64 were calculated using PROPKA[45]. The gp64 domain topology in **Supplementary Figure 1** was created with TopDraw[46]. **Supplementary Figure 2** was generated with CLUSTALX[47] and ESPript[48], and the heptad positions were assigned using TWISTER[49].

gp64 mutagenesis and syncytium-formation assays. Full-length gp64 was cloned into the baculovirus early promoter–based vector pIEx/Bac-1 for transient expression in insect cells. Mutations were generated using a QuickChange site-directed mutagenesis kit and confirmed by sequencing. Transfected *Sf*9 cells were rinsed once with PBS adjusted to pH 5.0 (with citric acid) and then incubated in PBS at pH 5.0 for 5 min at 22 °C. The acidic buffer was replaced with fresh tissue-culture medium and the cells were incubated for a further 2 h at 27 °C. Syncytium formation was assessed by microscopy with two independent observers. gp64 expression during the experiment was monitored by western blotting using the monoclonal anti-gp64 antibody B12D5. In addition, cells were analyzed for cell-surface staining by flow cytometry 2 d post-transfection using the B12D5 antibody and an anti-mouse phycoerythrin conjugate. The wild-type and mutated gp64 ectodomains were expressed as His-tag fusions in HEK 293T cells from pLHsec vector, and supernatant was analyzed by a western blotting using anti–His-tag antibody.

Accession codes. Protein Data Bank: coordinates for the crystal structure of baculovirus gp64 have been deposited with accession code 3DUZ.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

The work was supported by grants from the UK Medical Research Council to D.I.S. and I.M.J. J.K. was supported by a long term fellowship from the European Molecular Biology Organisation (EMBO). We thank M. Walsh for assistance during the data collection at the UK MAD beamline BM14 at the European Synchrotron Radiation Facility and R. Aricescu for help with syncytium-formation assays. We also thank L. Volkman (University of California) for the gift of the B12D5 monoclonal antibody and K. Harlos for the help with heavy-atom derivatives.

AUTHOR CONTRIBUTIONS

S.L. and I.M.J. expressed and purified the protein; J.K. and N.G.A.A. obtained crystals, and solved and refined the structure; J.K. and S.L. performed syncytium-formation assays; J.K., D.I.S. and I.M.J. analyzed the structure and wrote the manuscript.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Kielian, M. & Rey, F.A. Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat. Rev. Microbiol.* **4**, 67–76 (2006).
2. Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
3. Yin, H.S., Wen, X., Paterson, R.G., Lamb, R.A. & Jardetzky, T.S. Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature* **439**, 38–44 (2006).
4. Gibbons, D.L. *et al.* Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* **427**, 320–325 (2004).
5. Modis, Y., Ogata, S., Clements, D. & Harrison, S.C. Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313–319 (2004).
6. Heldwein, E.E. *et al.* Crystal structure of glycoprotein B from herpes simplex virus 1. *Science* **313**, 217–220 (2006).
7. Roche, S., Bressanelli, S., Rey, F.A. & Gaudin, Y. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* **313**, 187–191 (2006).
8. Roche, S., Rey, F.A., Gaudin, Y. & Bressanelli, S. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science* **315**, 843–848 (2007).
9. Blissard, G.W. & Rohrmann, G.F. Baculovirus diversity and molecular biology. *Annu. Rev. Entomol.* **35**, 127–155 (1990).
10. Blissard, G.W. & Wenz, J.R. Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. *J. Virol.* **66**, 6829–6835 (1992).
11. Hefferon, K.L., Oomens, A.G., Monsma, S.A., Finnerty, C.M. & Blissard, G.W. Host cell receptor binding by baculovirus gp64 and kinetics of virion entry. *Virology* **258**, 455–468 (1999).
12. Ho, Y.C. *et al.* Baculovirus transduction of human mesenchymal stem cell-derived progenitor cells: variation of transgene expression with cellular differentiation states. *Gene Ther.* **13**, 1471–1479 (2006).
13. Zeng, J., Du, J., Zhao, Y., Palanisamy, N. & Wang, S. Baculoviral vector-mediated transient and stable transgene expression in human embryonic stem cells. *Stem Cells* **25**, 1055–1061 (2007).
14. Kost, T.A., Condreay, J.P. & Jarvis, D.L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23**, 567–575 (2005).

15. Boublik, Y., Di Bonito, P. & Jones, I.M. Eukaryotic virus display: engineering the major surface glycoprotein of the *Autographa californica* nuclear polyhedrosis virus (AcNPV) for the presentation of foreign proteins on the virus surface. *Bio/Technology* **13**, 1079–1084 (1995).
16. Zhou, J. & Blissard, G.W. Mapping the conformational epitope of a neutralizing antibody (AcV1) directed against the AcMNPV gp64 protein. *Virology* **352**, 427–437 (2006).
17. Summers, M.D. & Volkman, L.E. Comparison of biophysical and morphological properties of occluded and extracellular nonoccluded baculovirus from *in vivo* and *in vitro* host systems. *J. Virol.* **17**, 962–972 (1976).
18. Zhang, Y., Kaufmann, B., Chipman, P.R., Kuhn, R.J. & Rossmann, M.G. Structure of immature West Nile virus. *J. Virol.* **81**, 6141–6145 (2007).
19. Plonsky, I. & Zimmerberg, J. The initial fusion pore induced by baculovirus gp64 is large and forms quickly. *J. Cell Biol.* **135**, 1831–1839 (1996).
20. Brown, J.H., Cohen, C. & Parry, D.A. Heptad breaks in α-helical coiled coils: stutters and stammers. *Proteins* **26**, 134–145 (1996).
21. Oomens, A.G., Monsma, S.A. & Blissard, G.W. The baculovirus gp64 envelope fusion protein: synthesis, oligomerization, and processing. *Virology* **209**, 592–603 (1995).
22. Kampmann, T., Mueller, D.S., Mark, A.E., Young, P.R. & Kobe, B. The role of histidine residues in low-pH-mediated viral membrane fusion. *Structure* **14**, 1481–1487 (2006).
23. Fredericksen, B.L. & Whitt, M.A. Vesicular stomatitis virus glycoprotein mutations that affect membrane fusion activity and abolish virus infectivity. *J. Virol.* **69**, 1435–1443 (1995).
24. Hannah, B.P., Heldwein, E.E., Bender, F.C., Cohen, G.H. & Eisenberg, R.J. Mutational evidence of internal fusion loops in herpes simplex virus glycoprotein B. *J. Virol.* **81**, 4858–4865 (2007).
25. Backovic, M., Jardetzky, T.S. & Longnecker, R. Hydrophobic residues that form putative fusion loops of Epstein-Barr virus glycoprotein B are critical for fusion activity. *J. Virol.* **81**, 9596–9600 (2007).
26. Monsma, S.A. & Blissard, G.W. Identification of a membrane fusion domain and an oligomerization domain in the baculovirus gp64 envelope fusion protein. *J. Virol.* **69**, 2583–2595 (1995).
27. Zhou, J. & Blissard, G.W. Identification of a gp64 subdomain involved in receptor binding by budded virions of the baculovirus AcMNPV. *J. Virol.* (2008).
28. Tani, H., Nishijima, M., Ushijima, H., Miyamura, T. & Matsuura, Y. Characterization of cell-surface determinants important for baculovirus infection. *Virology* **279**, 343–353 (2001).
29. Yin, H.S., Paterson, R.G., Wen, X., Lamb, R.A. & Jardetzky, T.S. Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proc. Natl. Acad. Sci. USA* **102**, 9288–9293 (2005).
30. Rossmann, M.G. & Argos, P. Exploring structural homology of proteins. *J. Mol. Biol.* **105**, 75–95 (1976).
31. Stuart, D.I., Levine, M., Muirhead, H. & Stammers, D.K. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6. *J. Mol. Biol.* **134**, 109–142 (1979).
32. Weissenhorn, W., Hinz, A. & Gaudin, Y. Virus membrane fusion. *FEBS Lett.* **581**, 2150–2155 (2007).
33. Lu, W., Chapple, S.D., Lissini, O. & Jones, I.M. Characterization of a truncated soluble form of the baculovirus (AcMNPV) major envelope protein gp64. *Protein Expr. Purif.* **24**, 196–201 (2002).
34. Walter, T.S. *et al.* Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64**, 14–18 (2008).
35. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
36. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
37. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2023–2030 (2003).
38. Abrahams, J.P. & Leslie, A.G.W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 30–42 (1996).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
40. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 458–464 (2005).
41. Cowtan, K.D. & Main, P. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 43–48 (1996).
42. Roversi, P., Blanc, E., Vonrhein, C., Evans, G. & Bricogne, G. Modelling prior distributions of atoms for macromolecular refinement and completion. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1316–1323 (2000).
43. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
44. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
45. Li, H., Robertson, A.D. & Jensen, J.H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704–721 (2005).
46. Bond, C.S. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311–312 (2003).
47. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
48. Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
49. Strelkov, S.V. & Burkhard, P. Analysis of α-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.* **137**, 54–64 (2002).

References

1. Kervestin, S. & Jacobson, A. NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* **13,** 700–712 (2012).
2. Kadlec, J., Izaurralde, E. & Cusack, S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat. Struct. Mol. Biol.* **11,** 330–337 (2004).
3. Fribourg, S., Gatfield, D., Izaurralde, E. & Conti, E. A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat. Struct. Biol.* **10,** 433–9 (2003).
4. Kadlec, J., Guilligay, D., Ravelli, R. B. & Cusack, S. Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *RNA* **12,** 1817–24 (2006).
5. Clerici, M. *et al.* Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.* **28,** 2293–2306 (2009).
6. Kadlec, J., Loureiro, S., Abrescia, N. G. a, Stuart, D. I. & Jones, I. M. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. *Nat. Struct. Mol. Biol.* **15,** 1024–1030 (2008).
7. Walter, T. S. *et al.* Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64,** 14–18 (2007).
8. El Omari, K. *et al.* Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1. *Nat. Commun.* **5,** 4874 (2014).
9. El Omari, K. *et al.* Pushing the limits of sulfur SAD phasing: De novo structure solution of the N-terminal domain of the ectodomain of HCV E1. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70,** 2197–2203 (2014).
10. Iourin, O. *et al.* Expression, purification and crystallization of the ectodomain of the envelope glycoprotein E2 from Bovine viral diarrhoea virus. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69,** 35–38 (2013).

V. Current Research

EMBL Grenoble (2008-2015) and IBS Grenoble (since 2015)

Structural and functional analysis of epigenetic regulators

In 2008, I returned back to EMBL in Grenoble where, after a short postdoc I was appointed as a staff scientist in the group of Stephen Cusack. Despite my postdoctoral experience in Oxford on structural analysis of viral proteins being very positive I realized that my major research interest is in studying eukaryotic gene expression and its links to human disease. Following my return to Grenoble I was given an opportunity to established a new, independent research line in the lab and I decided to work in the field of epigenetics and chromatin regulation focusing on structural analysis of epigenetic complexes. Then, in 2014 I obtained the ATIP/AVENIR start-up grant from the CNRS to be able set up an independent research team at the Institut de Biologie Structurale in Grenoble and further continue and expand my research on epigenetic regulatory complexes. I moved to the IBS in 2015.

Introduction

In eukaryotic cells, DNA is packaged in the nucleus in the form of chromatin. DNA wraps around octamers of core histone proteins to form nucleosomes that can further create higher-order structures as compact as metaphase chromosomes. Essentially all DNA-based processes including transcription, replication or DNA repair are controlled by a complex interplay between protein assemblies that manipulate DNA and local chromatin organization. Various post-translational modifications of histones or DNA help modify chromatin structure and are important for accessibility and recruitment of effector proteins such as transcription factors or chromatin remodellers (Kouzarides, 2007; Zentner and Henikoff, 2013). The best-characterized histone modifications include histone lysine acetylation and methylation. Histone acetylation has a potential of weakening histone-DNA interactions as well as nucleosome-nucleosome stacking, resulting in chromatin decompaction and DNA accessibility. This modification is also specifically recognized by bromodomain containing proteins. Histone methylation, depending on its position, can either repress or activate transcription and is recognized by proteins containing specialized domains such as PHD fingers, chromodomains or tudor domains.

The chromatin based (epigenetic) regulation of gene expression and other cellular processes with its implications in development or disease has been a major focus of molecular biology research in the last decade. General regulatory schemes are being uncovered mostly by genome-wide approaches. Chromatin modifications, their “writers”, “readers” and “erasers”, as well as individual pathways are being intensively studied (Figure 3). While it is now well established that most epigenetic regulators exert their function within multi-subunit assemblies, their detailed biochemical and structural characterization remains scarce. How different subunits are involved in modulating their activity and specificity as well as the mechanisms underlying epigenetic regulation often remain unclear. This information is essential if we want to understand the true nature of complex human diseases associated with epigenetic regulators.

We use an interdisciplinary approach to study essential chromatin modifying enzymes - the MOF acetyltransferase containing complexes and histone methyltransferases of the PRDM family. We combine biochemical and structural analyses with genetics and cell biology to unravel the molecular details of their epigenetic regulatory functions, their role in human disease and to fundamentally understand how chromatin modifying machines exert their control on cellular processes such as transcription or chromosomal recombination. We collaborate with leading cell biologists and geneticists in respective fields such as Asifa Akhtar (MPI Freiburg) and Bernard de Massy (IGH, Montpellier).

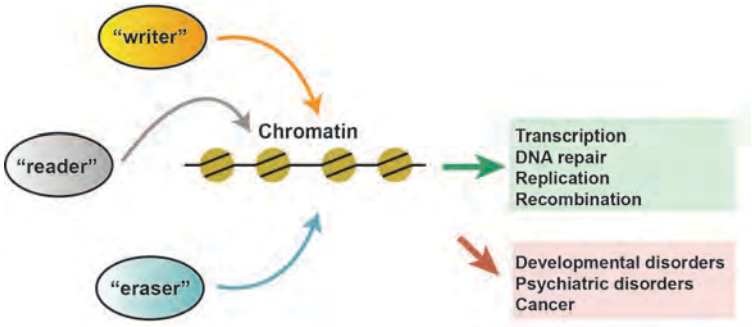


Figure 3. Chromatin regulatory complexes, that mediate post-translational modifications of chromatin (writers), that recognize (readers) or remove (erasers) those modifications, regulate most DNA associated processes. Their misregulation then often results in disease.

PRDM histone methyltransferases

As mentioned above, lysine methylation of histones is one the best characterised epigenetic chromatin modifications. The transfer of a methyl group from the cofactor S-adenosyl-L-methionine (SAM) to the target lysine is catalysed mostly by SET (**S**u(var), **E**(z) and **T**rithorax) domain containing proteins, a family of about fifty proteins in human (Schapira, 2011). While the classical SET domain enzymes are well characterized, little is known about the activity, structure or biochemistry of the PRDM sub-family of lysine methyltransferases, despite the ever-increasing list of their important functions. In human, the PRDM family consists of seventeen proteins that play important roles in a wide range of development processes controlling stem cell identity maintenance or cell differentiation and their deregulation can result in various cancers (Hohenauer and Moore, 2012). For example, PRDM1 is a master regulator of lymphoid cell differentiation (Hohenauer and Moore, 2012). PRDM14 together with PRDM1 and AP2γ are sufficient for primordial germ cell fate induction (Nakaki et al., 2013) and PRDM3 and PRDM16 are required for mammalian heterochromatin integrity (Pinheiro et al., 2012). PRDM proteins are characterized by the presence of an N-terminal PR (PRDI-BF1 and RIZ1 homology) domain followed by multiple zinc fingers, which confer DNA binding activity. An intrinsic methyltransferase activity has been shown for PRDM2, -3, -7, -8, -9 and -16 (Hohenauer and Moore, 2012). PR domains are only distantly related to the classical SET methyltransferase domains. The key signature motifs of SET domains are poorly conserved in the PRDM family and it remains unclear how the catalytic mechanism of PRDM proteins relates to that of SET domains.

PRDM9 has been intensively studied for its role in meiotic recombination that generates reciprocal exchanges between homologous chromosomes, is essential for proper alignment and segregation of chromosomes during meiosis, and is a major source of genome diversity. Errors in the meiotic recombination process lead to sterility or aneuploidy, and are a major cause of chromosome abnormalities such as the Down Syndrome. In humans, as in many mammals, recombination events concentrate to specific segments of the genome called recombination hotspots. PRDM9 was proposed to be a determinant of the positioning of meiotic recombination hotspots within the genome (Baudat et al., 2010; Myers et al., 2010). PRDM9 is only expressed in ovaries and testis and is essential for progression through meiotic prophase (Hayashi et al., 2005). In mouse, PRDM9 consist an N-terminal KRAB domain, PR/SET domain and an array of 12 Zn fingers at its C-terminus (Figure 4a). The Zn finger array was predicted and shown to bind DNA sequences enriched in the recombination hotspots. The residues responsible for the DNA binding are subject to rapid evolution and differ even in closely related species, presumably changing the PRDM9 binding sequence and hence positions of recombination hotspots (Grey et al., 2011; Patel et al., 2016; Ponting, 2011). PRDM9 has been suggested to catalyse methylation of double-methylated histone H3 Lysine 4 (H3K4me2) to H3K4me3 (Hayashi et al., 2005), an epigenetic modification that is believed to define recombination initiation, but the mechanism by which the methylation triggers DNA double strand break formation remains unknown. PRDM9 is also involved in hybrid sterility that prevents mixing of closely related species (Mihola et al., 2009). While our understanding of the role of PRDM9 at the genetic level has been

rapidly increasing, at the start of this project no biochemical or structural information on this protein was available. Similarly, most of the other PRDM proteins have never been characterized *in vitro* and the mechanistic details of their regulatory roles are often not known. The aim of this project is to shed light of the mechanistic detail underlying the various roles of the PRDM histone methyltransferase proteins. The project is being carried in collaboration with the group of Bernard De Massy (IGH Montpellier).

PRDM proteins - previous work

First, we focused on the histone methyltransferase (PR/SET) domain of mPRDM9 (Figure 4A). We solved its crystal structure in complex with the H3K4me2 substrate peptide and AdoHcy (Figure 4B)(Wu et al., 2013). This structure represents the first structure of a substrate and cofactor bound PR/SET domain of PRDM methyltransferase family and provides new insights into catalytic mechanism, substrate specificity and structural features of PRDM proteins. We could show that the genuine substrate of PRDM9 is H3K4 (rather than only H3K4me2) and that it possesses mono-, di- and trimethylation activities (rather than only trimethylation activity as reported previously (Hayashi et al., 2005). In collaboration with the group of Hong Wu (SGC, Toronto) we also determined the structure of free hPRDM9 in its autoinhibited form, where the catalytic site is blocked by the mobile post-SET domain, providing the first example of a complete rearrangement of the substrate and cofactor binding sites of a histone lysine methyltransferase (Figure 4C). We identified a conserved tyrosine in the catalytic site of PRDM9, whose mutation impairs the methyltransferase activity without any impact on the overall structure. Our collaborators then prepared a transgenic mouse carrying the mutated PRDM9 gene. They could show that while the mutated PRDM9 binds to the hotspot sequences as WT protein, the H3K4 methylation in those regions is lost and DNA recombination at these sites is impaired. This is the first direct evidence linking H3K4 methylation is responsible for DNA recombination during meiosis (unpublished).

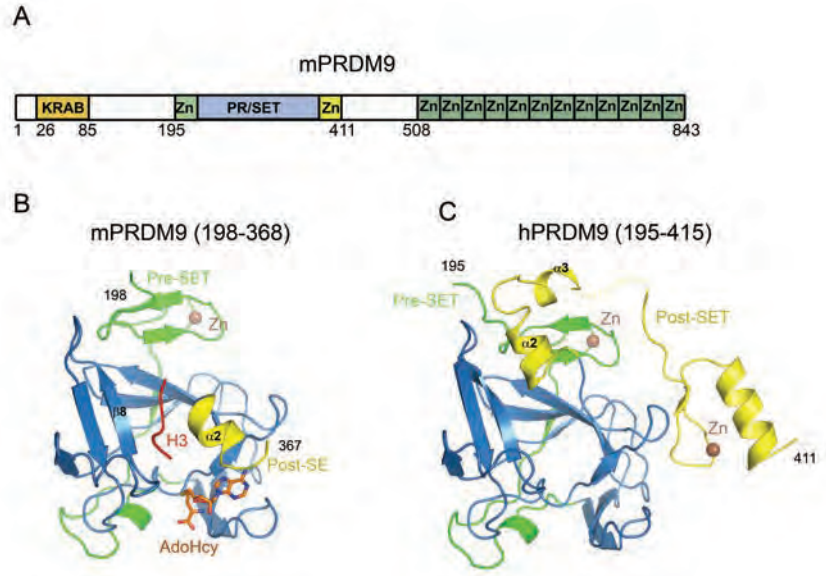


Figure 4. Crystal structures of the catalytic domain of PRDM9. (A) Schematic representation of the domain structure of mouse PRDM9. (B) Ribbon diagram of the mPRDM9 PR/SET domain in complex with H3K4me2 peptide and AdoHcy. The SET domain (residues 244-358) is shown in blue, pre-SET domain in green and post-SET in yellow. The H3K4me2 peptide (in red) binds between strand β8 and helix α2. (C) Structure of the catalytic domain of human PRDM9 in its unbound form colored as in B. Helix α2 is oriented in the opposite direction blocking the substrate binding cleft. The structure contains an additional Zn finger within its post-SET domain.

PRDM proteins - future plans

In addition to our structural analysis of the PRDM9 PR/SET domain, another group recently reported the structure of its Zn-finger array bound to DNA (Patel et al., 2016). Thus, apart from its

KRAB domain, the structure PRDM9 is now well known. We will now, in collaboration with de Massy group identify binding partners of PRDM9 and determine structures of the eventual complexes. This will be essential for our understanding of the mechanism by which the specific H3K4 methylation triggers DNA double strand break formation during meiosis in mammals. Eventual hypotheses based upon our structural work will be tested *in-vivo* in-house or in de Massy group.

Our structure of the PRDM9 PR/SET domain in complex with the H3K4me2 substrate peptide and AdoHcy is the first structure of a substrate and cofactor bound PR/SET domain of a PRDM protein. The PRDM9 structure can help us predict whether other PRDM proteins possess intrinsic methyltransferase activity or not. We will now attempt to structurally and functionally analyse the PR/SET domains, substrate specificity and the Zn finger mediated DNA binding of PRDM proteins such as PRDM1, and -16 which will provide mechanistic insights into their diverse roles, including stem cell maintenance and differentiation regulation.

Cell Reports
Report

Cell PRESS
Open ACCESS

Molecular Basis for the Regulation of the H3K4 Methyltransferase Activity of PRDM9

Hong Wu,[1] Nikolas Mathioudakis,[2,3] Boubou Diagouraga,[4] Aiping Dong,[1] Ludmila Dombrovski,[1] Frédéric Baudat,[4] Stephen Cusack,[2,3] Bernard de Massy,[4,*] and Jan Kadlec[2,3,*]
[1]Structural Genomics Consortium, University of Toronto, MaRS Center, South Tower, 7th Floor, 101 College Street, Toronto, ON M5G 1L7, Canada
[2]European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, 38042 Grenoble, France
[3]Unit for Virus Host-Cell Interactions, University Grenoble Alpes-EMBL-CNRS, 6 rue Jules Horowitz, 38042 Grenoble, France
[4]Institute of Human Genetics, UPR 1142, CNRS, 141 rue de la Cardonille, 34396 Montpellier, France
*Correspondence: bernard.de-massy@igh.cnrs.fr (B.d.M.), kadlec@embl.fr (J.K.)
http://dx.doi.org/10.1016/j.celrep.2013.08.035

SUMMARY

PRDM9, a histone lysine methyltransferase, is a key determinant of the localization of meiotic recombination hot spots in humans and mice and the only vertebrate protein known to be involved in hybrid sterility. Here, we report the crystal structure of the PRDM9 methyltransferase domain in complex with a histone H3 peptide dimethylated on lysine 4 (H3K4me2) and S-adenosylhomocysteine (AdoHcy), which provides insights into the methyltransferase activity of PRDM proteins. We show that the genuine substrate of PRDM9 is histone H3 lysine 4 (H3K4) and that the enzyme possesses mono-, di-, and trimethylation activities. We also determined the crystal structure of PRDM9 in its autoinhibited state, which revealed a rearrangement of the substrate and cofactor binding sites by a concerted action of the pre-SET and post-SET domains, providing important insights into the regulatory mechanisms of histone lysine methyltransferase activity.

INTRODUCTION

In humans, the PRDM family consists of 17 proteins that play important roles in a wide range of development processes, including stem cell identity maintenance or cell differentiation and their deregulation results in various cancers (Hohenauer and Moore, 2012). PRDM proteins are characterized by the presence of an N-terminal PR (PRDI-BF1 and RIZ1 homology) domain followed by multiple zinc fingers which confer DNA binding activity. PRDM2, -3, -8, -9, and -16 were shown to possess histone methyltransferase activity, whereas some other members were suggested to recruit histone-modifying enzymes (Eom et al., 2009; Hayashi et al., 2005; Hohenauer and Moore, 2012; Pinheiro et al., 2012). PR domains are only distantly related to the classical SET methyltransferase domains. The key signature motifs of SET domains are poorly conserved in the PRDM family, and the available crystal structures systematically lack bound substrate and cofactor. Thus, it remains unclear how the catalytic mechanism of PRDM proteins relates to that of SET domains (Hohenauer and Moore, 2012).

PRDM9 is the only member of the family (apart from its paralog PRDM7 in human) to contain a domain related to *Kr*üppel-*a*ssociated *b*ox (KRAB) often present in transcription repressors (Birtle and Ponting, 2006) but has been shown to catalyze methylation of H3K4me2 to H3K4me3 (Hayashi et al., 2005), a modification typically enriched at transcription start sites (Barski et al., 2007). PRDM9 may actually not be involved in transcription regulation but is a key factor in specifying the sites of meiotic recombination in mouse and human (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010). This function is mediated by the DNA binding specificity of its zinc finger array, and it has been proposed that PRDM9 binds to specific sites in the genome of oocytes and spermatocytes, where its methyltransferase activity leads to a local enrichment of H3K4me3 and recruits the meiotic recombination machinery (Grey et al., 2011). Whereas the importance of H3K4me3 catalyzed by Set1 in the activity of meiotic recombination sites was recently demonstrated in *S. cerevisiae* (Acquaviva et al., 2013; Sommermeyer et al., 2013), in mammals, the role of H3K4me3 in this process remains unclear. Furthermore, PRDM9 is involved in hybrid sterility, a function potentially linked to its role in recombination (Mihola et al., 2009). Here, we present crystal structures of its catalytic domain in an autoinhibited state and in complex with the H3K4me2 substrate peptide, providing insights into the substrate specificity, catalytic mechanism, and regulation of PRDM proteins.

RESULTS AND DISCUSSION

Structure of the PRDM9 PR/SET Domain in Complex with the H3K4me2 and AdoHcy

The structure of the PR/SET domain of mouse PRDM9 (mPRDM9) methyltransferase (residues 198–368) in complex with histone H3K4me2 peptide and AdoHcy (a methylation

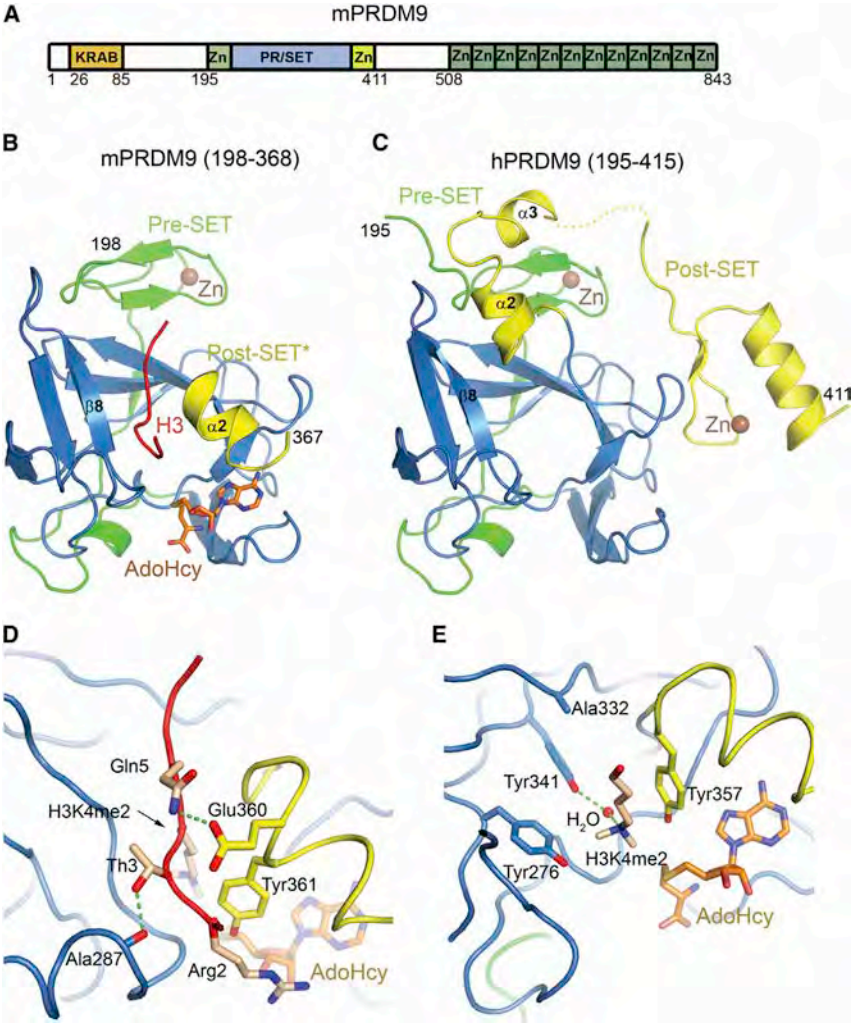


Figure 1. Crystal Structures of the Catalytic Domain of PRDM9

(A) Schematic representation of the domain structure of mouse PRDM9.

(B) Ribbon diagram of the mPRDM9 PR/SET domain in complex with H3K4me2 peptide and AdoHcy. The SET domain (residues 245–358) is shown in blue, pre-SET domain in green, and truncated post-SET (post-SET*) in yellow. The H3K4me2 peptide (in red) binds between strand β8 and helix α2.

(C) Structure of the catalytic domain of hPRDM9 in its unbound form colored as in (B). Helix α2 is oriented in the opposite direction blocking the substrate binding cleft. The structure contains an additional Zn finger within its post-SET domain.

(D) Details of H3K4me2 peptide recognition by mPRDM9. Only hydrogen bonds formed by peptide side chains are shown.

(E) The environment of H3K4me2 ε-amino group in the mPRDM9 catalytic site. Ala332 corresponds to F/Y switch residue Y305 in SETD7 (Del Rizzo et al., 2010).

See also Figure S1.

reaction product) was determined by X-ray crystallography (Figure 1). The asymmetric unit contains two mPRDM9 molecules: one in a ternary complex with the H3 peptide and AdoHcy (Figure 1B) and one in its unbound form (Figures S1A and S1B). Overall, the mPRDM9 PR/SET domain topology corresponds to other SET domains, with the conserved central SET domain fold (Figures 2A, 2B, and S2) flanked by pre-SET and post-SET regions. The pre-SET domain spans residues 198–244 and forms a Zn finger that is connected to the SET domain with a long linker (residues 222–244). The construct used contains only a short post-SET region fragment (residues 359–368; referred to as post-SET*) that is only folded in the presence of AdoHcy and the peptide (Figure S1B).

Although the two SET domain signature motifs (NHS/CxxPN and ELxF/YDY; x being any amino acid; Qian and Zhou, 2006) are not well conserved in the PRDM family, the AdoHcy is bound by mPRDM9 in a conventional way (Figures 2C and 2D), indicating that only Asn320 of the NHS/CxxPN motif (320-NCARDDEEQN in mPRDM9) is necessary for cofactor binding. Surprisingly, in PRDM3, -8, and -16 that were shown to possess methyltransferase activity (Eom et al., 2009; Hohenauer and Moore, 2012; Pinheiro et al., 2012), this invariant asparagine is replaced by an arginine or glutamine residue (Figure 2C). These proteins, thus, do not seem to have the potential of binding AdoHcy in the way that is highly conserved among SET-domain-containing methyltransferases and PRDM9. It is not clear whether arginine or glutamine residues can functionally substitute for Asn320. In PRDM4, for which the methyltransferase activity has not been reported, we could show that the corresponding arginine residue can block the AdoHcy binding site (Figure 2E).

The H3K4me2 peptide binds to PRDM9 in a cleft between strand β8 and helix α2 similar to other SET domains (Figures 1D and 1E). Electron density was interpretable for residues 1–7. In addition to backbone interactions with strand β8 and helix α2, the peptide also makes several side chain contacts with PRDM9 that might determine its specific recognition (Figure 1D). These include a hydrogen bond between Gln5 and the post-SET Glu360 and Thr3 interaction with backbone carbonyl of Ala287. Arg2 stacks against Tyr361 and makes a hydrogen bond with AdoHcy.

Substrate Specificity and the Active Site of PRDM9

PRDM9 was proposed to methylate H3K4me2, but not H3K4 nor H3K4me1 (Hayashi et al., 2005). We analyzed the ability of mPRDM9 (198–368) to bind unmodified H3K4 and H3K4me2 peptides by isothermal titration calorimetry and could show that mPRDM9 binds the two peptides with dissociation constants of 102 μM and 43 μM, respectively (Figures S3A and

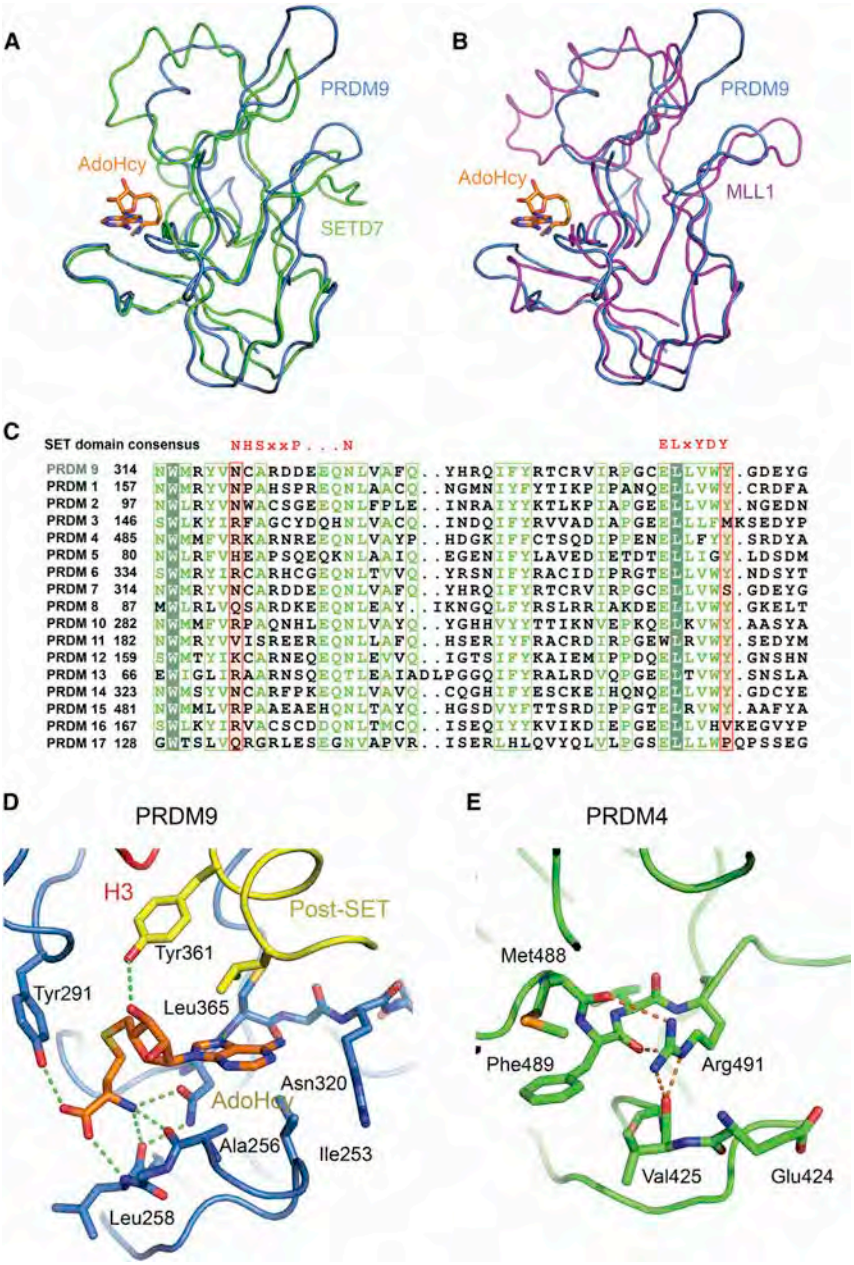


Figure 2. AdoHcy Binding by PRDM Proteins

(A) Comparison of the core SET domain of PRDM9 (residues 245–358) with residues 215–336 of SETD7 (PDB code 1O9S; root-mean-square deviation of 1.83 Å for 98 Cα atoms).

(B) Superposition of the core SET domain of PRDM9 onto MLL1 residues 3830–3945 (PDB code 2W5Z; root-mean-square deviation of 1.6 Å for 96 Cα atoms).

(C) Sequence alignment of PRDM proteins covering the two SET domain consensus regions (in red). Identical residues are in green boxes. Residues corresponding to PRDM9 Asn320 (AdoHcy binding) and Tyr357 (active site), invariant in other SET domains, are highlighted by a red frame.

(D) Details of AdoHcy binding by PRDM9. The AdoHcy interacts with Asn320, main chain atoms of the AGLG motif (Gly257, Gly259), Tyr 291 of the I-SET, and Tyr361 and Leu365 of the post-SET helix α2.

(E) Arg491 in PRDM4 (PDB code 3DB5) corresponding to PRDM9 Asn320 blocks the AdoHcy binding site.

See also Figure S2.

S3B). Additionally, MALDI-TOF mass spectrometry analysis revealed that mPRDM9 can mono-, di-, and trimethylate unmodified H3K4 peptide (Figures 3A and 3B). Finally, in a coupled fluorescent methylation assay, both H3K4 and H3K4me2 peptides were efficiently methylated by mPRDM9 (198–368), whereas H4K20 peptide was not (Figure 3C). The activity on H3K4 was lower but still significant, probably reflecting the lower binding affinity (Figure 3C). Together, these results clearly show that unmodified H3K4 is a genuine substrate of mPRDM9 and that mPRDM9 possesses mono-, di-, and trimethylation activities.

To explore more widely the substrate specificity of PRDM9, we tested the activity of mPRDM9 (198–368) on an array of 379 distinct peptides corresponding to different regions of H2A, H2B, H3, and H4 with various modifications (Figure S3C; Table S1). Significant methyltransferase activity was detected on H3 (1–19) peptides, confirming H3K4 (me0, -1, and -2) to be efficient substrates of PRDM9 (Figure 3D). Arg2 methylation had no significant effect on the H3K4 methylation within H3 (1–19) peptides carrying K9me3 modification (Figure 3D). Surprisingly, PRDM9 can also methylate H3K9 (me0, -1, and, -2) substrates (Figure 3D) and potentially H3K36 based on the signal observed on the H3 (26–45) peptide (Figure S3C). Although no enrichment for H3K9me3 could be detected at a mouse meiotic recombination site bound by PRDM9 (Buard et al., 2009), additional assays for substrate specificities in vitro and in vivo could certainly be interesting. Other SET domain methyltransferases, such as Ash1 from *D. melanogaster*, do display activities on various substrates which may be regulated in vivo (Beisel et al., 2002).

H3K4me2 is located in the channel formed by Trp293, Tyr357, and Tyr361. In the active site, the H3K4me2 ε-amino group is surrounded by three conserved tyrosine residues: Tyr357, Tyr276, and Tyr341 (Figure 1E). Invariant tyrosines corresponding to Tyr357 have been proposed to be directly involved in catalysis (Smith and Denu, 2009). We show that the Y357F mutation completely abolished the in vitro activity of mPRDM9 (Figures 3E and S3C). Similarly, Y276F and Y341F mutants

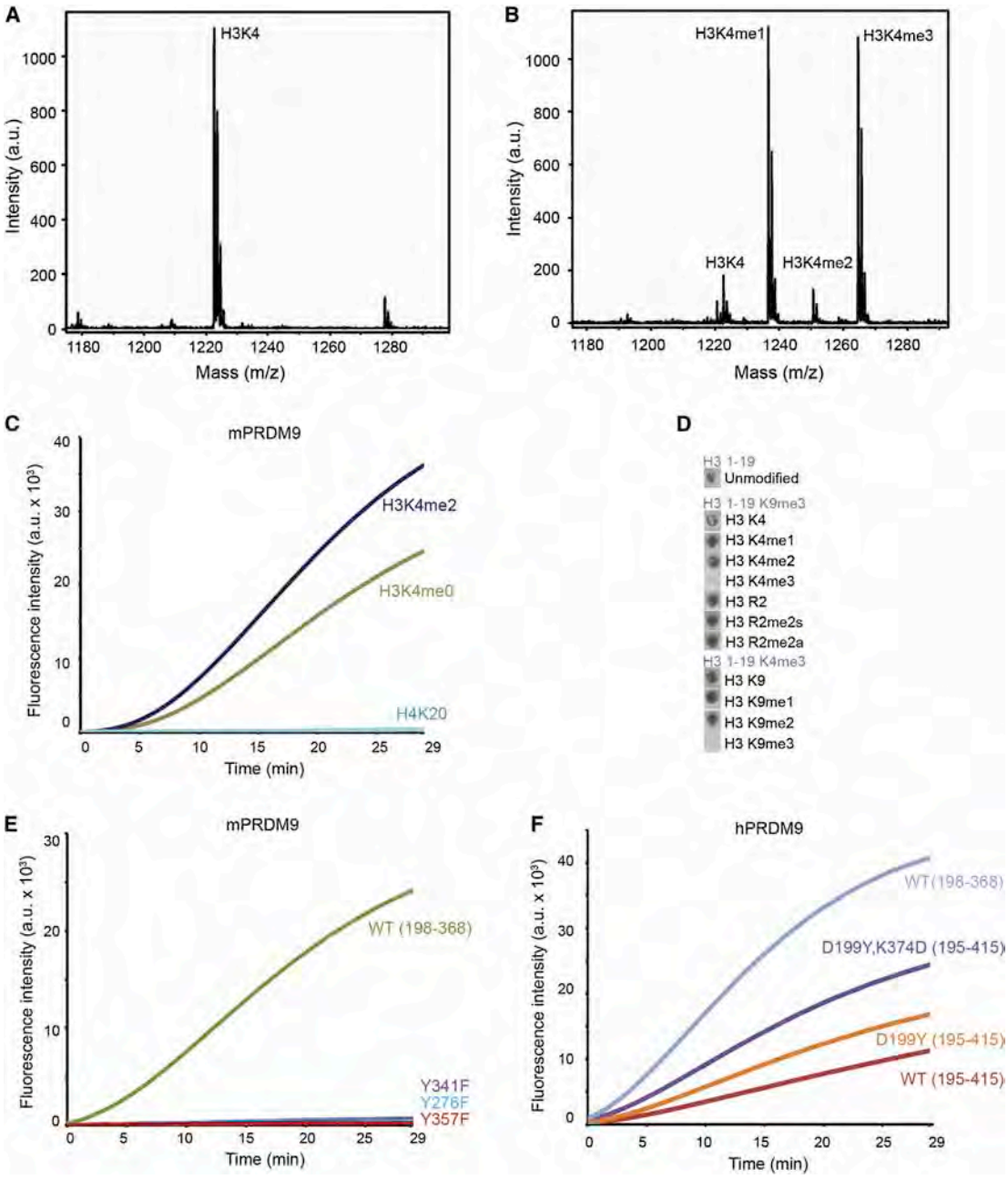


Figure 3. Substrate Specificity of mPRDM9

(A and B) MALDI-TOF mass spectrometry analysis of the H3K4 peptide methylation by PRDM9 (198–368) is shown. Whereas H3K4 peptide incubated in the absence of mPRDM9 is not methylated (A), after incubation with mPRDM9 (198–368), single, double, and triple methylation could be detected (B). a.u., arbitrary units.

(C) Coupled fluorescent methylation assay. H3K4 and H3K4me2 peptides are efficiently methylated by mPRDM9. No methylation is observed for H4K20 peptide.

(D) mPRDM9 methyltransferase activity on histone H3 peptides. Top panel shows methylation of H3 (1–19) peptide. Middle panel shows methylation of H3 (1–19) R8me2s/K9me3 peptides with various levels of K4 methylation and of K4me1/R8me2s/K9me3 peptide with indicated R2 modifications. The bottom panel shows R2me2s/R8me2s/K4me3 peptides with various levels of K9 methylation. No methylation is detected on H3 (1–19) K4me3/K9me3 peptides.

(E) Mutagenesis of key catalytic residues. Y357F, Y276F, and Y341F mutations abolish the methyltransferase activity of mPRDM9 (198–368).

(F) hPRDM9 (195–415) can methylate the H3K4 peptide. The methylation activity is ~4.4 times lower than for hPRDM9 (198–368). Mutations of Asp199 of the pre-SET domain and Lys374 of the post-SET domain, disrupting the pre-SET/post-SET-inhibitory interaction result in a higher methyltransferase activity (1.6 times higher for D199Y and 2.4 times higher for D199Y, K374D). The methyltransferase activity was calculated using linear portion of the curves between minute 7 and 15. Gel filtration and thermal shift measurement of melting temperature confirmed the structural integrity of these mutants (data not shown).

See also Figure S3.

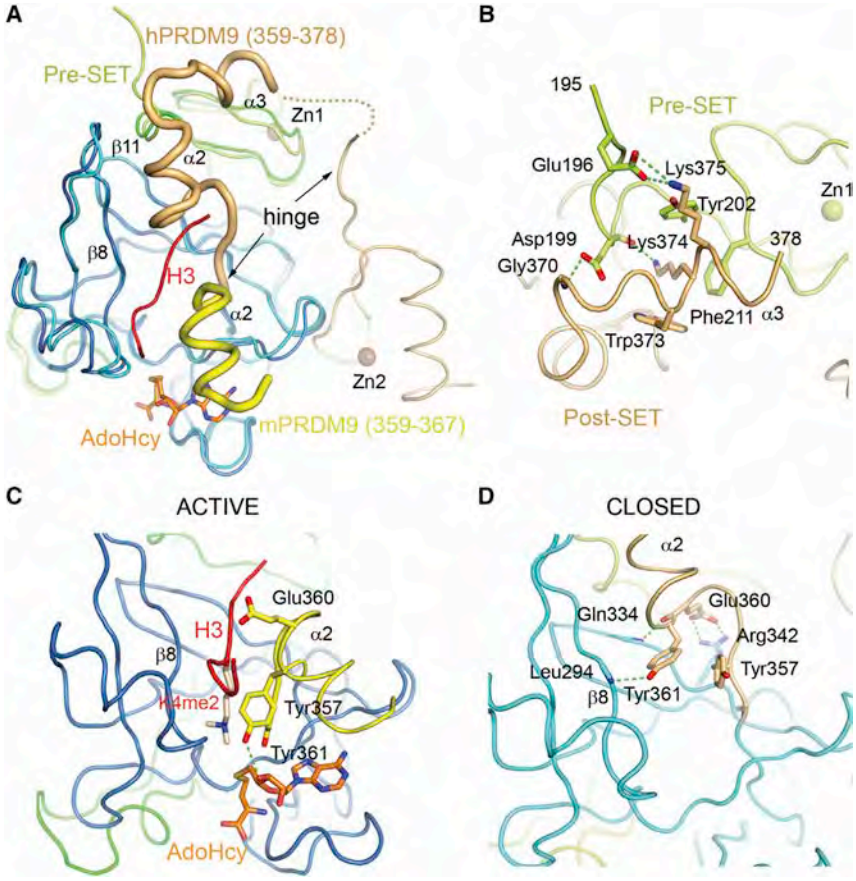


Figure 4. Conformational Changes of the PRDM9 Active Site

(A) Superposition of mPRDM9 (198–368) and hPRDM9 (195–415) structures. The post-SET domains are highlighted in yellow and brown, respectively. Proposed hinge regions are shown.

(B) Details of the interaction between the pre-SET and post-SET domain of hPRDM9. Gly370 forms a hydrogen bond with Asp199. The side chain of Lys374, which is surrounded by Tyr202, Phe211, and Trp373, binds to backbone carbonyl of Asp199. Lys375 makes a salt bridge with Glu196.

(C) Glu360 and Tyr361 participate in the substrate and cofactor binding in mPRDM9. Coloring corresponds to Figure 1B.

(D) Glu360 stabilizes the autoinhibitory conformation of hPRDM9 by a salt bridge with Arg342, and Tyr361 blocks the substrate binding cleft. The catalytic Tyr357 is also displaced.

See also Figure S4.

were inactive in methylation assay using the H3K4 peptide (Figure 3E). Surprisingly, Tyr276 was previously reported to be dispensable for methyltransferase activity on H3 (Hayashi et al., 2005). The integrity of the mutated proteins was verified by gel filtration and thermal shift measurement of melting temperature (Y276F decreased thermal stability of PRDM9 by 5°C; data not shown). Interestingly, Tyr357 is not conserved in PRDM3 and PRDM16 (Figure 2C). Thus, these PRDM proteins are either not truly active methyltransferases or they use different catalytic residues than those conserved in PRDM9 and all other SET domains. Tyr341 helps coordinate a water molecule in the active site (Figure 1E). In our structure, one of the methyl groups of the H3K4me2 would clash with the modeled position of the S-adenosyl methionine (AdoMet) donor methyl. Thus, we interpret our structure as being in a product conformation. Before the final methylation reaction, the lysine would need to be rotated within the catalytic site, possibly to position the clashing methyl group into that occupied by the active site water, as suggested for SETD7 (Del Rizzo et al., 2010). In zebrafish, both Tyr276 and Tyr341 are substituted with phenylalanine (Figure S4A), possibly indicating altered catalytic activity. Gly278 was suggested to be a key residue for the activity of mPRDM9 (Hayashi et al., 2005). The structure reveals that Gly278 is located just upstream of strand β7 forming a β sheet with β8 and β9 and is not in proximity to the substrate or cofactor binding site (11 Å from the H3K4me2 ε-amino group). Thus, this residue is unlikely to have a direct effect on catalysis but appears to be important for structural integrity of the domain because a G278A mutant cannot be expressed in soluble form in bacteria (data not shown).

Structure of the Autoinhibited PRDM9 PR/SET Domain

The post-SET domain is disordered in the unbound mPRDM9 molecule in the crystal (Figure S1B). Similarly, the post-SET domains of DIM-5 or MLL1 become fully folded only in the presence of the AdoHcy/AdoMet cofactor and substrate (Southall et al., 2009; Zhang et al., 2002). To better understand the role of the post-SET domain, we produced crystals of a larger fragment of mouse PRDM9 that, however, only diffracted to a low resolution. We were, however, able to solve the structure of human PRDM9 fragment spanning residues 195–415 (hPRDM9; 90% sequence identity with mPRDM9; Figure 1C). Compared to mPRDM9 (198–368), this fragment includes a slightly longer pre-SET region and an extended post-SET domain including a Zn finger, which the structure shows loosely associates with the SET domain via Glu326 and Arg345. Despite the presence of the substrate peptide and AdoHcy in the crystallization solution, hPRDM9 (195–415) crystallized in an unbound form. Unexpectedly, the hPRDM9 structure reveals that, in the absence of the substrate and cofactor, the post-SET region is ordered but undergoes a conformational change and binds across the SET domain, blocking the peptide binding cleft (Figure 4A). Consequently, also the AdoHcy binding site becomes incomplete suggesting that this is an autoinhibitory conformation. This position of the post-SET domain is stabilized by several interactions with the pre-SET domain Zn finger (Figures 4A and 4B). The last residue in the same position between the two structures is Val355. The following residues that form helix α2 (residues 359–364) in the ligand bound structure are rotated by ~180° around Val355

(Figure 4A). The new helix α2 is then formed by residues 362–367 packing against strands β8 and β11. Glu360 and Trp356 make backbone hydrogen bonds with Leu332 and Gln334 of β11. Glu360, which is involved in the substrate recognition in the complex structure, makes a salt bridge interaction with Arg342, whereas Tyr361 that forms the substrate lysine binding channel in the complex blocks the peptide binding cleft forming a hydrogen bond with Leu294 (Figures 4C and 4D). The following helix α3 packs against the pre-SET domain as shown in Figure 4B. We tested whether hPRDM9 (195–415) maintains methyltransferase activity and find that it is reduced from that of hPRDM9 (198–368) but still significant (Figure 3F). Thus, it seems that the presence of AdoMet and the peptide substrate is sufficient to overcome the inhibition and trigger hPRDM9 methylation activity. Additionally, we could show that mutations predicted to disrupt the inhibitory interaction between the pre-SET and post-SET domains (D199Y and K374D) increased the activity of hPRDM9 (195–415; Figure 3F). We propose that PRDM9 residues 355–356 and the linker between helix α3 and the post-SET Zn finger function as hinge regions that enable translocation of the helix α2 between the active conformation where Glu360 and Tyr361 participate in substrate lysine and cofactor binding and an inhibitory conformation stabilized by Glu360, where Tyr361 blocks the substrate binding site (Figure 4A). It remains to be established whether the transition between the stable rather than disordered conformation of the post-SET domain and its active state is regulated by other factors or posttranslational modifications during meiosis. This is a unique example of a complete rearrangement of the substrate and cofactor binding sites of a histone lysine methyltransferase by a concerted action of the pre-SET and post-SET domains.

EXPERIMENTAL PROCEDURES

Expression, Purification, and Crystallization

A His-tag fusion of mouse PRDM9 (residues 198–368) was expressed in *E. coli* BL21Star (DE3) from pProEXHTb vector. The protein was first purified by affinity chromatography using Ni^{2+} resin. After His-tag cleavage with tobacco etch virus (TEV) protease, the protein was further purified by a second Ni^{2+} column and size-exclusion chromatography. Pure PRDM9 was concentrated with 2 mM AdoHCy and 2 mM H3K4me2 peptide (ARTKme2QTARK-Y; Y added to facilitate quantification) to 9 mg/ml in a buffer containing 20 mM Tris, pH 7.0, 150 mM NaCl, and 5 mM β-mercaptoethanol. The best-diffracting crystals grew within 3 days at 5°C in a solution containing 0.2 M ammonium sulfate; 0.1 M Bis-Tris, pH 5.5; and 25% w/v polyethylene glycol (PEG) 3350. For data collection at 100 K, crystals were snap frozen in liquid nitrogen with a solution containing mother liquor and 30% (v/v) glycerol.

hPRDM9 (residues 195–415) was expressed as a His-tag fusion in *E. coli* BL21 (DE3) V2R-pRARE from pET28-MHL vector. Harvested cells were resuspended in phosphate-buffered saline, supplemented with 250 mM NaCl, 5 mM imidazole, 3 mM β-mercaptoethanol, 5% glycerol, 0.1% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, and 1 mM phenylmethanesulfonylfluoride. After affinity purification on 5 ml HiTrap Chelating column (GE Healthcare), the protein was further purified by size-exclusion chromatography. After TEV protease cleavage, hPRDM9 was purified to homogeneity by ion-exchange chromatography on Source 30Q column (10 × 10) (GE Healthcare). Pure hPRDM9 protein (8 mg/ml) was crystallized in a solution containing 23% PEG 3350; 0.2 M ammonium acetate; and 0.1 M BisTris, pH 5.5. The crystal was frozen in liquid nitrogen using 15% ethylene glycol as cryoprotectant.

hPRDM4 (residues 390–540) was expressed as a His-tag fusion in *E. coli* BL21 (DE3) Codon Plus RIL (Stratagene) from pET28a-LIC vector. The overexpressed protein was purified as described above for hPRDM9. Purified PRDM4 (10 mg/ml) was crystallized using hanging drop vapor diffusion method at 20°C by mixing 1.5 μl of the protein solution with 1.5 μl of the reservoir solution containing 23% PEG 3350; 0.2 M ammonium acetate; and 0.1 M BisTris, pH 6.5. The crystal was frozen in liquid nitrogen using Paratone-N as cryoprotectant.

Data Collection and Structure Determination

Crystals of mPRDM9 (198–368) belong to the space group $P2_12_12_1$ with unit cell dimensions a = 55.7 Å, b = 78.2 Å, and c = 107.6 Å. The asymmetric unit contains two PRDM9 molecules and has a solvent content of 60%. A complete native data set was collected to a resolution of 2.3 Å on beamline ID29 at the European Synchrotron Radiation Facility (ESRF). The data were processed using XDS (Kabsch, 2010). The structure was solved by molecular replacement with PHASER (McCoy et al., 2005) using the structure of PRDM11 (Protein Data Bank [PDB] code 3RAY; 44% sequence identity) as a search model. Using prime-and-switch density modification of program RESOLVE (Terwilliger, 2000) enabled to us to obtain a clearly interpretable map. The structure was built in COOT (Emsley and Cowtan, 2004) and refined in REFMAC5 (Murshudov et al., 1997) to final R factor of 20.8% and R_{free} of 24.5% (Table S2) with all residues in allowed (97% in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY (Davis et al., 2004).

X-ray diffraction data for hPRDM9 was collected at 100 K at beamline 19-ID of Advanced Photon Source at Argonne National Laboratory. Crystals of hPRDM9 (195–415) belong to the space group $P2_1$ with unit cell dimensions of a = 54.7 Å, b = 48.8 Å, c = 78.7 Å, and β = 100°. Data were processed using HKL-3000 suite (Otwinowski and Minor, 1997). The structure of hPRDM9 was determined by single-wavelength anomalous dispersion (SAD) phasing method using a SeMet-substituted crystal. SOLVE/RESOLVE (Terwilliger and Berendzen, 1999) were used for heavy atom site search, phase improvement, and initial model building. Program BUCCANEER (Cowtan, 2006) and ARP/wARP (Perrakis et al., 1999) were used for automatic model building. COOT (Emsley and Cowtan, 2004) was used for model building and visualization. Structure was refined with REFMAC5 (Murshudov et al., 1997) to final R factor of 19.6% and R_{free} of 26.4% with all residues in allowed (96% in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY (Davis et al., 2004). Crystal diffraction data and refinement statistics for the structure are displayed in Table S2. Representative parts of the $2F_o$–F_c electron density maps calculated using the refined models are shown in Figures S4B and S4C.

X-ray diffraction data for hPRDM4 were collected at 100 K on RIGAKU FR-E DW. Crystals of hPRDM4 belong to the space group *I*422 with unit cell dimensions of a = 107.3 Å, b = 107.3 Å, and c = 133.6 Å. Data were processed using HKL-2000 suite (Otwinowski and Minor, 1997). The structure of hPRDM4 was determined by SAD phasing method using a SeMet-substituted crystal. The structure was solved by using SHELXD (Sheldrick, 2008). COOT (Emsley and Cowtan, 2004) was used for model building and visualization. Structure was refined with REFMAC5 (Murshudov et al., 1997) to final R factor of 22.2% and R_{free} of 29.8% with all residues in allowed (92.1% in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY (Davis et al., 2004). Crystal diffraction data and refinement statistics for the structure are displayed in Table S2.

Methyltransferase Assays

Methyltransferase activity of PRDM9 was analyzed using a continuous, fluorescent-coupled assay SAMfluoro (G-Bioscience). The AdoHcy product of the methylation reaction is further converted by a supplied mixture of three enzymes to hydrogen peroxide that reacts with 10-acetyl-3,7-dihydroxyphenoxazine to produce a fluorescent compound Resorufin. The production of Resorufin was monitored for 30 min at room temperature using an Infinite 200 PRO plate reader (Tecan). The methylation reactions, containing 0.26 or 0.3 μM enzyme and 0.22 mM peptide substrate, were set up according to manufacturer instructions. The tested substrate includes H3K4 (ARTKQTARK-Y), H3K4me2 (ARTKme2QTARK-Y), and H4K20 (Y-GKGGKGLGKGGAKRHR KVLRD) peptides. All peptides include an extra tyrosine residue for quantification. All the reactions were performed at least in duplicates, and the differences between methylation rates calculated from the linear parts of corresponding curves (above background level) were within 10%.

For mass spectrometry analysis, 1 mM H3K4 peptide was incubated with 5 mM AdoMet in the presence or absence of 10 μM mPRDM9 (198–368) at room temperature for 2 hr in 20 mM Tris 7.0 and 150 mM NaCl. The reaction mixture was analyzed by MALDI-TOF mass spectrometry. The duplicate experiment yielded essentially identical results.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) experiments were performed at 25°C, using an ITC200 microcalorimeter (MicroCal). Experiments included 26 injections of 1.5 μl of 2.7 mM peptide solution into the sample cell containing 60 μM of mPRDM9 (198–368) in 20 mM Tris 7.0, 150 mM NaCl, and 5 mM β-mercaptoethanol. The initial data point was deleted, and the last point after saturation was subtracted from the data sets. Binding isotherms were fitted with a one-site binding model by nonlinear regression using Origin Software version 7.0 (MicroCal).

Methyltransferase Activity Assay on Peptide Array

Modified histone peptide arrays (Active Motif, catalog number 13005) were preincubated in methylation buffer (50 mM Tris-HCl, pH 9.0; 100 mM NaCl; 5 mM dithiothreitol; 10 μM $ZnCl_2$) for 20 min, followed by incubation with methylation buffer containing 0.58 μM of (methyl-3H)-S-adenosyl-L-methionine (Perkin Elmer) and 0.2 μM of either wild-type or Y357F mutant mPRDM9 (198–368) for 10, 30, or 60 min at ambient temperature. The arrays were washed four times for 5 min with washing buffer (50 mM NH_4HCO_3 and 0.1% SDS) and then dried and rinsed with Amplify NAMP100 solution (GE Healthcare). The arrays were completely dried and exposed to the Carestream Kodak Biomax MR films in darkness at −80°C for 1–3 days.

ACCESSION NUMBERS

Protein Data Bank coordinates for the crystal structure of the mPRDM9 (198–368) and hPRDM9 (195–415) have been deposited with accession codes 4C1Q and 4IJD, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.08.035.

ACKNOWLEDGMENTS

We thank the Grenoble European Molecular Biology Laboratory (EMBL)-European Synchrotron Radiation Facility (ESRF)-L'Institut Laue-Langevin (ILL)-Institut de Biologie Structurale (IBS) Partnership for Structural Biology for access to structural biology instrumentation, notably L. Signor of the IBS for his help with mass spectrometry, the High Throughput Crystallisation Laboratory group of EMBL Grenoble for performing initial screening crystallization trials, A. Palencia for help with the isothermal titration calorimetry, and J. Perard for help with the methylation assays. We thank the ESRF-EMBL Joint Structural Biology Group for access to and assistance on the ESRF synchrotron beamlines. This work was supported by the EU FP7-funded Network of Excellence EpiGeneSys awarded to S.C. and B.d.M. B.d.M. is supported by the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (09-BLAN-0269-01), and the Fondation pour la Recherche Médicale. "The SGC" is a registered charity (number 1097737) that receives funds from AbbVie, Boehringer Ingelheim, the Canada Foundation for Innovation, the Canadian Institutes for Health Research, Genome Canada through the Ontario Genomics Institute (OGI-055), GlaxoSmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda, and the Wellcome Trust (092809/Z/10/Z).

Received: May 21, 2013
Revised: July 13, 2013
Accepted: August 21, 2013
Published: October 3, 2013

REFERENCES

Acquaviva, L., Székvölgyi, L., Dichtl, B., Dichtl, B.S., de La Roche Saint André, C., Nicolas, A., and Géli, V. (2013). The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. Science *339*, 215–218.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science *327*, 836–840.

Beisel, C., Imhof, A., Greene, J., Kremmer, E., and Sauer, F. (2002). Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature *419*, 857–862.

Birtle, Z., and Ponting, C.P. (2006). Meisetz and the birth of the KRAB motif. Bioinformatics *22*, 2841–2845.

Buard, J., Barthès, P., Grey, C., and de Massy, B. (2009). Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. EMBO J. *28*, 2616–2624.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. *32*(Web Server issue), W615–W619.

Del Rizzo, P.A., Couture, J.F., Dirk, L.M., Strunk, B.S., Roiko, M.S., Brunzelle, J.S., Houtz, R.L., and Trievel, R.C. (2010). SET7/9 catalytic mutants reveal the role of active site water molecules in lysine multiple methylation. J. Biol. Chem. *285*, 31849–31858.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Eom, G.H., Kim, K., Kim, S.M., Kee, H.J., Kim, J.Y., Jin, H.M., Kim, J.R., Kim, J.H., Choe, N., Kim, K.B., et al. (2009). Histone methyltransferase PRDM8 regulates mouse testis steroidogenesis. Biochem. Biophys. Res. Commun. *388*, 131–136.

Grey, C., Barthès, P., Chauveau-Le Friec, G., Langa, F., Baudat, F., and de Massy, B. (2011). Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLoS Biol. *9*, e1001176.

Hayashi, K., Yoshida, K., and Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature *438*, 374–378.

Hohenauer, T., and Moore, A.W. (2012). The Prdm family: expanding roles in stem cells and development. Development *139*, 2267–2282.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr. D Biol. Crystallogr. *61*, 458–464.

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science *323*, 373–375.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G., and Donnelly, P. (2010). Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science *327*, 876–879.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Parvanov, E.D., Petkov, P.M., and Paigen, K. (2010). Prdm9 controls activation of mammalian recombination hotspots. Science *327*, 835.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F.M., Mittler, G., Genoud, C., Goyama, S., Kurokawa, M., et al. (2012). Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell *150*, 948–960.

Qian, C., and Zhou, M.M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cell. Mol. Life Sci. *63*, 2755–2763.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr. A *64*, 112–122.

Smith, B.C., and Denu, J.M. (2009). Chemical mechanisms of histone lysine and arginine modifications. Biochim. Biophys. Acta *1789*, 45–57.

Sommermeyer, V., Béneut, C., Chaplais, E., Serrentino, M.E., and Borde, V. (2013). Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. Mol. Cell *49*, 43–54.

Southall, S.M., Wong, P.S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Mol. Cell *33*, 181–191.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. *56*, 965–972.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Zhang, X., Tamaru, H., Khan, S.I., Horton, J.R., Keefe, L.J., Selker, E.U., and Cheng, X. (2002). Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. Cell *111*, 117–127.

Mof acetyltransferase containing complexes

MOF (Males absent On the First) belongs to the MYST family of histone acetyltransferases (HAT) including Tip60, MOZ, MORF and HBO1. These enzymes function within multi-subunit protein complexes that regulate their specificity and activity as well as playing a critical role in various DNA-linked cellular processes. Anomalous activities of these complexes have been connected to human disease, including cancer (Avvakumov and Côté, 2007).

MOF was first shown to play a crucial role in the dosage compensation mechanism in *Drosophila*, specifically acetylating lysine 16 of histone H4 (H4K16) on the male X chromosome (Conrad and Akhtar, 2011)(Keller and Akhtar, 2015). It is also responsible for a large fraction of H4K16 acetylation (H4K16ac) in mammalian cells (Taipale et al., 2005). MOF is essential for the maintenance and self-renewal of pluripotent embryonic stem cells (Li et al., 2012) and MOF-deficient mouse embryos fail to develop beyond blastocyst stage (Thomas et al., 2008). MOF and H4K16 acetylation also regulate autophagy (Füllgrabe et al., 2013) and are critical for the DNA damage response and double strand break repair (Li et al., 2010; Sharma et al., 2010). In addition, MOF acetylates acetylates several non-histone targets, including tumor suppressor p53, whose acetylation on lysine 120 induces apoptosis (Li et al., 2009). Misregulation of MOF and H4K16ac is implicated in many cancers (Su et al., 2016).

For a long time MOF has been known to be a component of the dosage compensation complex (DCC), also known as the MSL complex involved in dosage compensation in *Drosophila* (Conrad and Akhtar, 2011). MOF was then also shown to exist within the NSL (Non-Specific Lethal) complex involved in global transcription regulation (Raja et al., 2010).

Dosage compensation complex (MSL)

In *Drosophila*, as well as in mammals, females are characterized by two X chromosomes, whereas male cells contain only a single X chromosome. To achieve similar expression levels of X-linked genes in males and females, the X chromosomes are involved in a process of dosage compensation. In *Drosophila* males, the balance is achieved by an approximately two-fold up-regulation of transcription of X-chromosomal genes (Conrad and Akhtar, 2011; Keller and Akhtar, 2015). The process of dosage compensation in *Drosophila* is an established model of chromosome-wide transcription regulation at the chromatin level. A key role in this mechanism is played by the MSL complex (Figure 5), a large assembly consisting of five male-specific lethal proteins (MSL1, MSL2, MSL3, MOF, maleless (MLE) RNA helicase, and two functionally redundant long non-coding RNAs (lnc-RNA: roX1 and roX2) (Conrad and Akhtar, 2011; Keller and Akhtar, 2015). The MLE helicase is not a core MSL member but rather remodels specific stem-loop structures of roX RNAs before their integration into the complex (Ilik et al., 2013; Maenner et al., 2013). MSL complex specifically binds to and acetylates hundreds of sites (H4K16) on the male X chromosome. According to the current model, the MSL proteins and roX RNAs first assemble on numerous sequences called high affinity sites, followed by spreading of the complex onto the rest of the X chromosome (low affinity sites)(Alekseyenko et al., 2008; Straub et al., 2008; Villa et al., 2016)(Keller and Akhtar, 2015). The action of the MSL complex seems to eventually enhance transcription elongation of X-linked genes (Larschan et al., 2011).

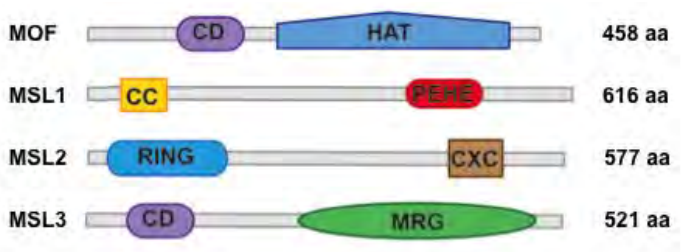


Figure 5. Domain structure of the human MSL complex subunits .CD – chromo barrel domain, CC – coiled coil .

Despite the intense research in this field, many important questions remain unresolved. We still do not fully understand how the specific recruitment of the MSL complex to the X chromosome is achieved. The mechanism of spreading of the complex along the X chromosome is not understood. It still remains to be determined how the presence of the MSL complexes on the male X chromosome ultimately results in a 2-fold activation of transcription. The molecular architecture of the MSL complex is not known. The role of the MSL complex is in mammals is not know.

MSL complex - previous work

The MOF-containing complexes are multi-protein assemblies. Individual subunits contain long, poorly conserved linker regions between the predicted domains, providing these complexes with a high level of flexibility (Figure 5). While this intrinsic flexibility is likely to be functionally relevant, as the complexes function in context of chromatin, for structural biology such large, inherently flexible complexes represent a considerable challenge, as they are often present in multiple conformations. Our strategy was to structurally analyse smaller MSL sub-complexes, and use the obtained information for rational analysis of the full complex. Indeed, so far we were able to determine crystal structures of all defined sub-complexes within the MSL complex. In the meantime, other groups published structures of the MOF and MSL3 chromobarrel domains and the MSL2 CXC domain (Kim et al., 2010; Nielsen et al., 2005; Zheng et al., 2014). Consequently, the MSL complex is now one of the best structurally characterized epigenetic regulator complexes (Figure 6).

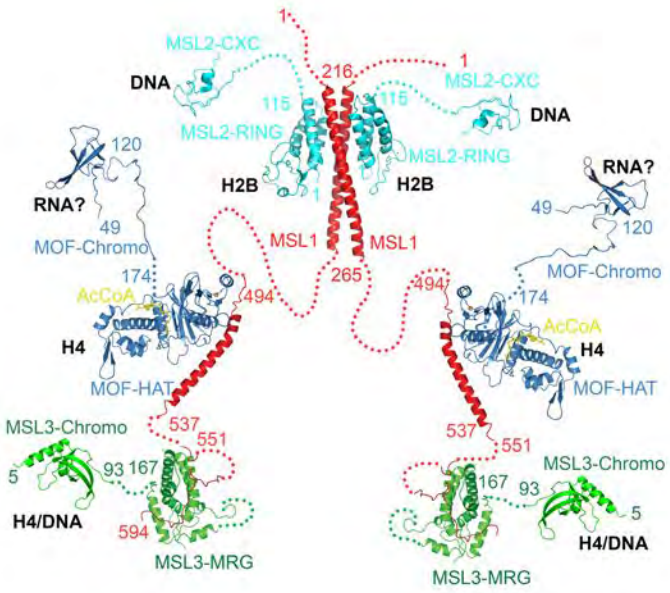


Figure 6. Summary of the structural information on the human MSL complex. A schematic model showing all known structures of the MSL sub-complexes and domains including our structures of the MSL1/MSL2 tetramer (Hallacli et al., 2012) and structures of MSL1 in complex with the MRG domain of MSL3 and the HAT domain of MOF (Kadlec et al., 2011). Also shown are structures of chromo-barrel domains of MSL3 (Kim et al., 2010) and MOF (Nielsen et al., 2005) and the CXC domain of MSL2 (Zheng et al., 2014). These structures, together with sequence alignments and secondary structure and disorder predictions suggest that the MSL complex is characterized by ordered functional domains separated by extensive natively disordered and flexible regions (dotted line). As it is currently unknown whether any additional interactions exist among the individual proteins and their domains, this model should be regarded only as schematic. Putative histone and DNA/RNA binding sites are labelled.

First, we determined two crystal structures of consecutive fragments of human MSL1 C-terminus with either the HAT (histone acetyltransferase) domain of MOF and acetyl-coenzyme A (Ac-CoA) or the MRG domain of MSL3 (Kadlec et al., 2011) (this work included supervision of a Master student Michael Lipp, University of Munich). These structures provided new insight into the regulation of the catalytic mechanism of MOF by the auto-acetylation of a conserved lysine residue within the

catalytic site and represent the first structures of HAT and MRG domains in complex with their binding partners. This work also participated in development of a new humidifier device for room temperature diffraction data collection (Sanchez-Weatherby et al., 2009).

Next, we determined the crystal structure of the complex between the predicted coiled-coil region of human MSL1 and the N-terminal portion of MSL2 (Hallacli et al., 2012) (co-supervision of an EMBL PhD student Michael Lipp, first co-author). Unexpectedly, this structure revealed that these two proteins form a heterotetramer, where two MSL1 subunits form a dimeric coiled-coil which serves as a binding platform for two molecules of MSL2. The fact that MSL1 dimerizes, completely changed our view of the dosage compensation complex structure and assembly, as it suggest that the entire complex contains all the subunits in pairs, including MSL3 and MOF. We generated mutants in *Drosophila* MSL1 based upon the structures that selectively disrupt MSL3, MOF and MSL2 binding and the MSL1 dimerization. We showed that MSL1 dimerization is required for incorporation of MSL2 into the MSL complex. Using ChIP (chromatin immunoprecipitation) assays we showed that MSL2 and the MSL1 dimerization are required for targeting of the complex to the X-chromosome, while MSL3 and MOF are required mostly for binding to low affinity sites. Finally, we showed that loss of MSL1 dimerization is lethal in flies (Hallacli et al., 2012). Interestingly, recent structure of the MSL CXC domain in complex with the high affinity site DNA sequence revealed that MSL2 recognizes its target sequence as a dimer (Zheng et al., 2014)

MSL complex - future plans

All our structures of the MSL sub-complexes were determined using human proteins. Compared to their *Drosophila* counterparts, the human proteins are significantly shorter with many of the disordered regions missing. While the *Drosophila* MSL complex contains also the MLE RNA helicase and roX1 and roX2 non-coding RNAs, these are most likely not present in the human complex. It has been proposed that within the *Drosophila* complex the MSL2 interactions with roX RNA depends on a prior remodelling of stem-loop structures within roX by MLE (Ilik et al., 2013; Maenner et al., 2013). Recently, the structure of the MLE helicase in complex with the corresponding roX RNA motif was reported (Prabu et al., 2015). The architecture of human MSL complex thus seems a bit more simple and more suitable for structural analysis.

The next step towards understanding the MSL complex architecture will be to obtain structural information on the entire complex. We will now produce the ternary MSL1/MSL2/MSL3/MOF complex in insect cells by co-expression using the Multibac system. We will use our structural data on individual sub-complexes to rationally engineer individual subunits. Once sufficient amount of homogenous material is obtained we will use it in several experiments. Apart from crystallization trials the complex will also be biophysically characterized. Mass spectrometry analysis combined with chemical crosslinking will be used to reveal any additional interactions among the MSL subunits. The complex will also be analysed using cryo-electron microscopy. The structures of the individual sub-complexes will then be docked into the resulting map to reveal the overall architecture of the complex. Finally, we will also study the interaction of the MSL complex with a nucleosome. The interaction between the MSL complex and the nucleosome will probably require specific histone modifications, which still need to be characterized. But this might be an efficient way to stabilize such a flexible complex. Hypotheses derived from our structural work will then be tested in functional experiments in *Drosophila* cell lines or transgenic flies in the Akhtar lab. Thus, this project will significantly contribute to a better understanding of the mechanism of dosage compensation in *Drosophila* as a unique model of epigenetic transcription regulation of an entire chromosome.

nature structural & molecular biology

Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1

Jan Kadlec[1,2,5], Erinc Hallacli[3–5], Michael Lipp[1,2], Herbert Holz[3,4], Juan Sanchez-Weatherby[1,2,4], Stephen Cusack[1,2] & Asifa Akhtar[3,4]

The male-specific lethal (MSL) complex is required for dosage compensation in *Drosophila melanogaster*, and analogous complexes exist in mammals. We report structures of binary complexes of mammalian MSL3 and the histone acetyltransferase (HAT) MOF with consecutive segments of MSL1. MSL1 interacts with MSL3 as an extended chain forming an extensive hydrophobic interface, whereas the MSL1-MOF interface involves electrostatic interactions between the HAT domain and a long helix of MSL1. This structure provides insights into the catalytic mechanism of MOF and enables us to show analogous interactions of MOF with NSL1. In *Drosophila*, selective disruption of Msl1 interactions with Msl3 or Mof severely affects Msl1 targeting to the body of dosage-compensated genes and several high-affinity sites, without affecting promoter binding. We propose that Msl1 acts as a scaffold for MSL complex assembly to achieve specific targeting to the X chromosome.

Dosage compensation is an essential process that equalizes the expression levels of X-chromosomal genes between males and females. In *Drosophila* males, dosage compensation results in an approximately two-fold upregulation of the transcription of X-linked genes[1]. In mammals, the balance in X-chromosomal gene expression is achieved by stochastic inactivation of one of the female X chromosomes[2]. Dosage compensation of X-chromosomal genes in *Drosophila* has been extensively studied as a model of chromosome-wide transcription regulation by histone hyperacetylation[3,4].

In *Drosophila*, the process is mediated by the dosage-compensation complex (DCC), also known as the MSL complex, which consists of at least five male-specific lethal proteins (Msl1, Msl2, Msl3, maleless (Mle) and males-absent-on-the-first (Mof)) and two functionally redundant noncoding RNAs (roX1 and roX2)[3]. In humans, an equivalent complex is formed from the counterparts of at least four *Drosophila* Msl proteins (MSL1, MSL2, MSL3 and MOF), although no associated RNA has yet been identified[5–8]. The *Drosophila* MSL proteins and roX RNAs are proposed to assemble and coat the X-chromosome in a process involving at least two steps. First, numerous high-affinity sites that are enriched for GA repeat sequences, including the roX genes, are occupied. MSL complexes then spread from these sites to the rest of the X chromosome including many active genes[9]. The MSL complex is responsible for acetylation of histone H4 at lysine 16 (H4K16ac) on the X chromosome mediated by Mof[10–12]. The details of the recruitment of the MSL complex to the X chromosome and the exact mechanism of dosage compensation remain poorly understood.

Human and *Drosophila* MSL1 consist of 614 and 1,039 amino acid residues, respectively, with no known globular domains predicted. In *Drosophila*, Msl1 was shown to interact with Msl2 via its N-terminal putative coiled coil[13] and with Msl3 and Mof via a conserved C-terminal region called the PEHE domain[14]. Msl3 (521 residues in human) interacts via its *MORF4*-related gene family (MRG) domain with Msl1 (ref. 15), whereas its N-terminal chromo-barrel domain binds nucleic acids[15] and is required for the spreading of the MSL complex along the X chromosome[16]. Mof (458 residues in human) is a member of the MYST family of acetyltransferases[10–12]. The Mof chromo-barrel domain is required for RNA binding[17], and the zinc finger within the HAT domain interacts with Msl1 (ref. 14). To have full enzymatic activity and specificity, Mof is required to be in a complex with Msl1 and Msl3 (ref. 14). Recently, it has been found that MOF is also a key component in a second, distinct chromatin-modifying complex called the NSL complex, wherein it interacts with NSL1, which has sequence similarity with MSL1 (refs. 7,18,19).

Although MSL complex members have been intensely studied over the last decade, the detailed molecular interactions within the complex remain unknown. As a first step toward achieving an atomic-resolution understanding of the architecture and function of the MSL complex, we determined crystal structures of two mammalian subcomplexes that MSL1 forms with the MOF HAT domain and with the MRG domain of MSL3. Based on the structural results, we carried out mutagenesis in the *Drosophila* orthologs to selectively disrupt the interaction of Msl1 with either Msl3 or Mof *in vivo*, thus demonstrating that the residues critical for the observed protein-protein interactions are conserved throughout evolution. Furthermore, using the sequence similarity between MSL1 and NSL1, we could demonstrate that mutation of corresponding residues in NSL1 also selectively

[1]European Molecular Biology Laboratory, Grenoble, France. [2]Unit of Virus Host-Cell Interactions, Université Joseph Fouier–European Molecular Biology Laboratory–Centre National de la Recherche Scientifique (UJF-EMBL-CNRS), Unité Mixte Internationale 3265, Grenoble, France. [3]European Molecular Biology Laboratory, Heidelberg, Germany. [4]Present addresses: Max-Planck-Institut für Immunbiologie und Epigenetik, Freiburg im Breisgau, Germany (E.H., H.H., A.A.); Diamond Light Source Ltd., Didcot, UK (J.S.-W.). [5]These authors contributed equally to this work. Correspondence should be addressed to A.A. (akhtar@immunbio.mpg.de) or S.C. (cusack@embl.fr).

Received 3 March 2010; accepted 20 October 2010; published online 9 January 2011; doi:10.1038/nsmb.1960

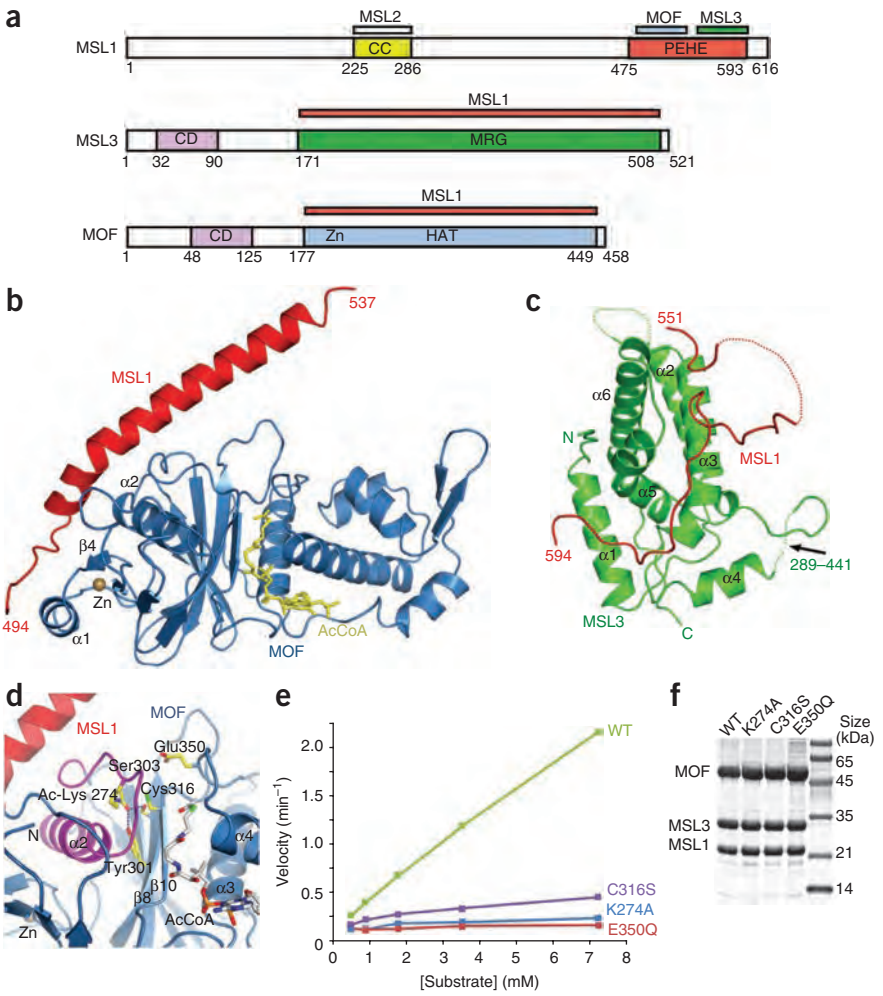


Figure 1 Crystal structures of the MSL1–MSL3 and MSL1–MOF subcomplexes. (**a**) Schematic representation of the domain structure of mouse MSL1 (which is essentially identical to human MSL1), MSL3 and MOF. Domain colors correspond to the ribbon diagram in **b** and **c**. The red, blue and green bars indicate MSL1, MOF and MSL3 interacting regions, respectively, as defined in this work and in ref. 13. CC, coiled coil; CD, chromo-barrel domain. (**b**) Ribbon diagram of the mammalian MSL1–MOF–acetyl-CoA complex. The HAT domain of MOF (residues 174–458) is shown in blue. The MOF secondary structures interacting with MSL1 are labeled. (**c**) Ribbon representation of the complex between MSL1 and MSL3. The MSL3 MRG domain (residues 167–288, 442–517) is shown in green, and its secondary structures are labeled. The disordered regions in MSL1 and MSL3 are shown as dots. The arrow indicates the location where residues 289–441 were deleted. (**d**) Ribbon diagram of the catalytic site of the MOF HAT domain. Acetyl-CoA is shown as sticks, with the acetyl moiety close to Cys316 and the catalytic Glu350. The hairpin structure formed by residues 257–281, shown in magenta, interacts with MSL1 (details in **Fig. 2**) and harbors acetylated lysine 274 (Ac-Lys274) that binds to β8. (**e**) The acetylation activity of human wild-type (WT), K274A, C316S and E350Q MOF was assessed within the MOF–$MSL1_{471–616}$–$MSL3_{167–289,442–517}$ complex in a fluorescence based HAT assay (DNTB assay), measuring the production of CoA during the acetylation of a histone H4 N-terminal peptide. The average of two initial velocity measurements is plotted against the substrate concentration. The rates obtained between duplicate runs were typically within 10% of each other. (**f**) Qualitative assessment of the purity of MSL complexes shown by Coomassie staining.

affects its interaction with MOF. Notably, we showed using chromatin immunoprecipitation (ChIP) assays that selective disruption of Msl1 interaction with either Mof or Msl3 severely affects targeting of Msl1 to the coding regions and to the 3′ end of X-linked genes, whereas Msl1 binding to promoters is largely independent of Msl3 or Mof *in vivo*. Finally, we showed that high-affinity sites differ in their requirement of Msl3 and Mof for Msl1 recruitment.

RESULTS

Structure of the MSL1–MOF complex

The structure of a complex between the HAT domain of human MOF (residues 174–458), the N-terminal part of the MSL1 PEHE region (residues 470–540) and acetyl-CoA was determined by X-ray crystallography at 2.8-Å resolution and refined to an R_{free} of 25.6% and an *R*-factor of 22.3% (**Fig. 1** and **Table 1**).

As with other MYST-family HAT domains, the MOF structure consists of a central core that participates in cofactor binding with flanking N- and C-terminal regions (**Fig. 1b** and **Supplementary Fig. 1**). In the N-terminal segment, residues 205–233 form a zinc-binding module, with the absolutely conserved Cys210, Cys213, His226 and Cys230 coordinating the Zn atom.

First, we confirmed that Glu350, in a position corresponding to the putative catalytic Glu338 of Esa1 (ref. 20) (**Fig. 1d**), is also likely to be catalytic residue in MOF, as the E350Q mutation essentially abolished the MOF HAT activity in fluorescence-based acetylation assays on a histone H4 N-terminal tail peptide (**Fig. 1e**,**f**).

Helix α2 and the downstream chain form a hairpin-like structure (residues 257–281), within which Lys274 showed an additional electron density at its tip that could be modeled as an acetyl group (**Fig. 1d** and **Supplementary Fig. 2a**). Lys274 acetylation was confirmed by mass spectrometry analysis (**Supplementary Fig. 2b**). The acetylated lysine is involved in the otherwise hydrophobic interface of the hairpin and strand β8 of the core β-sheet (**Fig. 1d**). Consistent with this, K274A MOF was substantially less stable in a thermal denaturation assay (**Supplementary Fig. 2c**,**d**) and was not active in acetylation assays (**Fig. 1e**,**f**). Notably, Lys274 was acetylated even in the E350Q mutants, probably through a residual HAT activity of this mutant (**Supplementary Fig. 3a**,**b**). Thus, the acetylation of Lys274 seems to be important for structural integrity and proper positioning of the residue 257–281 hairpin, which, in analogy to the structure of the Gcn5 HAT domain in a complex with H3 peptide[21], might be involved in substrate binding. Structurally important lysine acetylation in the core of the HAT domain appears to be a conserved feature at least among mammalian MYST proteins, as the corresponding lysine residues in Tip60 and MOZ are also modeled as acetylated, forming equivalent hydrogen bonds in their deposited structures (PDB codes 2OU2 and 2OZU, respectively). It remains to be established whether this modification is involved in regulation of the HAT activity of MYST acetyltransferases, as proposed for p300/CBP[22].

The next important residue to consider in the catalytic site of MOF was the conserved Cys316. The corresponding cysteine in Esa1 was shown to be acetylated, and a two-step catalytic mechanism was

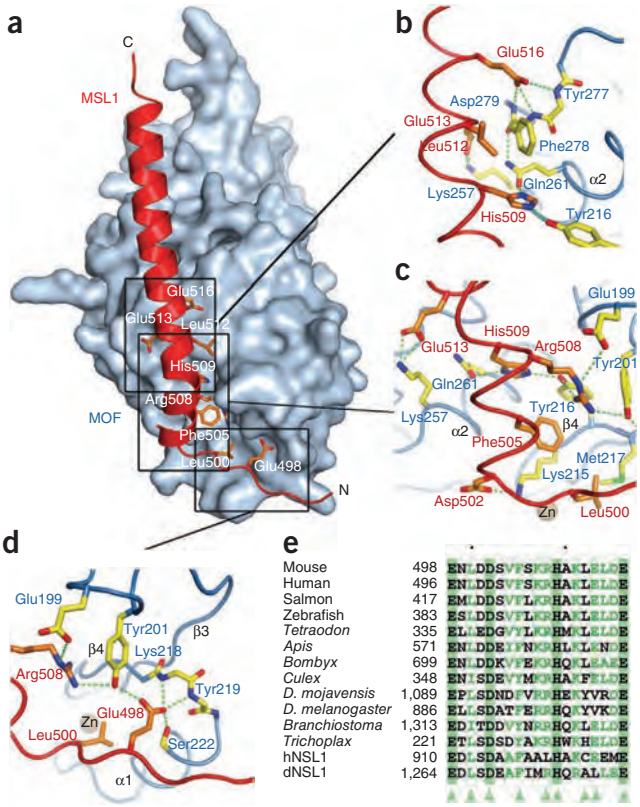


Figure 2 Structure of the MSL1–MOF complex. (**a**) Cartoon representation of MSL1 (residues 471–539, shown in red) in complex with the MOF HAT domain (shown as solvent-accessible surface in blue). MSL1 residues involved in the interaction are drawn as orange sticks. The three overlapping details of the interface shown in **b**–**d** are localized on the structure by the black boxes. In the detailed figures, MSL1 residues are shown in orange (labeled in red) and MOF residues in yellow (labeled in blue). (**b**) The interaction between MSL1 and the MOF hairpin including helix α2 (residues 257–281). MSL1 Glu516 forms hydrogen bonds with main chain amide groups of Tyr277, Phe278 and Asp279 (for clarity, side chains of Tyr277 and Asp279 are omitted). (**c**) Contacts of MSL1 with helix α2 and strand β4 of the MOF zinc finger. Hydrogen bonds of MSL1 Asp502, Arg508, His509 and Glu513 with MOF residues are shown in green. Phe505 inserts into a hydrophobic pocket of MOF. (**d**) Details of the interaction between MSL1 and the zinc finger of MOF (α1, β4). Glu498 hydrogen bonds main chain amide groups of Lys218 and Tyr219 as well as hydroxyl groups of Tyr201 and Ser222 of MOF (side chains of Lys218 and Tyr219 are not shown). (**e**) Sequence alignment of the MSL1 and NSL1 proteins. Only the sequence of the fragment involve in the interaction with MOF is shown. Identical residues are in green boxes. Green triangles indicate the interacting residues.

proposed for MYST acetyltransferases, whereby the acetyl moiety is transferred from acetyl-CoA to the substrate lysine via a cysteine residue[23]. Contrary to this, a previous biochemical study showed that Cys304 of Esa1 is dispensable for catalysis[24]. In our MOF structure, Cys316 is not acetylated, and mass spectrometry analysis identified only chymotryptic fragments containing unmodified Cys316 (data not shown). Notably, we showed that the C316S mutant is still partially active in the HAT assay (**Fig. 1e**,**f**) and that Lys274 is acetylated in this mutant (**Supplementary Fig. 3c**). The lower activity of the mutant (**Supplementary Fig. 3d**–**f**) might reflect local conformational changes due to the substitution. These data suggest that MOF does not use the two-step catalytic mechanism originally proposed for Esa1.

Next, we analyzed the interaction interface between the MOF HAT domain and MSL1. The MSL1 fragment forms a loop (residues 494–501) followed by a 52-Å-long helix (residues 502–533). Both elements interact extensively with the N-terminal part of the MOF HAT domain (**Figs. 1b** and **2** and **Supplementary Fig. 4**), with numerous, mainly charged contacts existing between the two molecules. The complex interface buries 878 Å^2 of MOF and 1,050 Å^2 of MSL1. In MOF the interaction involves three regions: helix α1 and strand β4 of the zinc finger, an upstream loop connecting β2 and β3 (residues 197–205), and the aforementioned hairpin (residues 257–281). Details of these interactions are shown in **Figure 2**. In MSL1, the key interacting residues include Glu498, Asp502, Arg508, His509, Glu513 and Glu516, which form multiple hydrogen bonds and salt-bridge interactions with MOF. Additionally, Leu500, Phe505 and Leu512 are inserted in hydrophobic pockets in the center of the interface. All the MSL1 interacting residues are very well conserved among species, reflecting the importance of this interaction for the functional integrity of the MSL complex (**Fig. 2e** and **Supplementary Fig. 5**). Mapping MOF residue conservation across species onto the surface of the MOF HAT domain revealed that the MSL1-interacting region is rather well conserved, despite the fact that the interface involves several main chain contacts (**Supplementary Figs. 4a** and **6**). The key MSL1 binding residues in MOF are Glu199, Tyr201, Tyr216, Gln261 and Phe278. Because MSL1 fragments interacting with MOF and MSL3 are unstable when expressed alone, further binding-affinity measurements could not be performed *in vitro*.

MYST acetyltransferases function exclusively within multiprotein complexes. Given the high structural similarity among the MYST family members, a corresponding surface including the zinc finger is available for protein-protein interactions in other MYST proteins as well (**Supplementary Fig. 4b**,**d**). Indeed, within human MYST acetyltransferases, most of the residues corresponding to the MSL1 binding surface of MOF are conserved (**Supplementary Figs. 4c** and **7**). Consistent with this, a two-hybrid screen has shown that the zinc finger of human HBO1 (MYST2) HAT is essential for the interaction with MCM2 of the minichromosome maintenance complex[25]. An MCM2 L222A mutant deficient in HBO1 binding could be reverted by mutations of Ile380 in HBO1 (ref. 25), corresponding to MOF Ser222 on helix α2, which is directly involved in the interaction with MSL1 (**Supplementary Fig. 7**). It is thus likely that HBO1 uses a similar surface for interaction with MCM2.

Structure of the MSL1–MSL3 complex

The MSL3 construct used here corresponds to MSL3 isoform c (residues 167–517), including the predicted MRG domain (**Fig. 1a**). Compared to the sequence of the known structure of the MRG domain of MRG15 (PDB entry 2F5J)[26], the human and *Drosophila* MSL3 domains contain two poorly conserved insertions with no predicted secondary-structure elements (residues 223–250 and 290–441 in human MSL3; **Supplementary Fig. 8**). To obtain diffracting crystals, the longer insertion was replaced with an eight-residue linker. This construct still bound efficiently MSL1 in a binary or ternary complex with MOF (**Supplementary Fig. 9**). The structure of the complex between the MSL1 PEHE region (residues 545–597) and $MSL3_{167–289,442–517}$ was determined at 3 Å resolution. The refinement resulted in an R_{free} of 25.3% and an R-factor of 23.2% (**Table 1**).

The MSL3 MRG domain structure is similar to that of MRG15 (r.m.s. deviation 0.99 Å for 151 Cα atoms; **Fig. 1c**). Electron density is missing for the short insertion connecting α2 and α3 (residues 224–245). The 151-residue region deleted from the MSL3 construct links helices α4 and α5 (**Fig. 1c**). Interpretable electron density

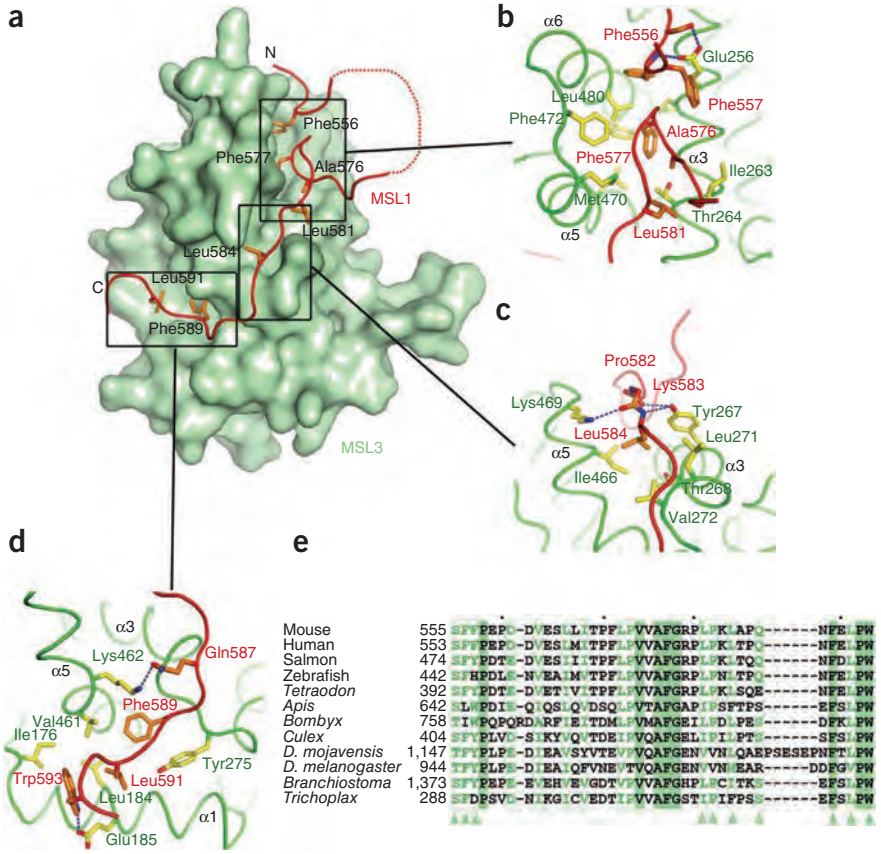


Figure 3 Structure of the MSL1–MSL3 complex. (**a**) Structure of MSL1 (residues 545–597, shown in red) in complex with the MSL3 MRG domain (shown as solvent-accessible surface in green). MSL1 residues involved in the interaction are drawn as orange sticks. The three black boxes correspond to the detailed views of the complex interface shown in **b**–**d**, where the MSL1 residues are shown in orange (labeled in red) and MSL3 residues in yellow (labeled in green). (**b**) Ala576 and Phe577 insert into hydrophobic pocket formed by helices α3, α5 and α6. Additionally, the N-terminal Phe556 and Phe557 bind on the top of Phe577, reinforcing this interaction. MSL3 Glu256 of helix α3 hydrogen bonds Ser555 and the main chain of Phe556. (**c**) The MSL1 binds between helices α5 and α3. Leu584 inserts into a cavity between the helices. MSL1 forms several main chain interaction with MSL3 Tyr267 and Lys469 (side chains of Pro582 and Lys583 are not shown). (**d**) MSL1 Phe589 and Leu591 insert into pocket formed by hydrophobic residues of helices α5 and α1 and a linker between helices α3 and α4. Additionally, Gln587 and Trp593 form hydrogen bonds with Lys462 and Glu185, respectively. (**e**) Sequence alignment of the MSL1 fragment involved in the interaction with MSL3. Identical residues are in green boxes. The interacting residues are indicated with green triangles.

was observed for residues 551–558 and 564–594 of MSL1. In the crystal, the N- and C-terminal parts of the MSL1 PEHE region bind to two distinct MSL3 molecules, which we interpret as being a result of nonphysiological domain swapping (see Online Methods and **Supplementary Fig. 10a**). We showed by multiangle laser light scattering (MALLS) that a ternary complex of MOF HAT–MSL1 PEHE–MSL3$_{167–289,442–517}$ is formed with an apparent 1:1:1 stoichiometry (**Supplementary Fig. 10b**).

MSL1 wraps around the MSL3 MRG domain as an extended chain (**Figs. 1c** and **3a**). The complex interface buries 1,307 Å^2 of MSL3. MSL1 forms numerous hydrophobic as well as several charged interactions with MSL3. The crucial interacting residues of MSL1 are four highly conserved phenylalanines (Phe556, Phe557, Phe577 and Phe589) that insert into different hydrophobic pockets on MSL3 (**Fig. 3**). The C-terminal part of the MSL1 peptide forms a short hairpin harboring Ala576 and Phe577 that inserts into a cavity formed by hydrophobic residues of the helical hairpin α5-α6 and a perpendicular helix α3. This interface is reinforced by the N-terminal part of the MSL1 peptide, which binds on the top of the MSL1 hairpin with Phe556 positioned between helices α6 and α3. MSL1 then folds around the last turn of α3 to place Phe589 and Leu591 into another hydrophobic surface formed by helices α5 and α1. Details of these interactions are given in **Figure 3**. Essentially all MSL3 and MSL1 residues involved in the interaction are well conserved among species (**Fig. 3e** and **Supplementary Figs. 5**, **8** and **10c**,**d**).

The MSL1–MSL3 structure is, to our knowledge, the first reported MRG domain in complex with its binding partner. Previous studies revealed that mutations of MRG15 residues corresponding to MSL3 Leu480 and Phe484, which form the hydrophobic pocket surrounding MSL1 Phe577 and Phe556, abolish the interaction of MRG15 with MRGBP[27]. Additionally, the same surface is involved in MRG15 dimerization, in which this hydrophobic pocket accommodates Tyr276 and Leu279 of another protomer[26]. MRG15 residues corresponding to the hydrophobic surface binding MSL1 Phe589 are involved in the interaction with the N terminus of PAM14 (ref. 26). Thus, similar hydrophobic surfaces appear to be generally used for protein-protein interactions in MRG domains. Mapping of phylogenetically conserved residues onto the surface of MSL3 revealed an additional highly conserved region formed by charged residues of helices α2 and α3 (**Supplementary Fig. 10e**) that might be involved in another protein-protein or protein–nucleic acid interaction.

The HAT activity of MOF is enhanced by MSL1 and MSL3

In *Drosophila*, the Mof HAT activity is enhanced by the presence of Msl1 and Msl3 (ref. 14). We were interested in identifying the minimal regions required for this activity and in determining whether the fragments used for crystallization are sufficient for enhanced acetylation. We therefore copurified mammalian complexes containing the MOF HAT domain with various MSL1 PEHE and the MSL3 MRG domain fragments (**Supplementary Fig. 9a**,**b**). The full-length *Drosophila* trimer complex containing Mof, Msl1 and Msl3 was used as a positive control (**Supplementary Fig. 9c**). Consistent with previous observations, the presence of Msl1 and Msl3 enhanced the HAT activity of *Drosophila* Mof on nucleosomal substrates (**Fig. 4a**; compare lanes 11 and 12 with lanes 9 and 10). Notably, the mammalian complex containing the human MOF HAT, MSL1 PEHE domain and MSL3 MRG domain was sufficient for enhancing the acetylation activity (**Fig. 4a**, lanes 5 and 6). However, removing the large insertion within the MSL3 MRG domain (residues 290–441) did not enhance HAT activity to a similar extent, indicating that this segment is required for the full activation potential of MOF (**Fig. 4a**, lanes 7 and 8). These results show that in the mammalian system as well, interaction of MOF with MSL3 via MSL1 enhances the HAT activity of MOF, whereas stimulation is not observed with a subcomplex containing only MOF and

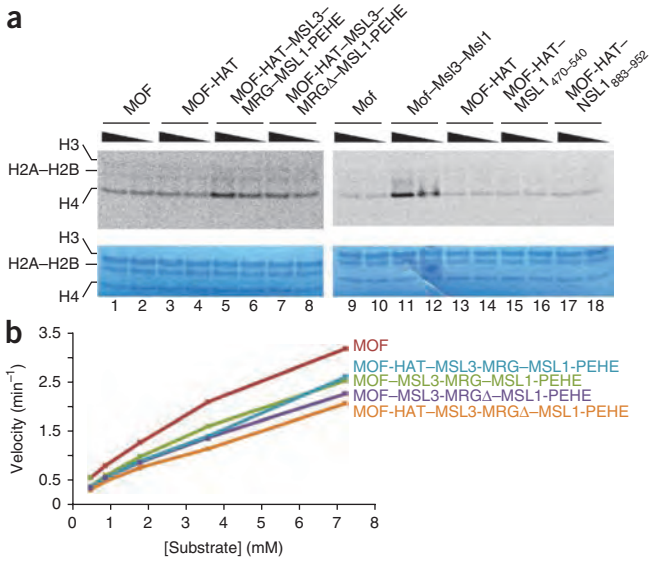


Figure 4 Acetylation activity of the MSL subcomplexes. (**a**) Acetylation assay on native nucleosomes. Each reaction contains 1,500 pM and 750 pM of the indicated mammalian and *Drosophila* MOF or MSL subcomplexes (**Supplementary Fig. 9a–c**), respectively, and 1.5 μg of native nucleosomes purified from MCF-7 cells. Upper panels show the radioactive signals, and lower panels show the Coomassie staining of the nucleosomes used in the reactions. The MSL1 PEHE domain contains residues 471–616, the human MSL3 MRG domain contains residues 167–517, and MSL3-MRGΔ contains residues 167–289 and 442–517. (**b**) Fluorescence-based HAT assay (carried out as in **Fig. 1e**) of the acetylation activity of human MSL subcomplexes containing either the full-length MOF or its HAT domain, the MSL1 PEHE domain and the MSL3 MRG domain or its deletion derivative MSL3-MRGΔ (**Supplementary Fig. 9d**).

$MSL1_{470–540}$ (**Fig. 4a**, lanes 15 and 16). We also tested the HAT activity of these ternary MSL complexes containing MOF HAT or full-length MOF in fluorescence-based assays on a histone H4 N-terminal peptide. No increase in activity was observed for any of the complexes compared to MOF alone, confirming that boosting takes place only in the presence of the entire nucleosome (**Fig. 4b**).

MSL1 mutagenesis

To test the importance of the principal interacting residues for the stability of the MSL1–MOF and MSL1–MSL3 complexes and to identify MSL1 mutants that prevent complex formation, we generated several constructs which we predicted to disrupt key interactions (**Fig. 5a** and **Supplementary Fig. 11**). We coexpressed the same protein fragments used for crystallization and performed pulldown assays with histidine (His)-tagged MSL1. We mutated three MSL1 residues that seemed essential for the binding of MOF. Each of these single point mutations, E498R, F505R and H509R, completely abolished the interaction *in vitro* (**Fig. 5b**, lanes 1–3).

To test the interaction between MSL1 and MSL3, we first prepared mutations in the short MSL1 hairpin (F577E, A576E). Both mutants copurified with MSL3 just like the wild-type protein, suggesting that the remaining contacts are sufficient for binding (**Fig. 5c**, lanes 1 and 2). Indeed, the MSL3 binding region of *Drosophila* Msl1 was originally mapped to residues 973–1039 (ref. 14), which overlap with only 14 residues of our construct (including Phe589), suggesting that residues 584–597 are sufficient to bind MSL3. F577E F589E and A576E F589E double mutations substantially reduced the binding (**Fig. 5c**,

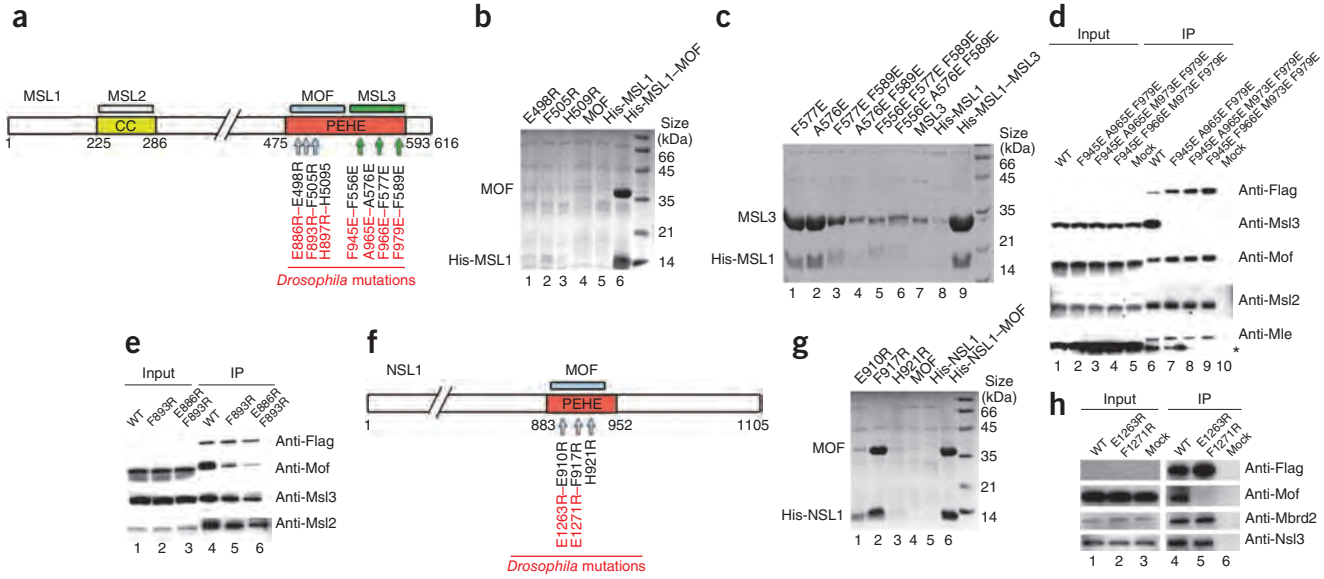


Figure 5 Mutagenesis of the MSL1 and NSL1 residues involved in MOF and MSL3 binding. (**a**) Schematic representation of domain structure of mouse MSL1. The binding sites for MSL2 (white), MOF (blue) and MSL3 (green) are shown. Individual mutations are indicated by arrows. (**b**) SDS-PAGE analysis of the binding of His-tagged $MSL1_{470–540}$ mutants (indicated above the lanes) to coexpressed $MOF_{174–458}$ after purification using Ni^{2+} resin. (**c**) Analysis of MSL1 mutant binding to MSL3 (as in **b**). (**d**) *Drosophila* Msl1 and Msl3 interaction *in vivo* in SL-2 cells transiently transfected with wild-type (WT) Ms11 and the indicated mutants. Shown are data from immunoprecipitation (IP) of whole-cell extracts taken 48 h after transfection using Flag-agarose resin. Mock lane represents the empty vector. The slight running difference between INPUT and IP lanes is due to different denaturing buffers. Anti-Flag antibody was used to detect exogenous MSL1 proteins. Transient transfections are always below the limit of detection in the INPUT lanes. Asterisk in MLE protein blot is a nonspecific cross-reacting band. (**e**) *Drosophila* Msl1 and Mof interaction *in vivo*: the experiment was performed as described in **d**. (**f**) Schematic representation of domain structure of human NSL1. The binding site for MOF is shown in blue above the PEHE region. Individual mutations in these regions are indicated by arrows. (**g**) MOF binding of $NSL1_{883–952}$ mutants, tested as in **b**. (**h**) *Drosophila* Nsl1-Mof interaction *in vivo*. Shown are data from IP of transiently expressed mutant carried out as in **d**,**e**.

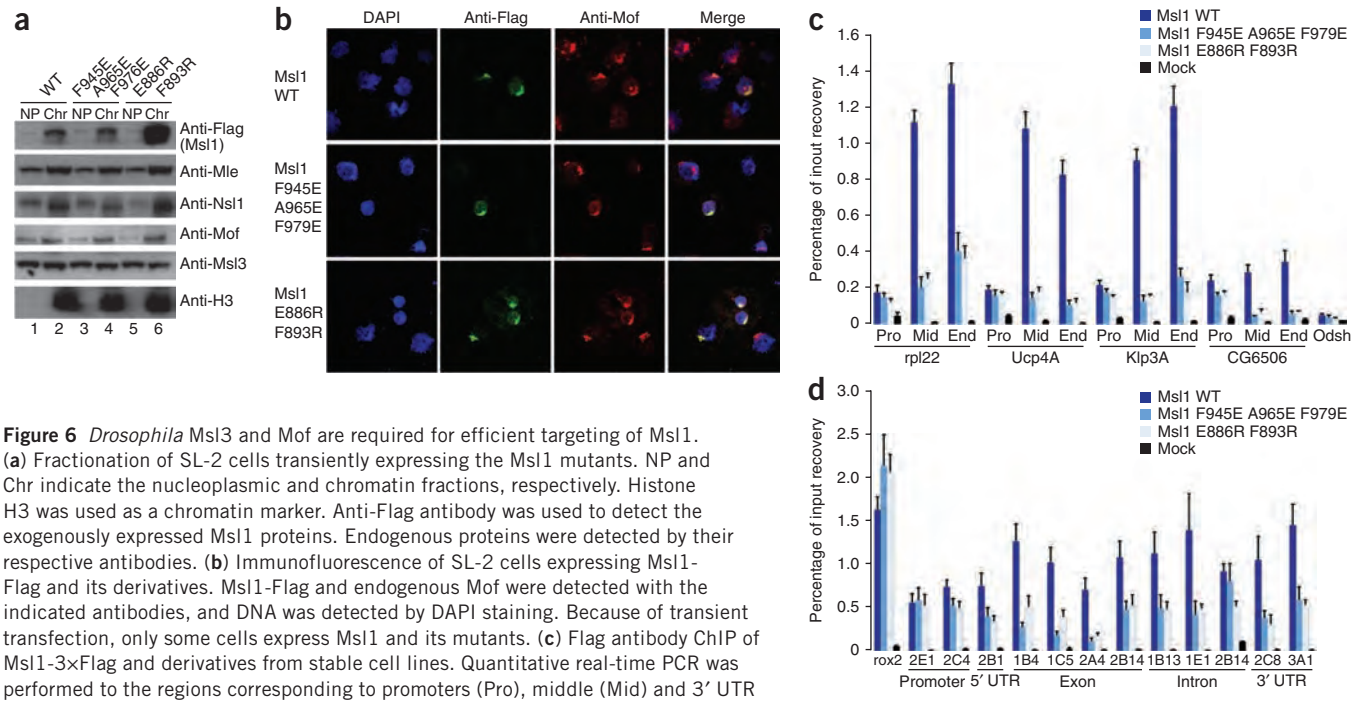


Figure 6 *Drosophila* Msl3 and Mof are required for efficient targeting of Msl1. (**a**) Fractionation of SL-2 cells transiently expressing the Msl1 mutants. NP and Chr indicate the nucleoplasmic and chromatin fractions, respectively. Histone H3 was used as a chromatin marker. Anti-Flag antibody was used to detect the exogenously expressed Msl1 proteins. Endogenous proteins were detected by their respective antibodies. (**b**) Immunofluorescence of SL-2 cells expressing Msl1-Flag and its derivatives. Msl1-Flag and endogenous Mof were detected with the indicated antibodies, and DNA was detected by DAPI staining. Because of transient transfection, only some cells express Msl1 and its mutants. (**c**) Flag antibody ChIP of Msl1-3×Flag and derivatives from stable cell lines. Quantitative real-time PCR was performed to the regions corresponding to promoters (Pro), middle (Mid) and 3′ UTR of the genes (End). All tested genes are dosage compensated and located on the X chromosome. ODSH upstream region was used as a negative control for the binding of the MSL complex. Each bar represents the average of four independent immunoprecipitations (IPs), and error bars indicate the s.d. Mock sample corresponds to wild-type cells. (**d**) Flag antibody ChIP of Msl1-3×Flag and derivatives on high-affinity sites. Thirteen high-affinity sites were chosen from ref. 9 and were grouped according to their genomic positions. The error bars indicate the s.d. of four independent IPs. Note that the two 2B14 sites are different (see **Supplementary Table 1** for the primer pairs used in the study).

lanes 3 and 4), and the interaction was essentially abolished in triple mutants with substitutions that interfere with hydrophobic contacts in the three distinct MSL3 regions (F556E A576E F589E and F556E F577E F589E; **Fig. 5c**, lanes 5 and 6).

To study the integrity of the mutated proteins and the impact of the MSL1 mutations on the incorporation of MOF and MSL3 into the MSL complex *in vivo*, we prepared corresponding mutations in full-length MSL1. Because the MSL proteins are evolutionary conserved and their role is better understood in *Drosophila*, we studied the effect of these mutations using *Drosophila* Msl1 (for the comparison of corresponding amino acid mutations between human and *Drosophila*, see **Fig. 5a**). To assure a complete loss of Msl3 binding, we introduced an additional mutation M973E (L584E in human). We expressed Msl1-Flag proteins in SL-2 cells and immunoprecipitated the corresponding MSL complexes using an anti-Flag antibody–coupled resin. All tested Msl1 mutations in the C-terminal part of the PEHE region were unable to copurify endogenous Msl3, whereas they had no effect on the remaining MSL components, Mof, Msl2 and Mle (**Fig. 5d**, lanes 7–9). In the case of Mof binding-site mutants, a reduced binding of Mof was obtained for the single F893R mutation in the N-terminal portion of the Msl1 PEHE region. Further reduction was observed for the E886R F893R double mutant (**Fig. 5e**, lanes 5 and 6). A partial reduction was also observed for Msl3 and Msl2 incorporation, suggesting that the presence of Mof in the complex might be important for Msl1 stability (**Fig. 5e**). These results indicate that Msl3 and Mof are incorporated into the MSL complex via the Msl1 scaffold and show that at least Msl3 can be disassembled from the complex without any apparent effect on the molecular interactions of other members of the MSL complex.

The fact that the Msl1 F945E A965E F979E triple mutant immunoprecipitates Mof but not Msl3 raises questions about the importance of the interaction between Mof and Msl3 (ref. 28). To further investigate the putative Mof-Msl3 interaction[14], we performed MSL complex reconstitution assays with the full-length proteins expressed in Sf21 insect cells. In the presence of Msl1, Mof clearly copurifies with Msl3-Flag (**Supplementary Fig. 11d**, lane 4), whereas in the absence of Msl1, the interaction could be seen only using western blot detection (**Supplementary Fig. 11d**, lane 2, and **Supplementary Fig. 11e**, lane 2), suggesting that the Msl3-Mof interaction does occur *in vitro*, albeit substantially more weakly than within the trimeric complex.

NSL1 and MSL1 bind MOF in a similar manner

Notably, the key interacting residues of MSL1 are also conserved in human and *Drosophila* NSL1 (also known as MSLv1 in human) proteins (**Fig. 2e**). Human NSL1 has recently been shown to interact with MOF within the NSL complex, which is involved in acetylation of p53 and H4K16 in male and female cells[7,18,19]. Indeed, a corresponding fragment of human NSL1 (883–952) is sufficient to form a stable complex with the MOF HAT domain *in vitro* (**Supplementary Fig. 12**). The MSL1 Arg508 forming three hydrogen bonds with MOF (**Fig. 2c**,**e**) is substituted with Leu920 in human NSL1; however, this arginine remains conserved in *Drosophila* Nsl1. Given the high sequence identity between MSL1 and NSL1 in this region, it is likely that their modes of interaction with MOF are very similar. We did not observe any increase of the MOF HAT activity in the presence of $NSL1_{883–952}$ alone, as had been seen in the presence of $MSL1_{470–540}$ (**Fig. 4a**).

To investigate the interaction between NSL1 and MOF, we prepared mutations E910R, F917R and H921R in human NSL1 and first tested the ability of these mutants to interact with the MOF HAT domain *in vitro*. H921R abolished and E910R considerably reduced the NSL1 binding to MOF (**Fig. 5f** and **Fig. 5g**, lanes 1 and 3). Next, we investigated whether NSL1 uses the same interaction surface for MOF contact *in vivo*. For this purpose, amino acids predicted to interact with Mof

Table 1 Data collection and refinement statistics

	MSL1–MOF–AcCoA	MSL1–MSL3
Data collection		
Space group	$I4_122$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	180.9, 180.9, 80.7	75.8, 127.1, 79.6
α, β, γ (°)	90, 90, 90	90, 118.4, 90
Resolution (Å)	48–2.7 (2.8–2.7)[a]	46–3.0 (3.12–3.0)
R_{merge}	8.6 (80.7)	3.4 (51.9)
$I/\sigma I$	14.05 (2.16)	16.37 (2.04)
Completeness (%)	99.1 (98.6)	93.5 (94.9)
Redundancy	4.0 (4.1)	2.2 (2.2)
Refinement		
Resolution (Å)	42–2.7	46–3.0
No. reflections	17,569	23,794
R_{work} / R_{free}	22.3 / 25.6	23.2/ 25.3
No. atoms		
Protein	2,641	6,492
Ligand	50	–
Zinc ion	1	–
Water	10	–
B-factors		
Protein	61	100
Ligand	45	–
Ion	56	–
Water	39	–
R.m.s. deviations		
Bond lengths (Å)	0.006	0.009
Bond angles (°)	0.945	1.118

[a]Values in parentheses are for highest-resolution shell.

were mutated in full-length *Drosophila* Nsl1, and the mutant proteins were expressed in SL-2 cells. Remarkably, the *Drosophila* Nsl1 mutant E1264R F1271R showed a severe loss of Mof interaction (**Fig. 5h**. lane 5), whereas the Mbdr2-Nsl3 interaction was preserved. This indicated that MOF uses similar surfaces for the integration into either NSL or MSL complexes in *Drosophila* cells. We therefore propose that specificity of Mof targeting depends on the differential interactions of Msl1 or Nsl1 with other members of the respective complexes.

Targeting of MSL1 to X-linked genes requires MOF and MSL3

On the basis of our high-resolution structures, we were able to design mutations in Msl1 that selectively disrupt its interaction with Mof or Msl3. We next investigated the impact of such mutations on the recruitment of Msl1 to X-chromosomal target genes. For these *in vivo* experiments, we selected Msl1 F945E A965E F979E, which lacks the interaction with Msl3, and Msl1 E886R F893R, which shows compromised Mof binding. To test whether the wild-type Msl1-Flag and its mutant derivatives are incorporated or targeted to chromatin, we initially performed cell fractionation assays (see **Supplementary Methods**). In wild-type cells, all MSL members have both nucleoplasmic and chromatin distributions, with an enrichment in the chromatin-bound pool (**Supplementary Fig. 13a**,**b**). Transiently expressed Msl1-Flag and the Msl1 F945E A965E F979E and E886R F893R mutants were detected mostly in the chromatin fractions, indicating that our constructs were incorporated into chromatin much like endogenous MSL complexes (**Fig. 6a**). Next, we tested whether X-chromosomal targeting is affected by the disruption of Msl1 interaction with either Mof or Msl3. So as not to exceed physiological protein levels, we expressed the constructs under the control of the copper-inducible MtnB promoter in uninduced conditions and used an anti-Flag antibody to visualize the exogenous Msl1-Flag and mutant derivatives by immunofluorescence (IF) microscopy. All constructs were correctly targeted to the X chromosome, showing colocalization with endogenous Mof (**Fig. 6b**).

Because immunofluorescence microscopy does not provide sufficient resolution to observe targeting to individual loci, we performed ChIP of the Msl1 derivatives on X-linked genes. For this purpose, we generated stable SL-2 cell lines that express 3×Flag epitope–tagged Msl1 and confirmed that Msl1-3×Flag showed binding profiles similar to those of the endogenous Msl1 (**Supplementary Fig. 13c**,**d**). ChIP experiments revealed that in contrast to the wild-type Msl1, Msl1 E886R F893R and Msl1 F945E A965E F979E showed considerably reduced binding on the body of X-linked genes (**Fig. 6c**). We also observed a low but consistent Msl1 signal around promoters of X-linked genes. Notably, this signal remained largely unaffected in the Msl1 mutants (**Fig. 6c**).

We next asked whether the compromised chromatin binding of the Msl1 mutants is restricted to low-affinity sites or whether targeting to high-affinity sites is also impaired. For this purpose, we chose 13 recently mapped high-affinity sites[9] and compared the binding profiles of wild-type Msl1 and its mutant derivatives to these sites. The roX2 gene was used as a control because it is a high-affinity site for MSL complex assembly and Msl1 binding on this site is independent of Msl3 or Mof[29]. Because high-affinity sites are located at different loci on the X chromosome, we separated them into positional categories (promoter proximal, 5′ untranslated region (UTR), exon, intron and 3′ end) in order to investigate any site-specific differences. Notably, this analysis revealed that disruption of Mof or Msl3 interactions also affected optimal binding of Msl1 to these high-affinity sites, especially those located away from promoter regions (**Fig. 6d**). However, for sites that are promoter proximal, such as 2E1 and 2C4, MSL1 mutants remained bound at comparable levels to the wild-type Msl1 (**Fig. 6d**).

DISCUSSION

In this study, we report the first crystal structures of two mammalian MSL subcomplexes containing two consecutive fragments of MSL1. These structures reveal how MSL1 uses short interacting peptides for the recruitment of MSL3 and MOF into the dosage-compensation complex. The conserved interactions of MOF with MSL1 and NSL1 explain how MOF can interact with two separate proteins complexes. It is tempting to speculate whether other proteins may also interact with MOF in a similar fashion.

Mutations of key interacting residues in *Drosophila* Msl1 revealed important information regarding its targeting to X-linked genes. In particular, we draw three important conclusions from these data. First, there appears to be a separate mode of Msl1 binding at the promoter region of target genes (regardless of low or high affinity) that seems to be independent of Msl3 or Mof interaction. These observations are noteworthy because Msl1 binding to promoters has not been appreciated previously. It will be interesting to investigate in future how other members of the MSL complex, such as Msl2, may influence targeting to specific chromatin regions in a similar manner.

Second, Msl1 binding to the coding regions of genes, as well as to high-affinity sites, requires efficient interaction with Msl3 or Mof. The latter observation is intriguing because high-affinity sites traditionally have been defined as sites where partial complexes of Msl1–Msl2 can bind in the absence of other components such as Msl3 or Mof[1,9]. However, our data revealed that optimal Msl1 binding to some high-affinity sites also requires Msl3 or Mof, suggesting that there is a more complex targeting mechanism than has been appreciated to date.

Third, these data reveal that there are qualitative differences among the high-affinity sites and that one possible cause of these variations could be

the positional clue within the chromatin environment. These site-specific chromatin characteristics, such as nucleosome-depleted promoter regions or differential chromatin marks within the coding regions[1], may work in combination with the postulated high-affinity site sequences[9].

Taken together, our data provide concrete structural evidence of the scaffolding role of Msl1 in the assembly and function of the MSL complex in *Drosophila* and mammals. Moreover, insights gained at atomic resolution provide us with the unique possibility of investigating the importance and mechanism of individual MSL complex members for transcription regulation and dosage compensation. Future structural investigations, including examination of additional components of the MSL complex, promise to broaden our understanding of how this chromatin remodeling machine is targeted to the X chromosome and upregulates transcription by two-fold.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Atomic coordinates and structure factors for the MSL1–MSL3 and MSL1–MOF complexes have been deposited with the Protein Data Bank under accession codes respectively 2Y0N and 2Y0M.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank K. Dzeyk of EMBL Heidelberg and I. Berard of the Institut de Biologie Structurale (IBS) Grenoble for their help with mass spectrometry; the High Throughput Crystallisation Laboratory (HTX) group of EMBL Grenoble for performing initial screening crystallization trials; M. Jamin's group for help with the MALLS experiment; the European Synchrotron Radiation Facility (ESRF)-EMBL Joint Structural Biology Group for access to and assistance on the ESRF synchrotron beamlines; P. Tropberger from R. Schneider's laboratory (Max Planck Institute of Immunology, Freiburg) for providing nucleosomes; and D. Panne and members of both laboratories for critical reading of the manuscript. E.H. is a Darwin trust fellow.

AUTHOR CONTRIBUTIONS

J.K., E.H., S.C. and A.A. designed the project and wrote the manuscript; J.K. and M.L. produced and crystallized the proteins; J.K. solved, refined and analyzed the structures, prepared and tested the mutants *in vitro*, and carried out HAT assays on the H4 peptide; J.S.-W. and J.K. performed dehydration of MSL1–MSL3 crystals; H.H. performed HAT assays on nucleosomes; and E.H. carried out all *in vivo* experiments.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Straub, T. & Becker, P.B. Dosage compensation: the beginning and end of generalization. *Nat. Rev. Genet.* **8**, 47–57 (2007).
2. Heard, E. & Disteche, C.M. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* **20**, 1848–1867 (2006).
3. Mendjan, S. & Akhtar, A. The right dose for every sex. *Chromosoma* **116**, 95–106 (2007).
4. Gelbart, M.E. & Kuroda, M.I. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* **136**, 1399–1410 (2009).
5. Smith, E.R. *et al.* A human protein complex homologous to the *Drosophila* MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol. Cell. Biol.* **25**, 9175–9188 (2005).
6. Marín, I. Evolution of chromatin-remodeling complexes: comparative genomics reveals the ancient origin of "novel" compensasome genes. *J. Mol. Evol.* **56**, 527–539 (2003).
7. Mendjan, S. *et al.* Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol. Cell* **21**, 811–823 (2006).
8. Taipale, M. *et al.* hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell. Biol.* **25**, 6798–6810 (2005).
9. Alekseyenko, A.A. *et al.* A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* **134**, 599–609 (2008).
10. Smith, E.R. *et al.* The *Drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol. Cell. Biol.* **20**, 312–318 (2000).
11. Akhtar, A. & Becker, P.B. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol. Cell* **5**, 367–375 (2000).
12. Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A. & Lucchesi, J.C. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J.* **16**, 2054–2060 (1997).
13. Scott, M.J., Pan, L.L., Cleland, S.B., Knox, A.L. & Heinrich, J. MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in *Drosophila*. *EMBO J.* **19**, 144–155 (2000).
14. Morales, V. *et al.* Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *EMBO J.* **23**, 2258–2268 (2004).
15. Morales, V., Regnard, C., Izzo, A., Vetter, I. & Becker, P.B. The MRG domain mediates the functional integration of MSL3 into the dosage compensation complex. *Mol. Cell. Biol.* **25**, 5947–5954 (2005).
16. Sural, T.H. *et al.* The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nat. Struct. Mol. Biol.* **15**, 1318–1325 (2008).
17. Akhtar, A., Zink, D. & Becker, P.B. Chromodomains are protein-RNA interaction modules. *Nature* **407**, 405–409 (2000).
18. Li, X., Wu, L., Corsa, C.A., Kunkel, S. & Dou, Y. Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. *Mol. Cell* **36**, 290–301 (2009).
19. Raja, S.J. *et al.* The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol. Cell* **38**, 827–841 (2010).
20. Yan, Y., Barlev, N.A., Haley, R.H., Berger, S.L. & Marmorstein, R. Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. *Mol. Cell* **6**, 1195–1205 (2000).
21. Rojas, J.R. *et al.* Structure of *Tetrahymena* GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* **401**, 93–98 (1999).
22. Wang, L., Tang, Y., Cole, P.A. & Marmorstein, R. Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function. *Curr. Opin. Struct. Biol.* **18**, 741–747 (2008).
23. Yan, Y., Harper, S., Speicher, D.W. & Marmorstein, R. The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. *Nat. Struct. Biol.* **9**, 862–869 (2002).
24. Berndsen, C.E., Albaugh, B.N., Tan, S. & Denu, J.M. Catalytic mechanism of a MYST family histone acetyltransferase. *Biochemistry* **46**, 623–629 (2007).
25. Burke, T.W., Cook, J.G., Asano, M. & Nevins, J.R. Replication factors MCM2 and ORC1 interact with the histone acetyltransferase HBO1. *J. Biol. Chem.* **276**, 15397–15408 (2001).
26. Zhang, P. *et al.* The MRG domain of human MRG15 uses a shallow hydrophobic pocket to interact with the N-terminal region of PAM14. *Protein Sci.* **15**, 2423–2434 (2006).
27. Bowman, B.R. *et al.* Multipurpose MRG domain involved in cell senescence and proliferation exhibits structural homology to a DNA-interacting domain. *Structure* **14**, 151–158 (2006).
28. Buscaino, A. *et al.* MOF-regulated acetylation of MSL-3 in the *Drosophila* dosage compensation complex. *Mol. Cell* **11**, 1265–1277 (2003).
29. Kelley, R.L. *et al.* Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* **98**, 513–522 (1999).

ONLINE METHODS

Expression, purification and crystallization. The MSL1–MOF and MSL1–MSL3 complexes were produced by coexpression in bacteria and purified as described in **Supplementary Methods**. The MSL1–MOF crystals grew at 20 °C in a solution containing 0.1 M sodium acetate (pH 5.0) and 1.0 M sodium formate and were cryoprotected with 30% (v/v) glycerol. The MSL1–MSL3 crystals were obtained at 15 mg ml^{-1} in a solution containing 0.1 M ADA (pH 6.5), 0.1 M Li_2SO_4 and 0.9 M $MgSO_4$ and initially diffracted to 8-Å resolution. After dehydration using the HC1b humidity control device and freezing in the presence of perfluoropolyether PFO-X125/03 (Lancaster Synthesis), the diffraction limit was extended to 2.8 Å[30].

Data collection and structure determination. Crystals of the $MOF_{174–458}$–$MSL1_{470–540}$ complex belong to the space group $I4_122$ with the unit cell dimensions a, b = 180.9 Å and c = 80.7 Å. The asymmetric unit contains one complex and has a solvent content of 70%. A complete native dataset was collected to a resolution of 2.8 Å on beamline ID14-EH4 at the ESRF (Grenoble, France). The data were processed using XDS[31]. Phases were obtained by molecular replacement using PHASER[32] with the deposited structure of the human MOF HAT domain (PDB code: 2GIV) as a search model. The initial map was improved using the prime-and-switch density modification option of RESOLVE[33]. After manual model rebuilding with COOT[34], the structure was refined using Refmac5 (with TLS refinement)[35] to a final R-factor of 22.3% and R_{free} of 25.6% (**Table 1**) with all residues in allowed (97.7% in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY[36]. A representative part of the $2F_o - F_c$ electron density map calculated using the refined model is shown in **Supplementary Figure 14a**.

Crystals of the $MSL3_{167–289,442–517}$–$MSL1_{545–597}$ complex belong to the space group $P2_1$ with four complexes per asymmetric unit. A complete native dataset was collected to a resolution of 3.0 Å on beamline ID14-EH2 at the ESRF. The data was processed using XDS[31]. The structure was solved by molecular replacement with PHASER[32] using the structure of the human MRG15 MRG domain (PDB code: 2F5J[26]) as a search model. Probably because of weak intensity of reflections above 3.5 Å, the obtained electron density appeared to be of a lower resolution than 3 Å. B-factor sharpening was therefore used to improve map quality[37]. The structure was built in COOT[34] and refined with Refmac5 (using TLS refinement)[35] to a final R-factor of 23.2% and R_{free} of 25.3% (**Table 1**) with 99.87% of residues in allowed (95.9% in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY[36]. A representative part of the –60 $Å^2$ B-factor–sharpened $2F_o - F_c$ electron density map calculated using the refined model is shown in **Supplementary Figure 14b–d**. The Wilson B-value for this dataset was determined to be 107 $Å^2$ using a maximum likelihood–based method as implemented in PHENIX[38]. Accordingly, the mean value for the isotropic individual B-factor for the final model is 100 $Å^2$. Interpretable electron density is observed for residues 551–558 and 564–594 of MSL1. In the crystal, the N- and C-terminal parts of MSL1 bind to two distinct MSL3 molecules, a result that we interpret as a result of domain swapping. It is unclear whether the MSL1 N terminus is swapped between two or four molecules (see **Supplementary Fig. 10a**). Buried surface areas of protein-protein interactions were calculated using the PISA web server at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)[39].

His-tag pulldown assays. The MSL1 and NSL1 and their mutated versions were coexpressed with MSL3 or MOF in bacteria, and the resultant complexes were purified using Ni^{2+} resin; see **Supplementary Methods**.

Thermal denaturation assay. Thermal denaturation (thermal shift) assays were performed as described in **Supplementary Methods**.

Expression of Msl1 and Nsl1 mutants in *Drosophila* SL-2 cells. *Drosophila* Msl1 was expressed as a C-terminal Flag fusion in the pAc5.1/V5-His A vector. Nsl1 was expressed as an N-terminal Flag fusion in a modified pBSactTAP vector. Transient transfection of SL-2 cells was done with Qiagen Effectene Transfection Reagent. Details are given in **Supplementary Methods**.

HAT assays. The activity of human MSL subcomplexes and MOF mutants in the active site was assessed within the MOF–MSL1–MSL3 complex in DTNB (5,5′-dithiobis(2-nitrobenzoic acid) HAT assays. The activity of human and *Drosophila* MSL subcomplexes was also assayed using native mono- and dinucleosomes obtained from MCF-7 cells. Further details can be found in **Supplementary Methods**.

***In vitro* reconstitution assay.** The reconstitution of *Drosophila* MSL subcomplexes produced in insect cells is described in **Supplementary Methods**.

Fractionation of SL-2 cells. SL-2 cell fractionation was based on swelling the cells in hypotonic buffer and vortexing in the presence of mild detergent. The supernatant is cytoplasmic fraction and the pellet is nuclei. The nucleoplasmic fraction was obtained by salt extraction, and the remaining chromatin was solubilized by benzonase treatment. See **Supplementary Methods** for details.

Generation of SL-2 stable cells lines. The cells were transfected with 0.5 µg of DNA with Qiagen Effectene Transfection Reagent, and selection was carried out with 1 mg ml^{-1} Geneticin for 2 weeks. Details are given in **Supplementary Methods**.

Immunofluorescence for SL-2 cells. SL-2 cells were swollen in 500 µl 0.5% sodium citrate for 7 min and loaded through a single-chamber Cytospin tunnel. The cells were spun for 10 min at 900 r.p.m. in a Cytospin (Thermo Shandon) and visualized with a 63× objective. See **Supplementary Methods** for details.

ChIP protocol from SL-2 cells. ChIP protocol was carried out as in ref. 19 with minor modifications. See **Supplementary Methods** for a detailed protocol.

30. Sanchez-Weatherby, J. *et al*. Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1237–1246 (2009).
31. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
32. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 458–464 (2005).
33. Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
35. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
36. Davis, I.W., Murray, L.W., Richardson, J.S. & Richardson, D.C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
37. Brunger, A.T., DeLaBarre, B., Davies, J.M. & Weis, W.I. X-ray structure determination at low resolution. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 128–133 (2009).
38. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
39. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

Molecular Cell
Article

Msl1-Mediated Dimerization of the Dosage Compensation Complex Is Essential for Male X-Chromosome Regulation in *Drosophila*

Erinc Hallacli,[1,4] Michael Lipp,[2,3,4] Plamen Georgiev,[1] Clare Spielman,[1] Stephen Cusack,[2,3] Asifa Akhtar,[1,*] and Jan Kadlec[2,3,*]
[1]Max-Planck-Institut für Immunbiologie und Epigenetik, Stübeweg 51, 79108 Freiburg im Breisgau, Germany
[2]European Molecular Biology Laboratory, Grenoble Outstation
[3]Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265
6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France
[4]These authors contributed equally to this work
*Correspondence: akhtar@ie-freiburg.mpg.de (A.A.), kadlec@embl.fr (J.K.)
http://dx.doi.org/10.1016/j.molcel.2012.09.014

SUMMARY

The Male-Specific Lethal (MSL) complex regulates dosage compensation of the male X chromosome in *Drosophila*. Here, we report the crystal structure of its MSL1/MSL2 core, where two MSL2 subunits bind to a dimer formed by two molecules of MSL1. Analysis of structure-based mutants revealed that MSL2 can only interact with the MSL1 dimer, but MSL1 dimerization is MSL2 independent. We show that Msl1 is a substrate for Msl2 E3 ubiquitin ligase activity. ChIP experiments revealed that Msl1 dimerization is essential for targeting and spreading of the MSL complex on X-linked genes; however, Msl1 binding to promoters of male and female cells is independent of the dimer status and other MSL proteins. Finally, we show that loss of Msl1 dimerization leads to male-specific lethality. We propose that Msl1-mediated dimerization of the entire MSL complex is required for Msl2 binding, X chromosome recognition, and spreading along the X chromosome.

INTRODUCTION

Heterogametic organisms with unequal numbers of sex chromosomes have to go through a process called dosage compensation to equilibrate their transcriptional output. Diverse solutions to the dosage problem evolved in different organisms. *Drosophila melanogaster* males transcriptionally upregulate their single X chromosome roughly two times to compensate for the absence of an active homolog (Conrad and Akhtar, 2011), whereas in mammals, females inactivate one of the two X chromosomes (Augui et al., 2011). Dosage compensation not only balances sex differences, but has also been shown to equalize X to autosome ratios in mammals, *C. elegans*, and *Drosophila* (Deng et al., 2011; Kharchenko et al., 2011). Dosage compensation mechanisms provide an excellent model for studying chromosome-wide transcription regulation through epigenetic mechanisms (Gelbart and Kuroda, 2009).

In *Drosophila*, the Male-Specific Lethal (MSL) complex, also known as the dosage-compensation complex (DCC), mediates dosage compensation (Hallacli and Akhtar, 2009). The complex consists of at least five MSL proteins (Msl1, Msl2, Msl3, Maleless [Mle], and Males-absent-on-the-first [Mof]) and two redundant long noncoding RNAs (roX1 and roX2) (Ilik and Akhtar, 2009). An equivalent, highly conserved complex also exists in human, composed of at least MSL1, MSL2, MSL3, and MOF. Whether any noncoding RNAs reside in the mammalian MSL complex remains unknown (Marín, 2003; Smith et al., 2005; Kadlec et al., 2011; Wu et al., 2011). MSL1, predicted to contain no globular domains, serves as a scaffold of the MSL complex (Kadlec et al., 2011). It interacts with its conserved C-terminal region, called the PEHE domain, with the histone acetyltransferases (HAT) domain of MOF and the MRG domain of MSL3 (Kadlec et al., 2011). Both *Drosophila* and human MSL1 were proposed to interact with MSL2 via an N-terminal predicted coiled-coil region (Li et al., 2005; Wu et al., 2011). MSL2 consists of an N-terminal RING finger and a C-terminal cysteine-rich domain involved in DNA binding (Fauth et al., 2010). The best-studied catalytic activity of the complex is the MOF-mediated acetylation of histone H4 at lysine 16 (H4K16ac) on the X chromosome (Hilfiker et al., 1997; Akhtar and Becker, 2000; Smith et al., 2000). We have recently shown that Mof's enzymatic activity is tightly regulated for H4K16 acetylation, promoting enhanced loading of RNAPII at the promoters of X-linked genes (Conrad et al., 2012a, 2012b). Human MSL2 has also been shown to be an E3 ubiquitin ligase for lysine 34 of histone H2B (Wu et al., 2011), suggesting that other histone marks may cross-talk on the male X chromosome.

According to the current model, the MSL complex is first enriched on numerous GA repeat-rich sequences called high-affinity sites (HAS), such as roX genes, followed by spreading to the rest of the X chromosome in a sequence-independent manner (Alekseyenko et al., 2008). Interestingly, HAS can be qualitatively differentiated with respect to their requirement of either Msl3 or Mof and their genomic location (Kadlec et al., 2011).

To gain mechanistic insights into how Msl1 and Msl2 influence dosage compensation, we determined the crystal structure of

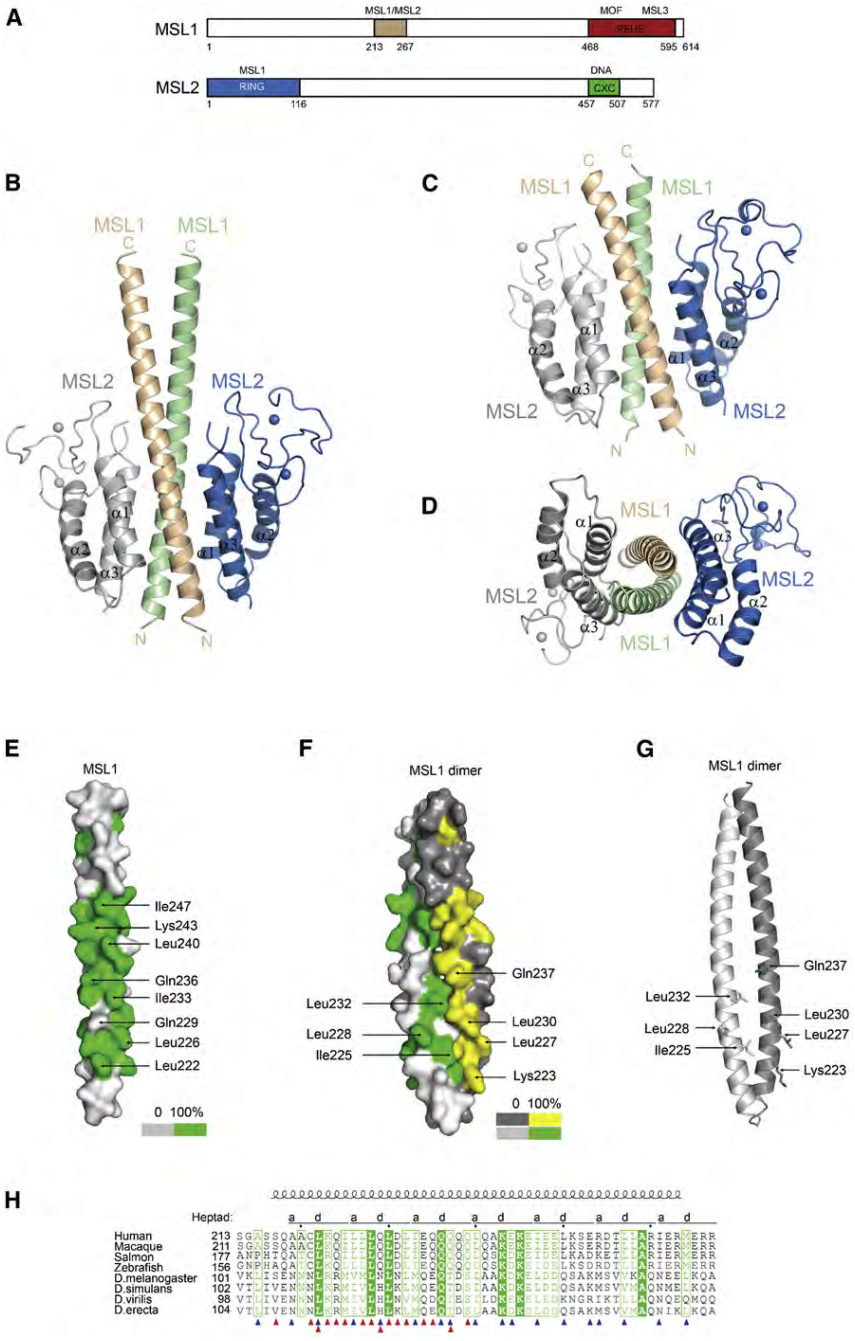


Figure 1. Crystal Structures of the MSL1/MSL2 Complex

(A) Schematic representation of the domain structures of human MSL1 and MSL2. The binding partners are indicated above individual domains.

(B) Ribbon diagram of the human $MSL1_{213-267}$/$MSL2_{1-116}$ complex. Two molecules of MSL1 form the central dimeric coiled coil (shown in brown and green). The N-terminal RING finger-containing domains of MSL2 are shown in blue and gray.

(C) Ribbon representation of the complex between shorter $MSL1_{213-252}$ and $MSL2_{1-116}$ in the same orientation as in (A).

(D) The $MSL1_{213-252}$/$MSL2_{1-116}$ structure rotated by 90° along the horizontal axis relative to (C).

(E) Surface representation of the MSL1 helix. Conserved surface residues, based on the sequence alignment in (H), are shown in green, indicating 100% conservation. Only conserved residues involving MSL1 dimerization are labeled.

(F) Surface representation of the MSL1 dimer forming composite binding sites for MSL2. Highly conserved residues are shown in green and yellow and residues binding MSL2 are labeled. Gln237 is also highlighted; as in *Drosophila*, its substitution with threonine is compensated by T7Q mutation in Msl2.

(G) Ribbon representation of the MSL1 dimer in the same orientation as in (F), showing the conserved MSL2 binding residues.

(H) Sequence alignment of MSL1 proteins comparing vertebrates and *Drosophila* species. Only the sequence of the coiled-coil region is shown. Identical residues are in green boxes, and conserved residues are shown in green. Blue triangles indicate residues involved in the MSL1 dimerization while red triangles show residues interacting with MSL2.

the complex between their highly conserved human orthologs, which unexpectedly show that MSL1 and MSL2 form a heterotetrameric core of the MSL complex. Functional experiments with structure-based *Drosophila* Msl1 mutants revealed that Msl1 dimerization is required for Msl2 and roX2 RNA binding, X chromosome recognition, and spreading along the X chromosome. Furthermore, we established a dimerization-, Msl2-, Msl3-, and Mof-independent binding of Msl1 at autosomal and X-linked promoters of male and female cells. Finally, we show that the Msl1 dimerization plays a vital role in vivo, as specific point mutations lead to male-specific lethality in flies.

RESULTS

MSL1 and MSL2 Form a Heterotetrameric Core of the MSL Complex

The complex between the predicted coiled-coil region of human MSL1 (residues 213–326) and the N-terminal portion of MSL2 (residues 1–116) was formed by coexpression in bacteria. Using trypsin-limited proteolysis, we identified a shorter MSL1 fragment spanning residues 213–267 that was sufficient for the MSL2 binding. The structure of this complex was determined by X-ray crystallography at a resolution of 3.5 Å (Figure 1). In order to improve the crystal quality, the MSL1 fragment was further shortened to residues 213–252, and a structure of its complex with $MSL2_{1-116}$ was solved at 3.25 Å resolution and refined to an R factor of 24.4% and R_{free} of 26% (Figures 1C and 1D). Msl1 and Msl2 proteins were

originally suggested to dimerize via their putative coiled-coil regions (Scott et al., 2000; Li et al., 2005). However, unexpectedly, both of our structures show that instead these two proteins form a heterotetrameric core of the MSL complex, where two MSL1 subunits form a dimeric coiled coil that serves as a binding platform for two molecules of MSL2 (Figure 1). Additionally, we confirmed by multiangle laser light scattering (MALLS) that the MSL1/MSL2 subcomplex has an apparent 2:2 stoichiometry also in solution (Figures S1A and S1B).

The MSL1-MSL1 Interface

The crystallized fragment of MSL1 (residues 213–267) forms a 75 Å long parallel dimeric coiled coil, where ten hydrophobic and four polar residues (Gln229, Gln236, Lys243, and Arg254), originally thought to be involved in the interaction with MSL2 (Li et al., 2005), pack in layers with a regular heptad (3–4) periodicity (Figures 1E–1H, S1C, and S1D). Additional stabilizing interactions between the helices are shown in Figure S1D. The dimer's two Glu229 and Glu236 residues form interhelical hydrogen bonds at its core (Figure S1D). The predicted heptad repeats in human MSL1 extend until residue 285, suggesting the MSL1-MSL1 coiled coil might be even longer. However, mass spectrometry analysis of a tryptic digest of the $MSL1_{213-310}$-containing complex revealed four sites accessible to trypsin at positions 266, 267, 272, and 273, indicating that this possible extension is less stable than the crystallized segment (Figures S1E and S1F). The heptad pattern between residues 215 and 233 was difficult to predict, as this region also contains a cluster of hydrophobic residues involved in the interaction with the helices of MSL2 (Figure 1H). A corresponding coiled coil in *Drosophila* Msl1 is located closer to the N terminus than in the human protein and is predicted to be longer (residues 100–176). We showed that $Msl1_{85-186}$ forms a stable complex with $Msl2_{1-192}$ in vitro (Figure S2A). Most of the residues involved in the MSL1 dimerization are highly conserved across species, reflecting the importance of this interaction for the functional integrity of the MSL complex (Figures 1E and 1H). Upon dimerization, the MSL1 coiled coil forms two composite, mostly hydrophobic binding sites for two molecules of MSL2, which are clearly identifiable by mapping of conserved residues onto the surface of the dimer (Figures 1F and 1G).

The Structure of MSL2

The structure of the N-terminal region of MSL2 consists of three long helices (α1, α2, and α3) forming a triple-stranded antiparallel coiled coil and a RING finger that is inserted between helix α2 and α3 (Figure 2A). $MSL2_{1-116}$ exhibits a sequence similarity to the N-terminal domain of BRCA1 ubiquitin ligase, mostly within the helix α2 and the RING finger. Superposition of the BRCA1 structure (PDB code 1JM7) onto MSL2 revealed that these two domains are indeed similar (Figures 2A and 2B; RMS deviation 2.1 Å for 77 Cα atoms). In the BRCA1 structure, the central RING finger is flanked only by two helices, corresponding to α2 and α3 of MSL2, which form a four-helix bundle with corresponding helices of BARD1 (Brzovic et al., 2001). Interestingly, in MSL2, the packing of helix α1 against helices α2 and α3 resembles the C-terminal helix of BARD1 binding to BRCA1 (Figures 2C and 2D). The MSL1 binding surface is then formed by helices α1 and α3.

The RING finger of MSL2 includes residues 42–93 and adopts a fold observed in other RING domain structures. The two Zn atoms are coordinated by absolutely conserved Cys44, Cys47, Cys67, and Cys70 and Cys62, His64, Cys81, and Cys84, respectively (Figure S2B). The N-terminal portion of the RING finger, which is in proximity of the helices of MSL2, is stabilized by interactions between conserved Leu50, which points toward a hydrophobic cavity between helices α2 and α3, and Gln63, which forms several hydrogen bonds with Val4 and Asn5 of α1. A solvent-exposed loop binding the second Zn atom (71–90) is the most flexible and poorly defined in both structures.

Human MSL2 was reported to be an E3 ligase for p53 and its RING finger to be indispensable for this activity (Kruse and Gu, 2009). Recently, MSL2 has been shown to ubiquitinate histone H2B, and its activity was greatly reduced by H64Y mutation within the RING finger (Wu et al., 2011). We thus analyzed the MSL2 RING finger structure with respect to its role in ubiquitination. The structure of c-Cbl in a complex with UbcH7 serves as a model for interactions between RING domains with E2 enzymes (Zheng et al., 2000). Ile383, Trp408, Pro417, and Phe418 of c-Cbl and equivalent residues in other RING proteins make a largely hydrophobic interacting surface contacting two loops of E2 (Figure 2E). Surprisingly, a corresponding interaction surface is not formed in the MSL2 structure. MSL2 lacks the central helix (downstream of Cys70) that is characteristic for most RING finger domains and that normally forms the E2 binding groove (Figures 2E–2G). Instead, a loop (residues 71–77) binds across the putative E2 binding surface, with Val46, Met75, and Met77 being buried in the interface (Figure 2F). This unusual conformation might be a result of the crystallization process, which would be in agreement with the high flexibility of this region seen in this and other RING structures. However, a similar positioning of this loop occurs also in the structure of promyelocytic leukemia proto-oncoprotein PML that equally lacks the central helix (buried Met38; Figure 2G) (Borden et al., 1995). Alternatively, this structure might represent an autoinhibited state of the RING finger, as recently characterized for TRAF2. This protein has an insertion in the loop that would apparently inhibit E2 binding; however, upon binding of a cofactor, TRAF2 is nevertheless active (Alvarez et al., 2010). It seems unlikely that the MSL2 RING finger could interact with an E2 in a way similar to other E3 enzymes without a local conformational change of the 71–77 loop that would make accessible the putative E2 binding surface (Figure S2C).

We tested the ubiquitination activity of *Drosophila* Msl2 using purified full-length protein expressed in insect cells in an in vitro ubiquitination assay (Figure 2H). In agreement with the literature, we could show that Msl2 can autoubiquitinate itself, which is a hallmark of E3 ligase proteins (Figure 2I). Msl1 did not exhibit any ubiquitination activity (data not shown). Interestingly, in the presence of the Msl2/Msl1 complex, both Msl1 and Msl2 were ubiquitinated, as the amount of unmodified proteins was rapidly decreasing in time (Figure 2J). This experiment indicates that Msl1 is a substrate of Msl2 in vitro and that the Msl1/Msl2 tetramer has significantly higher activity compared to Msl2 alone (Figures 2I and 2J). The increased activity of Msl1/Msl2 tetramer was also observed on Rpn10, a universal substrate for E3 ligases (Uchiki et al., 2009) (Figure S2F). Similarly, human MSL1/MSL2

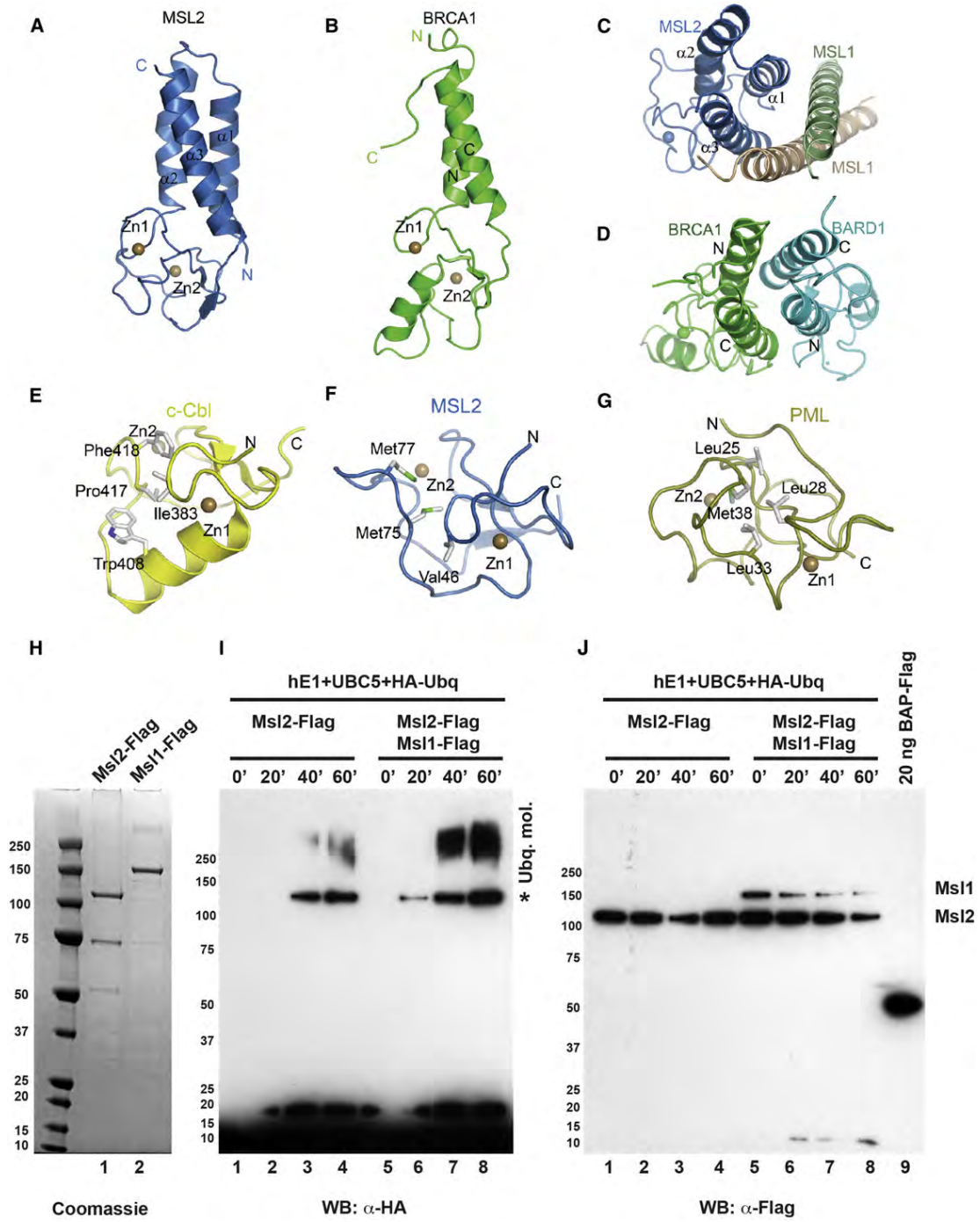


Figure 2. Structure of the N-Terminal Domain of MSL2 and Its Ubiquitination Activity

(A) Ribbon representation of the MSL2$_{1-116}$ structure. A RING finger coordinating two zinc atoms is inserted between helices α2 and α3.

(B) The MSL2 RING finger-containing domain is similar to the N terminus of BRCA1 (PDB code 1JM7).

(C and D) Comparison of the MSL1/MSL2 and BRCA1/BARD1 complexes. MSL2 interacts with MSL1 with α1 and α3 (C). Helix α1 of MSL2 packs against helices α2 and α3 in position equivalent to the one of the C-terminal helix of BARD1 interacting with BRCA1 (D).

(E) RING finger of c-Cbl (PDB code 1FBV). Ile383, Trp408, Pro417, and Phe418 form a hydrophobic groove involved in the interaction with UbcH7 E2 enzyme.

(F) RING finger of MSL2 (residues 42–93). Loop 71–77 binds across the putative E2 binding site.

(G) RING finger of promyelocytic leukemia proto-oncoprotein PML (PDB code 1BOR). The E2 binding site is obstructed in a way similar to MSL2.

(H) SDS-PAGE gel showing the purified *Drosophila* Msl1-Flag and Msl2-Flag expressed in insect cells. In lane 1, the 75 kDa band is a common contaminant of Msl2 purifications and 50 kDa band is a degradation product of Msl2.

has higher ubiquitination activity than MSL2 alone (Wu et al., 2011). We propose that this increased activity is achieved via Msl1-mediated dimerization of Msl2, which is in line with often-observed activity boost of E3 ubiquitin ligase dimers compared to monomers (Budhidarmo et al., 2012).

To understand the importance of the E3 ligase activity of Msl2, we prepared several mutations aimed to disrupt its interaction with E2 enzymes. We observed that triple mutation of the *Drosophila* counterparts of Val46, Met75, and Met77 of the loop occluding the putative E2 binding surface (Val43, Lys72, and Met74), but not single mutants, significantly reduced Msl2 E3 activity (Figures S2G and S2H). These mutations, however, also affected the overall Msl2 structure, as this mutant no longer interacted with Msl1 (Figure S2I).

The MSL1/MSL2 Interface

Msl2 was suggested to interact with Msl1 via its RING finger (Copps et al., 1998). In contrast, our structure shows that its interaction with the MSL1 dimer is exclusively mediated by helices α1 and α3, while the RING finger has no contact with MSL1. Interestingly, the putative role of the *Drosophila* Msl2 RING finger in the interaction with Msl1 was established by identification of 13 mutations, which in light of the present structure would nearly all destabilize the RING finger and thus probably also the entire Msl2 (Copps et al., 1998). Only two of these mutations (M14K and C107R) would probably directly affect the binding of helix α1 and α3 to Msl1. The helices of the two MSL2 molecules bind to MSL1 in an antiparallel fashion, forming an eight-helical bundle (Figure 3A) with multiple contacts within several hydrophobic and polar layers along the first three heptad repeats of MSL1. The complex interface buries 1180 Å^2 of the MSL1 dimer. The key interacting residues of MSL1 form a short, highly conserved cluster between Ser117 and Gln239 (Figures 1F–1H). The interacting residues of MSL2 are highlighted in Figure S2B and include conserved Tyr10, Arg15, Gln95, Cys102, and Tyr109. At the N-terminal end of MSL1, Ser117 forms a hydrogen bond with Gln112 of MSL2. Above, a mixed hydrophobic-polar layer is formed around MSL1 Leu222, where Cys221, Lys223, and Gln224 make several hydrogen bonds with Lys105, Glu108, and Tyr109 of α3 of MSL2 (Figure 3C). The central polar layer formed around MSL1 Gln229 and MSL2 Tyr10 (Figure 3D) is isolated on both sides from solvent by numerous hydrophobic residues of MSL1 and MSL2. Finally, the glutamine cluster at heptad 3, including Gln236 and Gln237, forms a network of hydrogen bonds with MSL2 (Figure 3B). The helical bundle is stabilized also on the exterior by salt bridges between Arg15 of MSL2 and Asp231 and Glu234 of MSL1.

In *Drosophila* Msl1 the important Gln237 is replaced by a threonine. Interestingly, Msl2 contains a compensatory threonine to glutamine mutation that might preserve the hydrogen bonding. The interaction network in the *Drosophila* complex is also likely to be maintained by L99Q mutation that compensates for Gln95 substitution for a methionine (Figures S2D and S2E). Thus, even though the level of conservation of the MSL1 coiled-coil region between human and *Drosophila* appears to be lower than in the case of the MSL3 and MOF binding regions (Kadlec et al., 2011), we believe the structure of the MSL1/MSL2 heterotetramer is very similar between these two species.

MSL1 Dimerization and MSL2 Binding Can Be Separated

Since the role of the MSL complex is better understood in *Drosophila*, and the key residues in all interaction interfaces MSL1 makes with MSL1, MSL2, MSL3, and MOF are evolutionarily conserved, we performed all our functional studies with *Drosophila* proteins in cell lines as well as transgenic flies. All the *Drosophila* Msl1 mutants used in this study and the corresponding mutations to mammalian counterparts are summarized in Figure 3E, and they all have a C-terminal 3xFlag epitope unless indicated otherwise.

In our previous study, we showed that the Msl3 and Mof interactions with Msl1 can be disrupted without any apparent influence on the other protein-protein interactions within the complex (Kadlec et al., 2011). To further support this finding and functionally separate the N-terminal interactions of Msl1 with Msl2 and the C-terminal interactions with Mof and Msl3 (through the PEHE region), we generated an Msl1 mutant (Msl1 mut.1) that binds neither Msl3 nor Mof (Figure 3F, lane 2, see also Figure S3A). Using coimmunoprecipitation with transiently expressed Msl1, we could show that the Msl1 interaction with Msl2 remains unaffected even when both Msl3 and Mof are eliminated from the complex (Figure 3F).

To test the dimerization of the full-length Msl1 in vivo, we transiently coexpressed the wild-type (WT) Msl1-Flag and Msl1-myc proteins and immunoprecipitated Msl1-Flag proteins using a Flag antibody-coupled resin. Indeed, the Flag-tagged Msl1 coimmunoprecipitated with Msl1-myc as well as Msl2, Msl3, and Mof (Figure 3F lane 1). Furthermore, we observed that Msl1 can dimerize even in the absence of Msl3 and Mof (Figure 3F lane 2). Next, we were interested in identifying Msl1 mutations that would disrupt its dimerization, without directly affecting the residues interacting with Msl2. Thus, we mutated either four or five residues at *a* or *d* heptad positions along the coiled coil to aspartates (Msl1 mut.2 and mut.3; Figures 1E, 1H, and 3E). Both mutants, although they were highly expressed, failed to copurify Msl1-myc and Msl2, while the interaction with Msl3 and Mof was unaffected (Figure 3F, lane 3 and 4). This experiment confirms that the interaction with Msl2 requires the entire composite Msl2 binding site formed by the Msl1 dimer (Figure 3A). We cannot exclude that the presence of two aspartate side chains (V114D, M121D) might affect the Msl1/Msl2 interface directly. It is important to note that neither the Msl1 dimerization nor Msl2 binding is required for the interaction with Msl3 and Mof. These results emphasize the modular nature of Msl1 interactions with different members of the MSL complex.

(I) In vitro ubiquitination assay with recombinant Flag-tagged Msl1 and Msl2. Equal amounts of proteins were assayed in 20 min time interval. HA antibody was used to determine the autoubiquitination of Msl2 (lanes 1–4) and ubiquitinated species of Msl1 and Msl2 (lanes 5–8). 0 time point indicates no ATP control. Asterisk indicates E1 enzyme ubiquitination.
(J) The same experiment as in (I) blotted for Flag antibody to monitor Msl1 and Msl2 species. Twenty nanograms BAP-Flag protein is used as an indicator of the amounts of Msl1 and Msl2 in this assay.

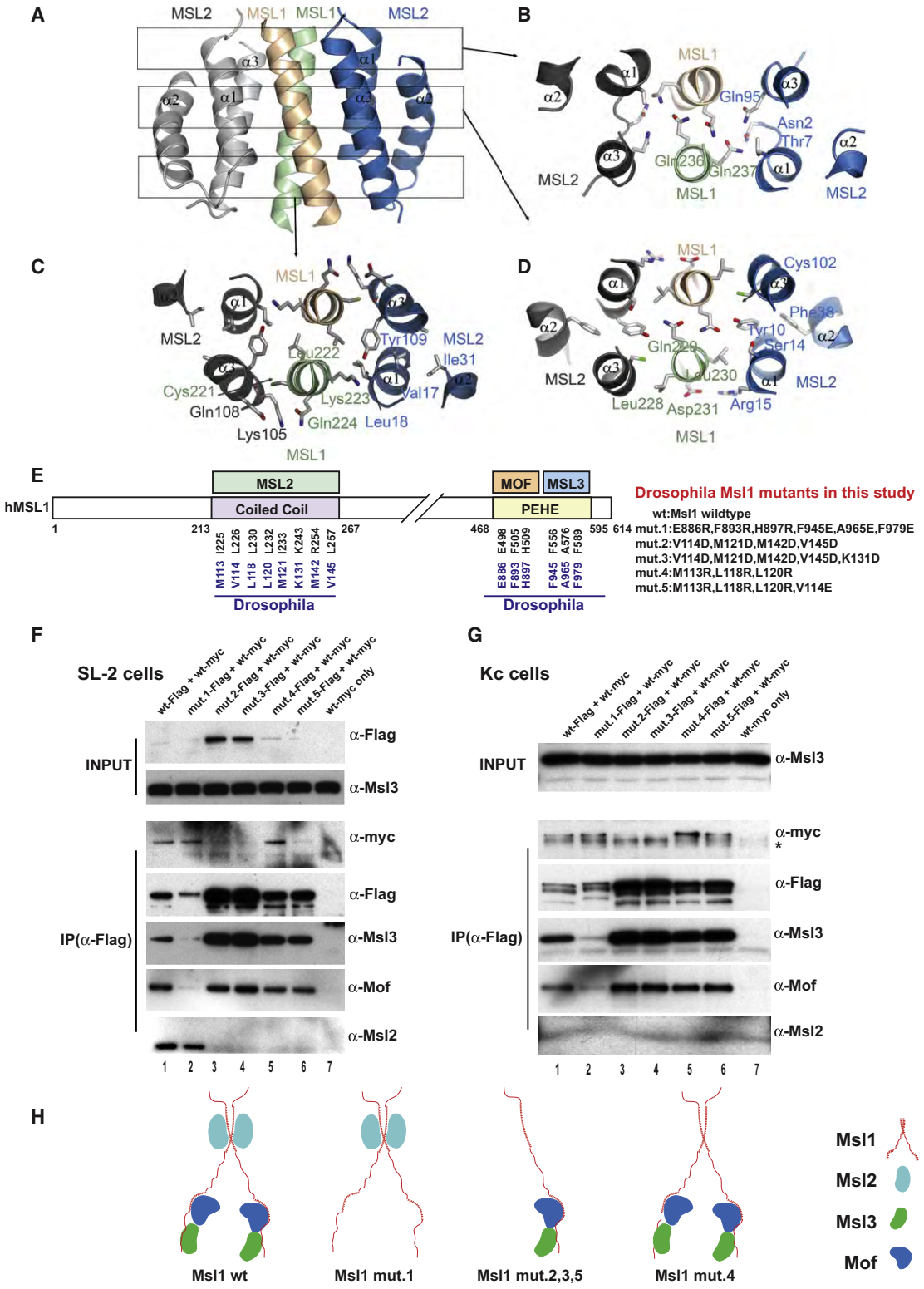


Figure 3. Details of the MSL1/MSL2 Interface and Analysis of Msl1 Derivatives in Cell Lines

(A) The helices of MSL1 and MSL2 form an antiparallel eight-helical bundle with many hydrophobic and polar interactions between the four molecules. The three details of the interface shown in (B)–(D) are localized on the structure by the black boxes.

(B) Highly conserved glutamine residues of the third heptad repeat of MSL1 (Gln236, Gln237) form several hydrogen bonds with Asn2, Thr7, and Gln95 of MSL2.

Next we designed a mutant that would not interact with Msl2 but would preserve the integrity of the Msl1 dimer. Thus, we mutated three residues in the Msl1/Msl2 interface that do not lie at *a* or *d* heptad positions to arginines (M113R, L118R, L120R: Msl1 mut.4; Figures 1F and 3E). The Msl1 mut.4 was still able to dimerize with Msl1-myc and bind Msl3 and Mof, while the interaction with Msl2 was lost (Figure 3F, lane 5), indicating that the presence of Msl2 is not required for the Msl1 dimerization. Finally, we showed that a single additional mutation in *a* heptad position (V114E) was sufficient to disrupt directly both Msl1 dimerization and Msl2 binding (Figure 3F, lane 6). Similar results were obtained when we repeated the coIP experiments with an HA-tagged WT Msl1 (Figure S3B). The hypothesis that Msl1 dimer can exist without Msl2 was further supported by coIP experiments in Kc cells, a cell culture model for *Drosophila* female cells, where Msl2 translation is inhibited (Figure 3G). The WT Msl1 dimerizes also in these cells (Figure 3G, lane 1), and similar effects as in SL-2 cells were also observed with mutants expressed in Kc cells. Interestingly, Msl1 mutants that lose the Msl2 interaction were consistently observed to be more abundant than the WT and mut.1, indicating a possible effect of Msl2 on Msl1 turnover, consistent with the in vitro ubiquitination experiment (Figures 2I and 2J). The schematic summary of the mutant Msl1-containing complexes is represented in Figure 3H. Taken together, these results conclusively show that Msl1 dimeric coiled coil is a platform for Msl2 interaction in vivo.

Msl1 Dimer Platform and Its Association with Msl2 Is Required for X Chromosome Recognition

In order to understand the importance of the Msl1 dimerization in X chromosome targeting, we tested our mutants for their interactions with chromatin by ChIP analysis in stable SL-2 cell lines. Similar expression level for each mutant was ensured by an inducible promoter (Figure S4). We used the Flag epitope for IP to selectively pull down mutant derivatives, avoiding endogenous Msl1. First we analyzed Msl1 binding to two HAS targets (roX2 and su(wa)) and several low-affinity sites within four X-linked genes (Figure 4A). Msl1 mut.1 ChIP shows that roX2 HAS binding is independent of both Msl3 and Mof, and su(wa) showed a reduced binding of the partial complex, whereas spreading across the body of the X-linked genes was completely lost. This result further supports our previous hypothesis that not all HAS are identical and show differential affinities toward various surfaces of the MSL complex (Kadlec et al., 2011). Strikingly, all the other mutants (Msl1 mut.3, mut.4, and mut.5) did not bind either to HAS or low-affinity site gene bodies. Exceptions were observed for the promoter regions of the same genes where binding remains unaffected (see below). In order to ensure that X chromosome recognition is lost starting from the HAS, we tested 12 more HAS targets determined by Kuroda and colleagues (Alekseyenko et al., 2008) (Figure 4B). Remarkably, all the tested targets show reduced binding of Msl1 mut.1 and completely abolished binding of the Msl1 mut.3 and mut.5. The loss of binding of Msl1 mut.4 importantly shows that Msl1 dimer per se cannot target the X chromosome, but requires the composite action with Msl2. Taken together, these results clearly indicate that Msl1 dimerization-mediated Msl2 binding is necessary for the recognition of X chromosomal genes.

roX2 RNA Integration Requires the Full Complex

The MSL complex contains two functionally redundant long noncoding RNAs, roX2 and/or roX1, implicated in spreading (Franke and Baker, 1999; Meller and Rattner, 2002). However, the actual mode by which the MSL complex binds RNA remains unknown (Lee et al., 1997; Akhtar et al., 2000; Li et al., 2005; Morales et al., 2005; Fauth et al., 2010). We used Msl1 mutants to study roX integration into the complex in vivo by RNA immunoprecipitation (RIP) method, where fixed complexes are pulled down and RNAs are quantitatively measured by quantitative PCR (Selth et al., 2009). First, we optimized the RIP protocol in SL-2 cells, where roX2 but not roX1 is expressed, using the Mle subunit as a bait protein, as its interaction with roX2 is well established (Lee et al., 1997). Mle bound roX2 in vivo, as verified by two different primer pairs and did not bind a nonspecific nuclear RNA, 7SK (Figure 4C). RIP by Flag antibody gave only background levels of signal from WT SL-2 cells (Figure 4C), ensuring the specificity of signals obtained from RIP of Msl1 mutants (Figure 4D). Rox2 binding to the exogenous WT Flag-tagged Msl1 was recapitulated from the stable cell line. While Msl1 mut.1 showed a significant reduction of RNA recovery, mut.3 and mut.4 completely lost the binding despite their equivalent expression levels (Figure S4). These results indicate that Msl1-Msl3-Mof trimeric complex (Msl1 mut.3) and hexameric complex (Msl1 mut.4) cannot bind roX2 in the absence of Msl2. The Msl3 or Mof proteins are also required for complete incorporation of the RNA, but their contribution is not detectable by this method when Msl2 is not present in the complex. Msl2 thus appears to be a key subunit for stable roX2 integration into the MSL complex.

Msl1 Binds to Promoters in Male and Female Cells

Reproducible Msl1 binding to the promoters of X-chromosomal genes and its independent nature from Msl3, Mof, Msl2, and

(C) Details of the MSL1/MSL2 interactions around MSL1 Leu222. Mostly hydrophobic layer formed by MSL1 Leu222 and Ly223 and MSL2 Val17, Leu18, Tyr109, and Ile31. This layer is part of a large hydrophobic core including also MSL1 Ile225 and Leu226 and MSL2 Leu35, Leu106, and Ile110 (not shown). MSL1 Cys221 and Gln224 form several hydrogen bonds with Lys105 and Glu108 of MSL2.
(D) Glutamines 229 forming an interhelical hydrogen bond within the MSL1 dimer also interact with Tyr10 of MSL2. Leu228 and Leu230 make hydrophobic contact with MSL2. Additionally, MSL2 Arg15 forms salt bridge interactions with Asp231 and Asp233 of MSL1.
(E) Mutated residues in *Drosophila* and their human homologs are represented on the human MSL protein scheme. *Drosophila* Msl1 mutants used in this study. All Msl1 mutants, including WT, have a C-terminal 3xFlag tag.
(F) Flag immunoprecipitation of Msl1 mutants in SL-2 cells. Msl1-Flag mutants are cotransfected with WT myc-tagged Msl1, and Flag beads were used for IP. Western blots were performed with the indicated antibodies. Flag and myc tag indicates C-terminal 3xFlag and 3xMyc tag, respectively.
(G) Same experiment in (F) performed in Kc cells. Asterisk in anti-myc blot indicates a contamination band. Msl2 absence is a marker for Kc cells.
(H) Schematic summary of WT or Msl1 mutants containing complexes derived from IP studies in (F) and (G).

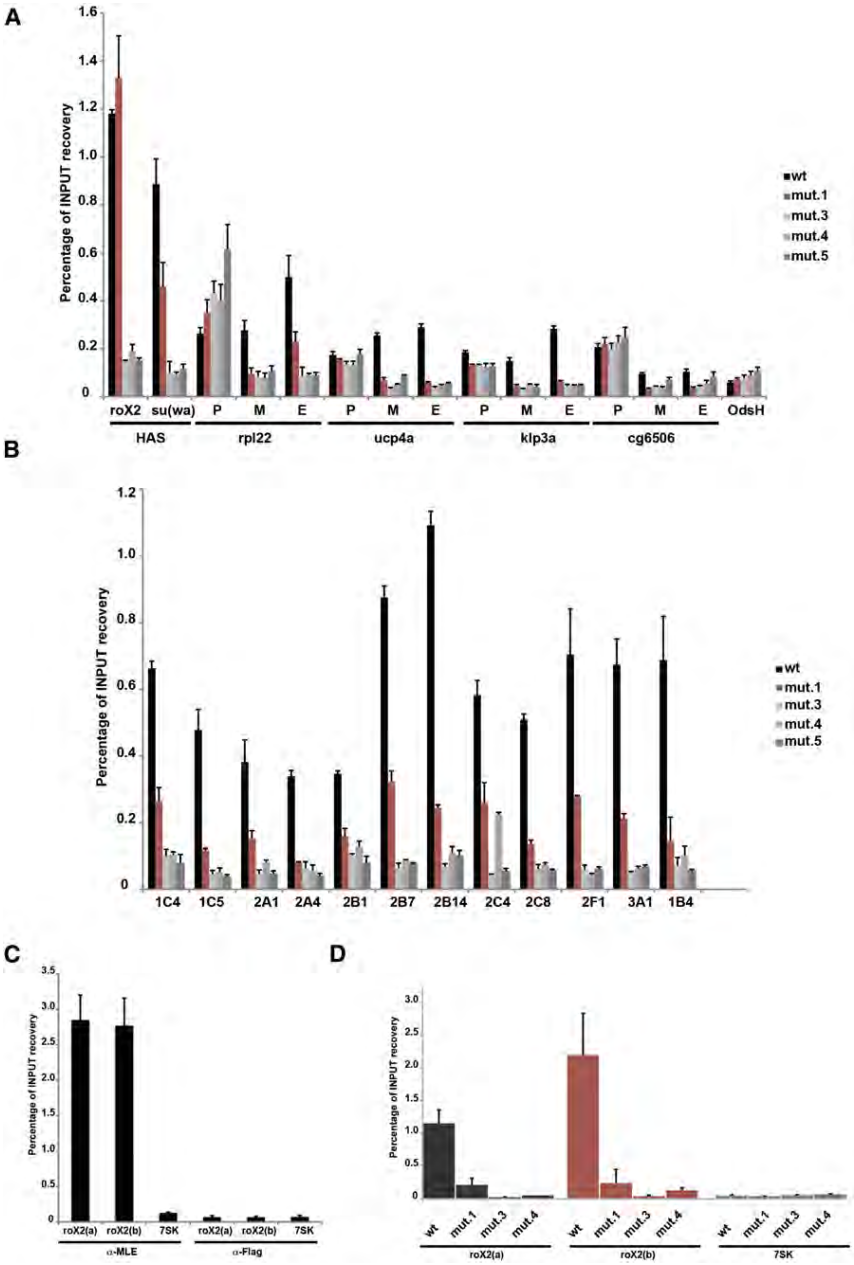


Figure 4. Chromatin and roX2 RNA Interactions of Msl1 Mutants

(A) ChIP of Msl1 mutants with Flag antibody in SL-2 stable cell lines. Two HAS and four X-linked genes were chosen as X chromosomal targets. OdsH target is used as a negative control. P, M, and E indicate promoter, middle, and end of the genes, respectively. The error bars represent the standard deviation of three independent experiments.

(B) Same experiment as in (A) is performed on 12 selected HAS.

(C) RNA immunoprecipitation (RIP) in SL-2 cells with Mle and Flag antibody. RIP on Mle protein is used as a positive control for roX2 RNA binding. Two different roX2 sites are quantitatively amplified (roX2 a, roX2 b). 7SK is used as a nuclear RNA negative control. RIP with Flag antibody is repeated on same targets in WT SL-2 cells to show background levels of RNA recovery. The error bars represent the standard deviation of three independent experiments.

(D) Flag RIP experiment in SL-2 stable cell lines that express Msl1 mutants. Two roX2 RNA target sites and a negative control RNA target (7SK) are amplified. The error bars represent the standard deviation of three independent experiments.

dimerization prompted us to hypothesize that this binding could be independent from its role in dosage compensation. In such a scenario, Msl1 might also be detectable at the promoters of autosomal genes, where dosage compensation does not occur. Indeed, by ChIP we detected significant enrichments at the promoters of eight random autosomal targets, while ORF binding was at the background level (Figure 5A, black bars). We also tested Msl1 chromatin interactions in female Kc cells and observed clear enrichments on the promoters of both X-linked and autosomal targets (Figure 5B). Since Mof is also present on autosomal promoters (Kind et al., 2008), we checked Msl3 systematically on the same autosomal and X-linked genes in males and females. Surprisingly, Msl3 was absent on autosomal promoters in male cells (Figure 5C) and on female promoters (data not shown) compared to X chromosomal targets. Msl3 occupancy at X-linked promoters was either absent or very low relative to Msl1. Taken together, these results suggest that Msl1 binding at the promoters is independent from its role in dosage compensation. The presence of Msl3 only on X chromosome ORFs provides a distinguishing feature for the X chromosomal genes versus autosomal targets with concomitant spreading of the MSL complex.

Msl1 Dimerization Is Essential for Male Viability

In order to assess the functional relevance of these *msl1* mutations in *Drosophila* in vivo, we generated transgenic flies expressing the mutant variants of *msl1* (WT Flag-tagged, mut.1, mut.3, mut.4, mut.5) in a spatiotemporally regulated manner using the UAS/Gal4 binary system. All transgenes were inserted in the same genomic location (65B2) by phiC31 integrase-mediated transformation to avoid the influence of position effects on gene expression and facilitate direct comparison upon phenotypic analysis (Groth, 2004). The fly system also enabled us to directly compare sex-specific effects of different mutations.

We first induced ectopic expression of these mutants in a WT background ubiquitously with a strong *tubulin-Gal4* driver at 25°C (Lee and Luo, 1999). Strikingly, expression of Msl1 mut.3,

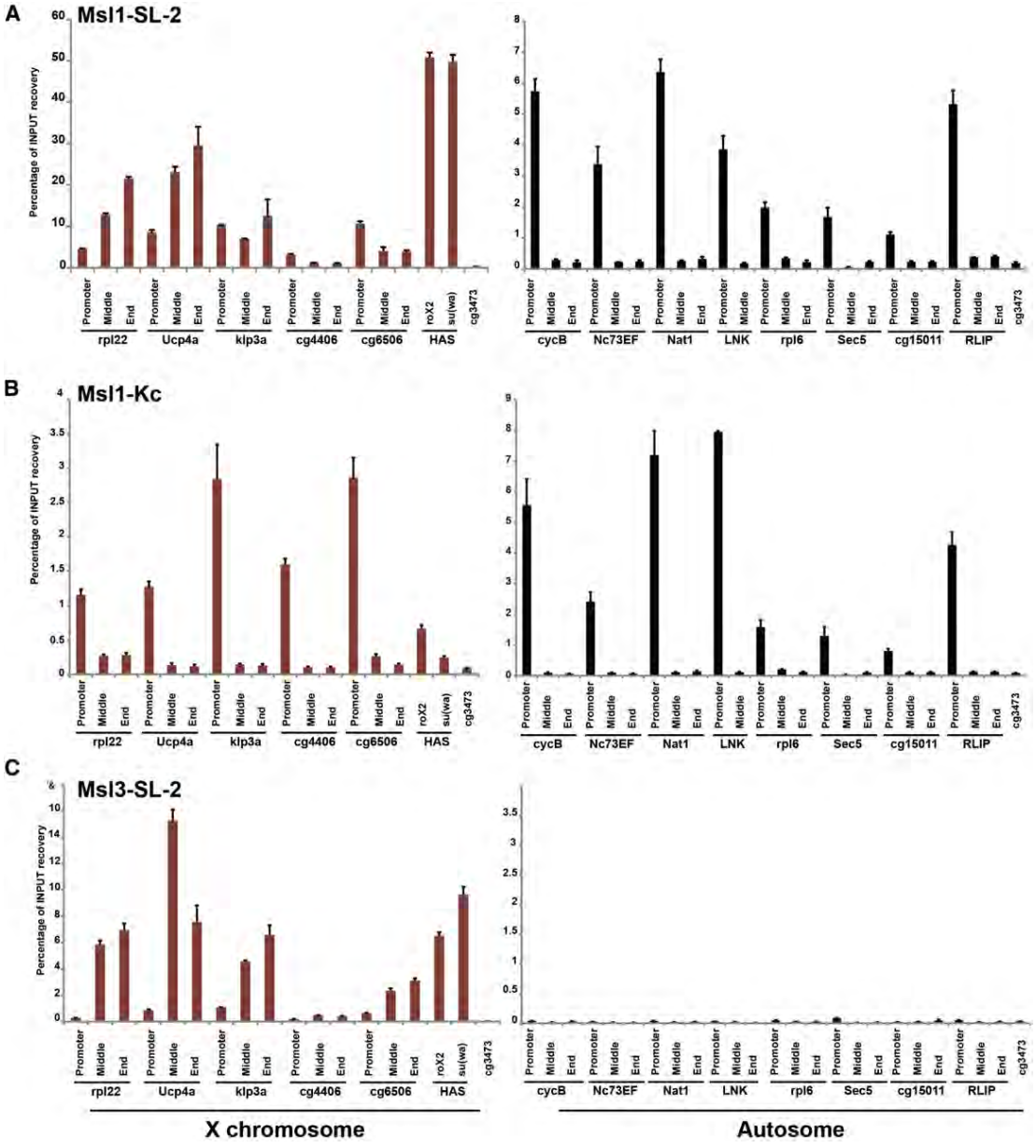


Figure 5. Msl1 Binds to Promoters of X and Autosomes in a Sex-Independent Manner

(A) ChIP of endogenous Msl1 in SL-2 cells. Five X-linked genes and two HAS are chosen for canonical X chromosome enrichment (red bars). Cg3473 is a negative control target site. Eight autosomal target sites are shown with black bars. The error bars represent the standard deviation of three independent experiments. Same experiment is performed for endogenous Msl1 in Kc cells (B) and endogenous Msl3 in SL-2 cells (C).

mut.4, and mut.5 caused both male and female lethality, whereas Msl1 mut.1 caused only male-specific lethality and WT Msl1 expression did not have any observable effects on viability (Figure 6A). Western blot analysis showed that Msl1 mut.3, mut.4, and mut.5, which lose Msl2 interaction, are more abundant than WT Msl1 and mut.1 (Figure 6B), as observed in our cell culture system, suggesting a possible downregulating effect of Msl2 on Msl1.

In order to ensure that lethality is not due to indirect effects of overexpression of the mutant proteins, especially for the coiled-coil Msl1 mut.3, mut.4, and mut.5, we repeated the experiment at 18°C, where *tubulin-Gal4*-induced transgene expression can be significantly decreased relative to 25°C (Mondal et al., 2007) (Figures 6C and 6D and Figure S5). Under these conditions, we observed that female viability is restored for Msl1 mut.3 and mut.4 and partially for mut.5, whereas male-specific lethality was still observed for all mutants (except escapers for Msl1 mut.1 and mut.4). Ectopic expression of WT Msl1 in these conditions had no effects on viability. These results show that dominant-negative effects of all mutations can be observed

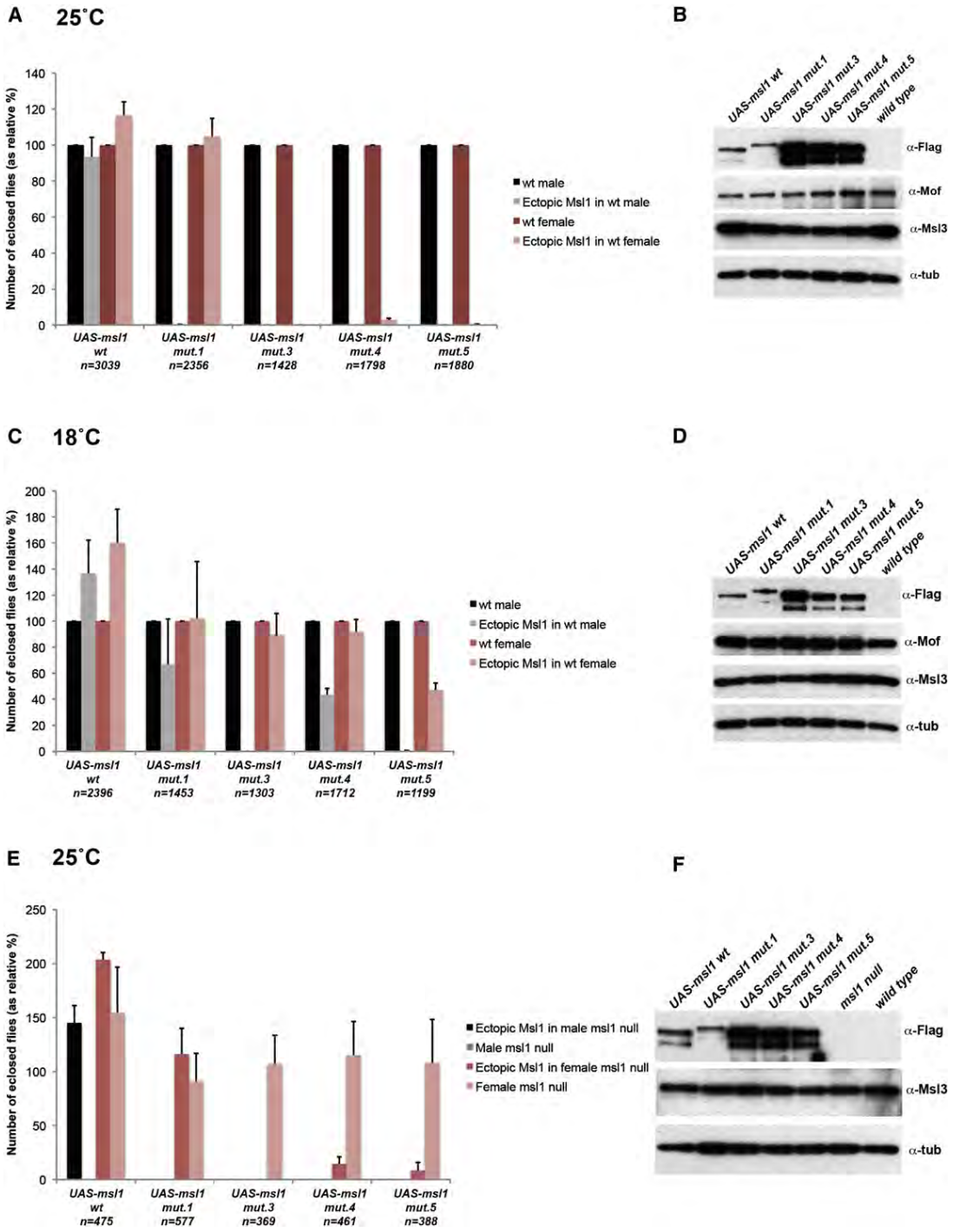


Figure 6. Viability of Adult Flies upon *tubGal4*-Induced Ectopic Expression of *UAS-msl1

(A) Ectopic expression of WT Msl1 and Msl1 mut.1, mut.3, mut.4, and mut.5 in a WT background at 25°C. The nonexpressing *TM6Tb/UAS-msl1** (* representing Msl1 WT, mut.1, mut.3, mut.4, or mut.5) flies were used as internal controls (males, black bars; female, red bars) and scored as 100% viable. The number of eclosed *tubGal4/UAS-msl1** flies (males, gray bars; female, pink bars) expressing the *UAS-msl1** transgene are represented as relative percentages to their nonexpressing siblings. The details of the fly cross are indicated in Experimental Procedures. The total number of flies counted for each cross is indicated underneath each group. The error bars represent standard deviations of three independent crosses.

(B) Western blots from protein extracts prepared from second instar larvae carrying different *UAS-msl1** transgenes, all C-terminal 3xFlag tagged. Flag antibody was used to probe exogenous Msl1; Mof and Msl3 protein levels are shown for comparison. Tubulin levels were used as a loading control (Anti-Tubulin, DM1A, Sigma).

exclusively in males at both temperatures, whereas females become sensitive to the levels of Msl1 mut.3, mut.4, and mut.5 at 25°C.

To assess the direct effect of the mutations, we expressed the Msl1 mutant variants in *msl1*^{L60}/*msl1*$^{\gamma 269}$ null mutant flies to reconstitute Msl1 function. As expected, in the absence of Msl1, females are viable whereas males die as third instar larvae or at early pupal stages (Figures 6E and 6F). At 25°C, *tubulin-Gal4*-induced ectopic expression of WT Msl1 rescued completely the *msl1* loss-of-function male-specific lethality (Figure 6E). Noticeably, none of the *msl1* mutants rescued male lethality (Figure 6E). Female viability dropped significantly in Msl1 mut.3, mut.4, and mut.5, similar to the dominant effect observed upon overexpression in a WT background (Figure 6E). At 18°C, *tubulin-Gal4*-induced ectopic expression of WT Msl1 rescued the *msl1* loss-of-function male-specific lethality only partially, and other mutants failed to do so (data not shown). These results clearly show that all of the residues that are determined from the crystal structure are absolutely essential for male viability.

DISCUSSION

Msl1 and Msl2 are essential core subunits of the dosage compensation complex that contains Msl3 and Mof as well as more peripherally bound Mle. Here we report the crystal structure of the human MSL1/MSL2 subcomplex, together with a detailed biochemical and functional mutagenesis analysis in vitro and in vivo. Our study revealed several important and intriguing features of the MSL complex architecture. Contrary to the expected coiled coil-based heterodimerization with MSL2 (Li et al., 2005) we show that the MSL1 coiled-coil region mediates MSL1 homodimerization. Putative self-association of *Drosophila* Msl1 via a glycine-rich region between residues 26 and 84 was proposed by Li et al., but its oligomeric state was not further characterized (Li et al., 2005). Our data clearly show that MSL1 forms dimers, and the evolutionary nonconserved glycine-rich region is dispensable for this interaction. The MSL1 dimeric coiled coil then binds two molecules of MSL2, forming a heterotetrameric core of the MSL complex. The structural data as well as immunoprecipitation experiments demonstrate that the Msl1 dimerization is required for the interaction with Msl2, since the single Msl1 molecule contains only half of the Msl2 binding site. The fact that the Msl1 dimerization can be observed in female cells, lacking Msl2, suggests that Msl2 is not required for the Msl1 dimer formation. This was additionally confirmed by mutagenesis of the Msl2 binding surface on Msl1. Finally, we show that both Msl1 dimerization and Msl2 binding are independent of the presence of Msl3 or Mof. Importantly, since all our structure-based Msl1 mutations specifically affect only the targeted interfaces, without influencing interaction with other MSL proteins and gene promoters, we believe that they do not significantly perturb the overall structure of Msl1 that is anyhow predicted to be mostly intrinsically disordered. We had reported earlier that MSL1 interacts with MSL3 and MOF with relatively short segments (20 and 40 residues, respectively), while the surrounding regions are poorly conserved and predicted to be unstructured (Kadlec et al., 2011). Consistent with this, the MSL1/MSL2 structure reveals yet another short, highly conserved region that is used both for MSL1 dimerization and the MSL2 binding, further reinforcing the scaffolding role of MSL1.

Msl1 and X Chromosome Recognition

Using our Msl1 mutants, we are able to create partial complexes and assess specifically the role of individual MSL subunits. We found that Msl1 per se cannot recognize X chromosomal features other than promoters (Figure S6). However, binding of Msl2 to the Msl1 dimer has two important consequences: rudimentary recognition of X chromosome and roX2 RNA integration into the complex. It is clear that HAS are not identical in terms of their affinity to the complex. Msl1 mut.1 ChIP experiment shows that roX2 HAS only requires Msl1/Msl2 while all other tested HAS show reduced level of the complex. We propose that chromatin regions like roX2 HAS are the "elementary high-affinity sites" where initial enrichment of Msl1 on X chromosome is mediated by a common action of Msl1/Msl2.

Our RIP results show that Msl1-Mof-Msl3 trimer or hexamer (corresponding to Msl1 mut.3 and mut.4) cannot bind roX RNA, indicating an active role of Msl2 in the binding. Loss of RNA signal in RIP experiments cannot be due to a loss of interaction with Mle, as Mle association with the complex is bridged by RNA (Richter et al., 1996; Copps et al., 1998; Smith et al., 2000). Interestingly, Msl1 mut.1 also shows significant loss of roX RNA interaction, implying that full integration happens only in the context of the whole complex. It is tempting to speculate that roX2 RNA interaction may enable crosstalk between the two distant N-terminal Msl1/Msl2 and C-terminal Msl1-Msl3-Mof catalytic centers of the MSL complex.

Dimerization of Msl1 Enables Spreading of the Complex along Gene Bodies

The fact that MSL1 dimerizes through such an extended interface (buried surface of 1340 Å^2) and the dimer formation is

(C) Recovery of female but not male viability upon weaker/lower ectopic expression of Msl1 mut.3, mut.4, and mut.5 in a WT background at 18°C. Crosses were performed as in (A), but the flies had been kept at 18°C.
(D) Western blots were done as in (B).
(E) Male lethality rescue assay by ectopic expression of WT Msl1 and Msl1 mut.1, mut.3, mut.4, and mut.5 in *msl1* null (*msl1*^{L60}/*msl1*$^{\gamma 269}$) background at 25°C. The number of eclosed *msl1*^{L60}/*msl1*$^{\gamma 269}$*; tubGal4/UAS-msl1** flies (males, black bars; females, red bars) expressing the *UAS-msl1** transgene in *msl1* null mutant background are represented as relative percentages to their nonexpressing heterozygous siblings (*msl1*$^{L60\ or\ \gamma 269}$ */CyO, GFP; UAS-msl1*/TM6Tb*). The nonexpressing *msl1* null siblings (*msl1*^{L60}/*msl1*$^{\gamma 269}$*; UAS-msl1*/TM6Tb)* (males, gray bars; females, pink bars) are shown as internal controls for the male-specific lethality of *msl1* loss of function. The cross is indicated in Experimental Procedures. The error bars represent standard deviations of three independent crosses.
(F) Tubulin antibody was used for loading control and Msl3 and Flag antibodies were probed to show the levels of expression of transgenic Msl1 and endogenous Msl3.

required for MSL2 binding dramatically changes our view on the dosage compensation complex structure and assembly. As it is possible to copurify the recombinant human MSL1/MSL2/MSL3/MOF complex from insect cells using Flag-tagged MSL2 (Wu et al., 2011) that can presumably only bind dimeric MSL1, it is very likely that the MSL complex contains all the subunits in pairs, including also MSL3 and MOF. We therefore propose that MSL complex binding to the open reading frames of the X-linked genes in *Drosophila* could happen through a dimer-dependent nucleosome engagement rather than via Msl1-mediated oligomerization of MSL complexes on chromatin (Li et al., 2005) (Figure S6). The presence of two copies of each of the chromatin-modifying or modification-binding domains of the complex would increase the number of possible, probably transient contacts with nucleosome(s), containing histones also in pairs. The *Drosophila* Msl1 scaffold is a large, mostly disordered protein (1039 residues) that provides the MSL complex with a high degree of flexibility. The Msl2 and Mof/Msl3 binding regions of Msl1 are separated by 720 poorly conserved, probably unstructured residues. It is thus possible that while some subunits are attached to chromatin, others, connected by the flexible Msl1 linker, can browse the surrounding nucleosomes for new attracting histone marks. The dimer-dependent spreading can also be deduced from the ChIP analysis of Msl3/Mof-deficient Msl1 mutant in the endogenous Msl1 background (Figure 4A). This mutant can still dimerize with the endogenous intact Msl1, albeit at low levels observed from our IP analysis (Figure 3F); however, it cannot spread to the open reading frames, which indicates that both copies of Msl3 and Mof are required for spreading.

Sex-Independent Binding of Msl1 at Promoters

The occurrence of Msl1 at the promoters in both sexes and its independence from other members of the complex for this binding suggests the possibility of an evolutionarily conserved function in higher eukaryotes. All complex members, except for Msl2 and Msl1, have origins traceable to yeast (Marín, 2003). The emergence of "Msl1-like genes," namely Msl1 and Nsl1 in *Drosophila*, both having a PEHE region to bind Mof through the same surface (Kadlec et al., 2011), seems to focus this ubiquitous acetyltransferase to promoter regions of a large portion of the *Drosophila* genome. Indeed, Mof binds to promoters in both sexes and is responsible for the promoter chromatin H4K16 acetylation (Kind et al., 2008). It was also observed that RNAi of Nsl1 or Msl1 does not completely diminish Mof occupancy at the promoter, probably because both proteins have complementary roles (Raja et al., 2010). It will be interesting to delineate possible functional interplay of Msl1 and Nsl1 at promoters as well as the distribution of Mof between these two proteins. It is important to note that Msl1 is not essential for female viability, possibly due to this redundancy between Msl1 and Nsl1 in terms of Mof recruitment to the promoters. Female viability decreases only when Msl1 mutants that have an intact PEHE region are expressed, probably causing mistargeting of Mof. In addition, no effect is observed upon expression of WT Msl1 or mut.1, strengthening the hypothesis that the observed female phenotypes are due to Mof rather than Msl1. One distinguishing factor between the promoter complex and the dosage compensation complex is Msl3, whose binding on the autosomal promoters was undetectable and on X-linked promoters was also very low.

In summary, our study enhances our perspective on the architecture of MSL complex and how this configuration could help spreading of MSL complex on the X chromosome. Future structural work incorporating the remaining MSL complex subunits, including RNA, is likely to reveal novel insights into the molecular mechanisms underlying this chromatin-binding complex.

EXPERIMENTAL PROCEDURES

Expression, Purification, and Crystallization

A His-tag fusion of a proteolytically stable fragment of the human MSL1 coiled-coli region (residues 213–267) and untagged N-terminal domain of $MSL2_{1-116}$ was coexpressed in bacteria. The purified complex crystallized in a solution containing 0.1 M HEPES (pH 7.0), 1.1 M sodium malonate (pH 6.5), and 0.8% Jeffamine ED 2001. The complex of $MSL1_{213-252}$ and $MSL2_{1-116}$ was produced in the same way, and crystals were obtained in 0.1 M MES (pH 6.0) and 1.6 M ammonium sulfate.

Data Collection and Structure Determination

The structure of the $MSL1_{213-267}$/$MSL2_{1-116}$ was solved by a zinc multiple anomalous dispersion (MAD) experiment and was refined to R factor of 25.6% and R_{free} of 29.6% (Table 1) with all residues in allowed (96.5% in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY (Davis et al., 2004). The structure of the $MSL1_{213-254}$/$MSL2_{1-116}$ was solved by molecular replacement using PHASER (McCoy et al., 2005) and the MSL1/MSL2 model obtained above as a search model. The structure was refined using Refmac5 (with TLS refinement) to R factor of 24.4% and R_{free} of 26%. All the residues are in allowed (96% in favored) regions of the Ramachandran plot. A representative part of the $2F_o - F_c$ electron density map calculated using the refined model is shown in Figure S7. Details of the crystallography procedures are available in Supplemental Experimental Procedures.

Flag Immunoprecipitations

Harvested cells were washed with cold PBS two times and resuspended in 1 ml HEMGT 150 (25 mM HEPES/NaOH 7.6, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% glycerol, 0.2% Tween-20, 150 mM NaCl, supplemented with protease inhibitor cocktail from Roche) buffer. After three freeze/thaw cycles in liquid N_2 and 37°C water bath, the extract was centrifuged for 30 min at 20,000× g. A bed volume of 30 μl M2-Flag agarose beads (Sigma) was incubated for 3 hr at 4°C. The beads were washed five times in HEMGT 250 and boiled with 40 μl 4X Laemmli Buffer.

Chromatin Immunoprecipitation in SL-2 Cells

ChIP was carried out as described in Kadlec et al., 2011 with a few modifications based on Chelex protocol adapted from Nelson et al., 2009. Details are available in Supplemental Experimental Procedures.

RNA Immunoprecipitation

RIP was carried out as described in Selth et al., 2009 with 25 million SL-2 stable lines that had been induced with $CuSO_4$ for 12 hr with indicated amounts in Figure S4.

Drosophila Crosses

In order to obtain flies ectopically expressing mutant *msl1* in a WT background, y^1 *w*; P{tubP-GAL4}LL7/ TM6B*, *P{Ubi-GFP.S65T}PAD2*, Tb^1 virgin females were crossed with males homozygous for the appropriate *UAS-msl1** transgenic insertion. For analysis in *msl1* null mutant background, y^1 *w*;* $msl1^{L60}$*/ CyO, P{ActGFP}JMR1; P{tubP-GAL4}LL7/ TM6B*, Tb^1 virgin females were

Table 1. Data Collection and Refinement Statistics

	$MSL1_{213-252}$ $MSL2_{1-116}$	$MSL1_{213-267}$ $MSL2_{1-116}$	Peak	Inflection point	Remote
Data collection					
Space group	$P3_221$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Cell dimensions					
a, *b*, *c* (Å)	101, 101, 88.6	104.6, 182.2, 89.5	104.1, 180.60, 89.4	104.3, 180.8, 89.7	104.8, 181.7, 90.3
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	31–3.25 (3.37–3.25)[a]	50–3.5 (3.65–3.5)	50–3.6 (3.73–3.6)	50–3.65 (3.75–3.65)	50–3.7 (3.83–3.7)
R_{merge}	6.6 (68)	7.3 (72)	6.8 (77)	6.6 (77)	6.8 (78)
I/σI	11.6 (1.9)	11 (1.8)	14.5 (1.9)	14.8 (1.9)	14.4 (1.9)
Completeness (%)	98.7 (99.9)	97.2 (98.2)	99.8 (99)	99.9 (99.9)	99.6 (96.4)
Redundancy	3.6 (3.8)	3.7 (3.7)	3.9 (3.9)	3.9 (3.9)	3.9 (3.7)
Refinement					
Resolution (Å)	31–3.25	45–3.5	–	–	–
No. reflections	8,005	20,450	–	–	–
R_{work}/R_{free}	24.4/26	25.6/29.6	–	–	–
No. atoms					
Protein	2,260	7,030	–	–	–
Zn ion	4	12	–	–	–
B factors					
Protein	57	79	–	–	–
Zn ion	72	84	–	–	–
Rmsds					
Bond lengths (Å)	0.009	0.007	–	–	–
Bond angles (°)	1.042	0.956	–	–	–

[a]Values in parentheses are for highest-resolution shell.

crossed with *y^1 w*; msl1$^{\gamma 269}$ cn^1 bw^1/CyO, P{ActGFP}JMR1; P{UAS-msl1*}65B2* males.

ACCESSION NUMBERS

Coordinates of $MSL1_{213-267}$/$MSL2_{1-116}$ and $MSL1_{213-254}$/$MSL2_{1-116}$ have been deposited to the Protein Data Bank with accession codes 4B86 and 4B7Y, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, one table, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.09.014.

ACKNOWLEDGMENTS

We are grateful to Andrea Pichler's lab for their insightful suggestions on the ubiquitin-related experiments and Sarah Toscano for help with transgenic flies. We thank the EMBL-ESRF-ILL-IBS Partnership for Structural Biology (PSB) for access to structural biology instrumentation, notably L. Signor of the Institut de Biologie Structurale (IBS) Grenoble for his help with mass spectrometry, the High Throughput Crystallisation Laboratory (HTX) group of EMBL Grenoble for performing initial screening crystallization trials, and J. Perard for help with the MALLS experiment. We thank the European Synchrotron Radiation Facility (ESRF)-EMBL Joint Structural Biology Group for access to and assistance on the ESRF synchrotron beamlines. We are grateful to members of both labs for critical reading of the manuscript and helpful suggestions. This work was supported by the EU FP7 funded Network of Excellence EpigGeneSys awarded to A.A. and S.C. E.H. was supported by Darwin Trust Fellowship.

Received: March 14, 2012
Revised: July 9, 2012
Accepted: September 10, 2012
Published online: October 18, 2012

REFERENCES

Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol. Cell *5*, 367–375.

Akhtar, A., Zink, D., and Becker, P.B. (2000). Chromodomains are protein-RNA interaction modules. Nature *407*, 405–409.

Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.-K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J., and Kuroda, M.I. (2008). A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. Cell *134*, 599–609.

Alvarez, S.E., Harikumar, K.B., Hait, N.C., Allegood, J., Strub, G.M., Kim, E.Y., Maceyka, M., Jiang, H., Luo, C., Kordula, T., et al. (2010). Sphingosine-1-phosphate is a missing cofactor for the E3 ubiquitin ligase TRAF2. Nature *465*, 1084–1088.

Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. *12*, 429–442.

Borden, K.L., Boddy, M.N., Lally, J., O'Reilly, N.J., Martin, S., Howe, K., Solomon, E., and Freemont, P.S. (1995). The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. EMBO J. *14*, 1532–1541.

Brzovic, P.S., Rajagopal, P., Hoyt, D.W., King, M.C., and Klevit, R.E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. Nat. Struct. Biol. *8*, 833–837.

Budhidarmo, R., Nakatani, Y., and Day, C.L. (2012). RINGs hold the key to ubiquitin transfer. Trends Biochem. Sci. *37*, 58–65.

Conrad, T., and Akhtar, A. (2011). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat. Rev. Genet. *13*, 123–134.

Conrad, T., Cavalli, F.M.G., Holz, H., Hallacli, E., Kind, J., Ilik, I., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. (2012a). The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. Dev. Cell *22*, 610–624.

Conrad, T., Cavalli, F.M.G., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. (2012b). Drosophila dosage compensation involves enhanced Pol II recruitment to male X-linked promoters. Science *337*, 742–746.

Copps, K., Richman, R., Lyman, L.M., Chang, K.A., Rampersad-Ammons, J., and Kuroda, M.I. (1998). Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. EMBO J. *17*, 5409–5417.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. *32*(Web Server issue), W615–W619.

Deng, X., Hiatt, J.B., Nguyen, D.K., Ercan, S., Sturgill, D., Hillier, L.W., Schlesinger, F., Davis, C.A., Reinke, V.J., Gingeras, T.R., et al. (2011). Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. Nat. Genet. *43*, 1179–1185.

Fauth, T., Müller-Planitz, F., König, C., Straub, T., and Becker, P.B. (2010). The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. Nucleic Acids Res. *38*, 3209–3221.

Franke, A., and Baker, B.S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol. Cell *4*, 117–122.

Gelbart, M.E., and Kuroda, M.I. (2009). Drosophila dosage compensation: a complex voyage to the X chromosome. Development *136*, 1399–1410.

Groth, A.C. (2004). Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage C31. Genetics *166*, 1775–1782.

Hallacli, E., and Akhtar, A. (2009). X chromosomal regulation in flies: when less is more. Chromosome Res. *17*, 603–619.

Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A., and Lucchesi, J.C. (1997). mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. EMBO J. *16*, 2054–2060.

Ilik, I., and Akhtar, A. (2009). roX RNAs: non-coding regulators of the male X chromosome in flies. RNA Biol. *6*, 113–121.

Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S., and Akhtar, A. (2011). Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. Nat. Struct. Mol. Biol. *18*, 142–149.

Kharchenko, P.V., Xi, R., and Park, P.J. (2011). Evidence for dosage compensation between the X chromosome and autosomes in mammals. Nat. Genet. *43*, 1167–1169, author reply 1171–1172.

Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell *133*, 813–828.

Kruse, J.-P., and Gu, W. (2009). MSL2 promotes Mdm2-independent cytoplasmic localization of p53. J. Biol. Chem. *284*, 3250–3263.

Lee, T., and Luo, L. (1999). Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. Neuron *22*, 451–461.

Lee, C.G., Chang, K.A., Kuroda, M.I., and Hurwitz, J. (1997). The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. EMBO J. *16*, 2671–2681.

Li, F., Parry, D.A.D., and Scott, M.J. (2005). The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. Mol. Cell. Biol. *25*, 8913–8924.

Marín, I. (2003). Evolution of chromatin-remodeling complexes: comparative genomics reveals the ancient origin of "novel" compensasome genes. J. Mol. Evol. *56*, 527–539.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr. D Biol. Crystallogr. *61*, 458–464.

Meller, V.H., and Rattner, B.P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J. *21*, 1084–1091.

Mondal, K., Dastidar, A.G., Singh, G., Madhusudhanan, S., Gande, S.L., VijayRaghavan, K., and Varadarajan, R. (2007). Design and isolation of temperature-sensitive mutants of Gal4 in yeast and Drosophila. J. Mol. Biol. *370*, 939–950.

Morales, V., Regnard, C., Izzo, A., Vetter, I., and Becker, P.B. (2005). The MRG domain mediates the functional integration of MSL3 into the dosage compensation complex. Mol. Cell. Biol. *25*, 5947–5954.

Nelson, J., Denisenko, O., and Bomsztyk, K. (2009). The fast chromatin immunoprecipitation method. Methods Mol. Biol. *567*, 45–57.

Raja, S.J., Charapitsa, I., Conrad, T., Vaquerizas, J.M., Gebhardt, P., Holz, H., Kadlec, J., Fraterman, S., Luscombe, N.M., and Akhtar, A. (2010). The nonspecific lethal complex is a transcriptional regulator in Drosophila. Mol. Cell *38*, 827–841.

Richter, L., Bone, J.R., and Kuroda, M.I. (1996). RNA-dependent association of the Drosophila maleless protein with the male X chromosome. Genes Cells *1*, 325–336.

Scott, M.J., Pan, L.L., Cleland, S.B., Knox, A.L., and Heinrich, J. (2000). MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. EMBO J. *19*, 144–155.

Selth, L.A., Gilbert, C., and Svejstrup, J.Q. (2009). RNA immunoprecipitation to determine RNA-protein associations in vivo. Cold Spring Harb. Protoc. *2009*, t5234.

Smith, E.R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R.G., Allis, C.D., and Lucchesi, J.C. (2000). The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. Mol. Cell. Biol. *20*, 312–318.

Smith, E.R., Cayrou, C., Huang, R., Lane, W.S., Côté, J., and Lucchesi, J.C. (2005). A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. Mol. Cell. Biol. *25*, 9175–9188.

Uchiki, T., Kim, H.T., Zhai, B., Gygi, S.P., Johnston, J.A., O'Bryan, J.P., and Goldberg, A.L. (2009). The ubiquitin-interacting motif protein, S5a, is ubiquitinated by all types of ubiquitin ligases by a mechanism different from typical substrate recognition. J. Biol. Chem. *284*, 12622–12632.

Wu, L., Zee, B.M., Wang, Y., Garcia, B.A., and Dou, Y. (2011). The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. Mol. Cell *43*, 132–144.

Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. Cell *102*, 533–539.

Non-specific lethal complex (NSL)

The NSL complex was first identified in *Drosophila melanogaster* and mammalian cells (Mendjan et al., 2006). It is a large multiprotein assembly that was shown to consist of seven proteins: MOF, NSL1, NSL2, NSL3, MBD-R2, WDS and MCRS2. The human NSL complex was suggested to contain the corresponding seven members (MOF, KANSL1, KANSL2, KANSL3, PHF20, WDR5, MCRS1) (Figure 7) and possibly two additional subunits, OGT1 (O-linked N-acetylglucosamine (O-GlcNAc) transferase) and HCF-1 (Host cell factor -1)(Cai et al., 2010). It remains unclear, whether OGT and HCF-1 are genuine members of the complex. So far these two subunits were not involved in our analyses.

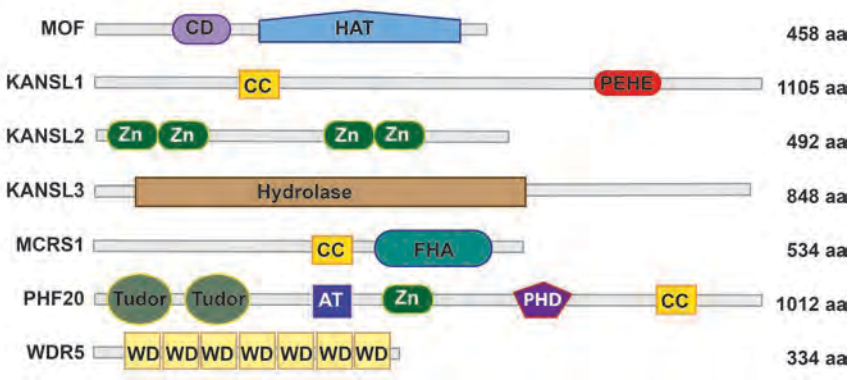


Figure 7. Schematic representation of the domain structure of the NSL complex subunits. HAT- histone acetyltransferase domain, WD- WD repeat, CC- coiled coil, CD- chromo barrel domain, AT- AT hook.

Several subunits of the NSL complex are also present in other chromatin complexes. As described above, MOF is found in the dosage compensation complex (MSL complex); WDR5 is also present in the MLL/COMPASS and ATAC complexes and MCRS1 and the Zn-coordinating domains of KANSL2 exist within the INO80 remodeller (Figure 8).

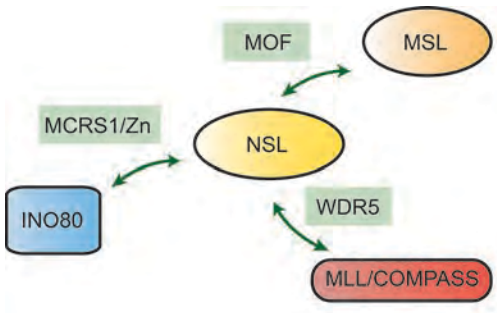


Figure 8. The NSL complex shares multiple subunits with other epigenetic complexes.

While the complex has been identified only recently, its subunits are known to be essential for various cellular processes. All the subunits are essential for viability in *Drosophila* or mammals (Badeaux et al., 2012; Mendjan et al., 2006). The best-characterized subunit is the histone acetyltransferase MOF that possess the only currently known catalytic activity of the complex (see above). In human, haploinsufficiency of KANSL1 causes a multisystem disorder called the Koolen-de Vries syndrome that is characterized by intellectual disability and has a prevalence of 1 in 16,000 (Koolen et al., 2012). KANSL2 gene is altered in patients with severe intellectual disability (Gilissen et al., 2014). The WDR5 subunit is an important regulator of the embryonic stem cell transcriptional network and PHF20 is required for somatic cell reprogramming (Ang et al., 2011; W. Zhao et al., 2013).

The NSL complex functions as a major transcription regulator as it associates with promoters of thousands of genes where it acetylates nearby H4 histones and loss of the NSL complex binding severely affects their expression levels (Chelmicki et al., 2014; Lam et al., 2012; Raja et al., 2010). The majority of the NSL complex-bound targets were shown to belong to housekeeping genes and the NSL complex was required for efficient recruitment of RNA polymerase II to their promoter regions (Feller et al., 2012; Lam et al., 2012). The NSL complex regulates transcription in embryonic stem cells and neuronal progenitors by targeting promoters and TSS-distal enhancers (Chelmicki et al., 2014). Interestingly, the NSL complex and its subunits were recently shown to have another role during mitosis, where they bind to and stabilize microtubule minus-ends (Meunier et al., 2015). The NSL complex subunits (including MOF, WDR5, KANSL1, PHF20 and MCRS1) are also associated with certain primary cancers. Loss of H4K16ac has been identified as a common hallmark in human cancer (Fraga et al., 2005) and MOF has been shown to be down-regulated in primary breast carcinoma and medulloblastoma (Pfister et al., 2008). In contrast, MOF overexpression was shown to correlate with increased cellular proliferation, oncogenic transformation and tumour growth (Gupta et al., 2008; Zhao et al., 2013). MOF is also for example overexpressed in human non-small cell lung cancer (Chen et al., 2014). KANSL1 is mutated in Down syndrome related acute megakaryoblastic leukemia (Yoshida et al., 2013). MCRS1 acts as an oncogene during fibroblast transformation, and this activity can be suppressed by the tumor suppressor PTEN (Okumura et al., 2005). MCRS1 overexpression has been documented in a variety of human cancers inducing proliferation of cancer cells.

While, new important roles and functions of the NSL complex and its subunits are being reported continuously, major questions related to this complex remain unanswered. It still remains unclear how the NSL complex mediates its functions. Is the histone acetylation the only catalytic activity of the complex and how it is regulated? How is the complex targeted to chromatin? Similarly to other chromatin-regulatory complexes, the molecular architecture of the complex is unknown. How do the individual subunits interact with each other and what are the dynamics of its composition and assembly? How does the NSL complex cross-talk with other chromatin complexes. The aim of our project is to structurally and functionally analyse the NSL complex to better understand its epigenetic regulatory functions and its role in human disease.

NSL complex - previous work

General overview of the strategy used for the structural analysis of the NSL complex

Epigenetic regulatory complexes represent a considerable challenge for structural analyses due to their large size and inherent flexibility. Most of the complexes are poorly characterized at biochemical level and their subunit composition is often not well established. With the recent exception of PRC2 (Jiao and Liu, 2015), essentially none of the large chromatin complexes has been fully structurally characterized at high resolution. Structures of complexes with a nucleosome are now becoming reality, but still limited to small domains.

Similarly, to the MSL complex, most of the NSL complex subunits are large multi-domain proteins (Figure 7). The individual subunits contain multiple domains linked by long, poorly conserved regions that are mostly predicted to be intrinsically disordered (Figures 7,9). These disordered linkers have often different lengths and sequence in different species and provide these complexes with a high level of flexibility. Given that there is essentially no biochemical data available on this complex, attempting directly structural analysis of the full complex might be too ambitious. We have thus decided to analyse, in parallel, individual binary interactions within the complex that will eventually be valuable for analysis of the larger modules or the full NSL complex.

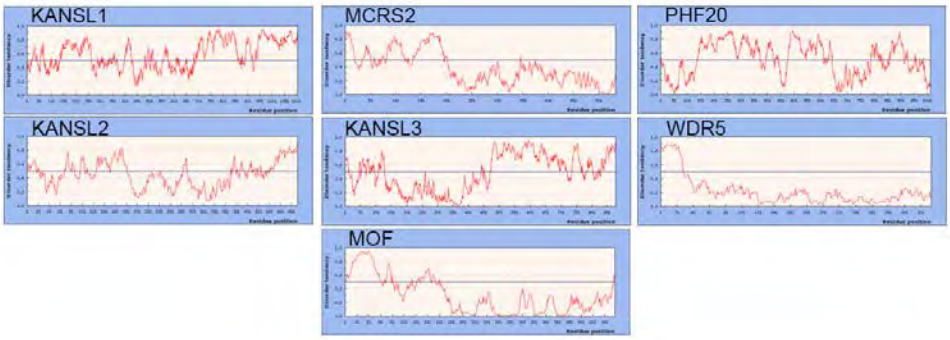


Figure 9. Disorder prediction of the NSL complex subunits using prediction server IUPRED (http://iupred.enzim.hu). The red line indicates the disorder tendency per residue in a primary sequence along x axis. Only WDR5 and MOF are predicted to be well ordered.

The strategy we have been following to obtain structural information on the molecular architecture of these complexes is described below:

1. Identify and characterise protein-protein interactions within the complex.
2. Obtain structures on individual sub-complexes.
3. Analyse hypotheses based on the structural work in functional assays. The information obtained at this level will significantly help our understanding of the role of individual proteins and their interactions as well as it will be essential for structural analysis of larger complexes.
4. Engineer individual subunits using information obtained above to prepare a less flexible full complex or individual modules (structurally/functionally autonomous parts of the complex).
5. Obtain structure of the modules and the full complex by hybrid methods.
6. Analyse interactions of the modules or the full complex with the substrate nucleosome.

NSL binary interactions

As a starting point for this analysis, each of the *Drosophila* or human NSL complex subunits was individually produced in insect cells as Flag fusion and co-purified with untagged versions of the other subunits using an anti-Flag resin. This way we identified several putative binary interactions among the subunits. The schematic representation of the interaction network within the NSL complex is shown in Figure 10A.

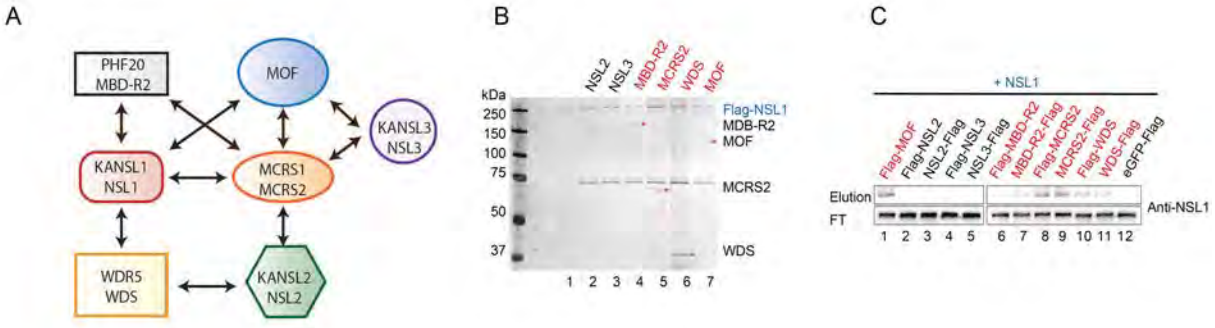


Figure 10. (A) Preliminary interaction map based on pull-down experiments with Flag tagged NSL proteins produced in insect cells in Akhtar lab. Both the human and *Drosophila* protein names are shown. (B, C) A representative pull-down assay result. (B). A silver stained SDS-PAGE gel. Flag-NSL1 was used as the bait and other untagged NSLs as prey. Interacting proteins are indicated by asterisks and highlighted in red. (C) Western

blot analysis of the reverse Flag pull-down assays of B using untagged NSL1 as the prey and either N-terminal or C-terminal Flag-tagged NSLs as bait.

KANSL1-MOF interaction

KANSL1 functions as a scaffold of the NSL complex interacting with the MOF, PHF20, WDR5 and MCRS1 subunits (Figure 11)(Dias et al., 2014). We have described the interaction of the KANSL1 PEHE domain with the HAT of MOF. Since the KANSL1/MOF complex resisted crystallization, we used for our analysis the structure of the complex between the related PEHE domain of MSL1 (of the MSL complex above) and MOF (Figure 11). All the key interacting residues are perfectly conserved in KANSL1 and we could show that their mutagenesis results in removal of MOF from the complex (Kadlec et al., 2011) (Figure 11).

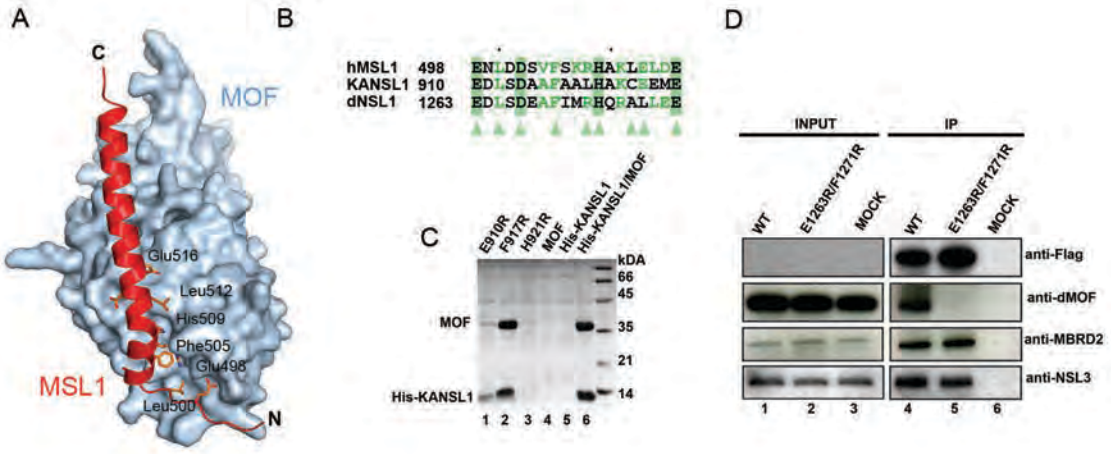


Figure 11. (A) Crystal structure of the MOF HAT domain in complex with the PEHE domain of the MSL1 subunit of the MSL complex. (B) Sequence alignment of human MSL1 and NSL1 proteins showing that the MOF-binding residues (green triangles) are conserved in KANSL1 and dNSL1. (C) SDS-PAGE analysis of the binding of His-tagged $KANSL1_{883–952}$ mutants (indicated above the lanes) to co-expressed $MOF_{174–458}$ after purification using Ni^{2+}resin. Unbound KANSL1 becomes unstable (lanes 3,4,5). (D) *Drosophila* NSL1 interactions *in vivo* in SL-2 cells transiently transfected with wild-type (WT) NSL1 and the indicated mutants. Shown are data from immunoprecipitation (IP) of whole-cell extracts taken 48 h after transfection using Flag-agarose resin. Mock lane represents an empty vector.

KANSL1-WDR5-KANSL2 sub-complex

We have also identified and structurally characterized the interactions of WDR5 with KANSL1 and KANSL2 (Dias et al., 2014). This work included supervision of an EMBL PhD student Jorge Dias and the Mittweida University Master student Janine Brettschneider. The structures of human and *Drosophila* KANSL1/WDR5/KANSL2 revealed how the β-propeller domain of WDR5 simultaneously recognizes the WIN (WDR5 interacting motif) of KANSL1 and a short hydrophobic motif of KANSL2. These interactions of WDR5 are very similar to its interactions with the MLL and RbBP5 proteins within the MLL/COMPASS histone methyltransferase complexes. The NSL and MLL/COMPASS complexes had previously been proposed to form a large WDR5–linked assembly that would possess both H3K4 methylation and H4 acetylation activities (Dou et al., 2005). However, our structural and biophysical characterization of these complexes clearly showed that the interactions of WDR5 within the MOF-containing NSL complex and MLL complexes are mutually exclusive and WDR5 thus cannot be the shared subunit linking them (Figure 12A,B). The mutagenesis of NSL1 residues interacting with WDR5 results in the elimination of WDR5 and NSL3 from the complex (Figure 12C). Using structure-based KANSL1 mutants in transgenic flies, the KANSL1–WDR5 interaction was shown to be

required for proper assembly, efficient recruitment of the NSL complex to target promoters, and fly viability (Figure 12D) (Dias et al., 2014).

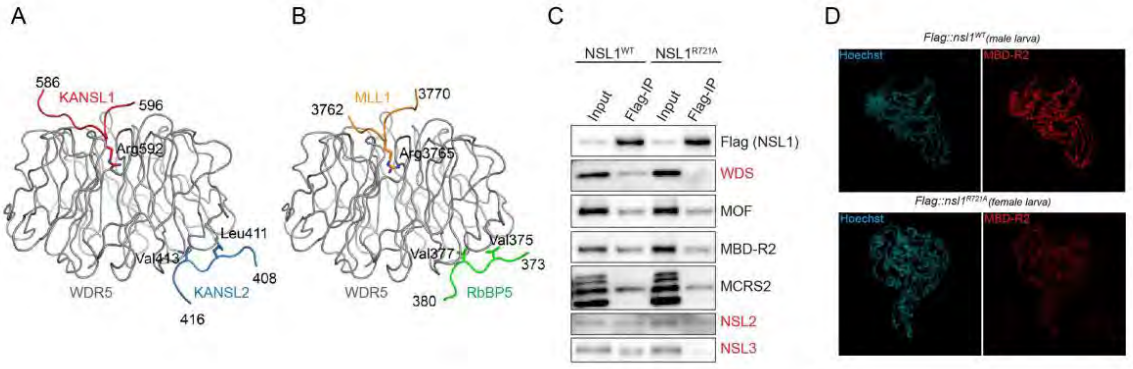


Figure 12. Comparison of the crystal structures of KANSL1-WDR5-KANSL2 (A) and MLL1-WDR5-RbBP5 (PDB code 3P4F) complexes (B). The two complexes were superimposed using WDR5. (C) Flag immunoprecipitation of wild-type (WT) *Drosophila* NSL1 and NSL1^{R721A} (Arg721 corresponds to human Arg592 mediating the interaction with WDR5) in SL-2 cells. Antibodies used for Western blot analysis are indicated. (D) Polytene chromosome immunostaining of *Flag::nsl1*^{WT} male and *Flag::nsl1*^{R721A} female third instar larvae using anti-MBD-R2 antibody.

NSL complex - future plans

The project has two main objectives:

1. Structural analysis of the NSL sub-complexes
We will biochemically, biophysically and structurally characterise individual NSL subunits, their domains and mutual complexes. The hypotheses based upon the structural work will be used in rational functional experiments in cell lines and animals. Together, this analysis will shed new light on function of the NSL members, but will also be essential for the subsequent structural analysis of the full NSL complex and its modules.

2. Structural analysis of the full NSL complex and its interaction with nucleosome
The aim will be to produce larger or full NSL complexes. The combination of the available data on individual NSL sub-complexes and the results generated in objective 1 will help us rationally engineer and produce the full NSL complex. The full NSL complex and/or its modules (e.g. chromatin binding or acetyltransferase) will be characterised and subject to structural analysis. We will also analyse the interactions of these complexes with the nucleosome as substrate.

1. Structural analysis of the NSL sub-complexes

KANSL2

KANSL2 contains two cysteine/histidine rich, possibly Zn-coordinating, regions of unknown fold and function (Figure 13). As described above, we had shown that a separate C-terminal hydrophobic DLDV motif interacts with the WDR5 subunit with dissociation constant of 8 μM (Dias et al., 2014) (Figure 11). A similar short motif of RbBP5 is involved in the recruitment of WDR5 into the MLL/COMPASS complex (Avdic et al., 2011; Odho et al., 2010). A recent work has also shown that WDR5 recognizes a similar motif of MYC and this interaction is important for the recruitment of MYC to chromatin (Thomas et al., 2015). When we mutated the KANSL2 residues critical for this interaction, neither KANSL2 nor WDR5 were eliminated from the complex (Dias et al., 2014). Given the relatively low affinity for WDR5, we believe that KANSL2 is maintained within this complex through its interactions with other NSL subunits.

In our preliminary interaction analysis of the NSL complex, KANSL2 interacts only with WDR5 and MCRS1 (Figure 10A). We performed pull-down experiments with various MCRS1 and KANSL2 deletion mutants and could show that KANSL2 interacts with a middle part of MCRS1 via its second Zn-binding domain (Figure 13). We also determined the crystal structure of the first Zn-binding domain of KANSL2 (Figure 13), which forms concentration dependent dimers. The full-length KANSL2 protein is difficult to express alone, but can be well co-expressed with MCRS1 in bacteria.

Based on these preliminary results we will now follow several directions. We will biophysically characterise the KANSL2/MCRS1 interaction and determine its crystal structure by X-ray crystallography. This will still require construct optimisation. The key interacting residues will be mutated and tested in rational functional assays in transgenic flies or mice where a wild type KANSL2 will be replaced with its mutated version using Cas9 technology (in collaboration with Akhtar lab). The targeting to chromatin will be analysed by ChIP and the profiles of target gene expression by quantitative RT-PCR analysis. If observations on selected genes look promising we will broaden the analysis by generating ChIP-seq and RNA-seq to obtain a genome-wide perspective. We will test the impact of these mutations on the assembly of the NSL complex using endogenous or *in-vitro* reconstituted complexes. We have already described structure-based mutations that specifically eliminate WDR5 (and KANSL3) or MOF from the complex (Dias et al., 2014; Kadlec et al., 2011). Antibodies against most NSL subunits are already available in the lab. As we can prepare protein fragments of all subunits, corresponding antibodies can easily be produced.

Interestingly, both the Zn-containing domains are conserved in the INO80D subunit of the INO80 chromatin remodeller that also contains MCRS1. We could demonstrate an equivalent interaction between INO80D and MCRS1. The minimal MCRS1/INO80D complex will be biophysically and structurally characterised. The structural information will be exploited in functional assays. This analysis will thus shed light on the role of these unusual Zn-binding domains and their interactions with MCRS1 in the chromatin targeting of both the NSL and INO80 complexes.

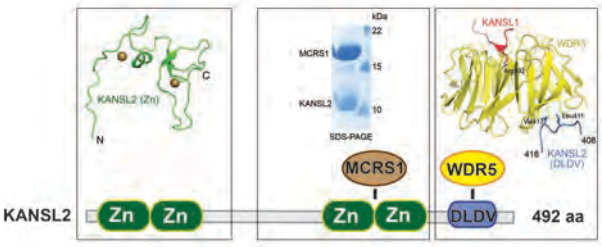


Figure 13. The domain structure of KANSL2. The N-terminal Zn domain structure is shown. The middle Zn finger domain directly interacts with MCRS1. The DLDV motif binds WDR5.

KANSL1 - scaffold of the complex

Predicted to be mostly intrinsically disordered and 120 kDa in size, KANSL1 is difficult to produce. Nevertheless, little amounts can be obtained from insect cells. To identify stable soluble fragments we used library based screening technique called **e**xpression of **s**oluble **p**roteins by **r**andom **i**ncrement **t**runcation (ESPRIT) in collaboration with Hart group at EMBL Grenoble. We screened over 50 000 clones of KANSL1 and eventually selected 50 most soluble constructs. Thus, we now have 20-30 kDa constructs covering its primary sequence with the exception of the coiled-coil region, which likely requires a binding partner to be soluble. We have recently shown that KANSL1 is the scaffold of the complex that interacts with the MOF, PHF20, WDR5 and MCRS1 subunits (Dias et al., 2014). We have demonstrated that KANSL1 interacts with WDR5 and MOF with short conserved motifs (Dias et al., 2014; Kadlec et al., 2011) (Figure 14). The predicted disorder and a sequence conservation indicate that also the remaining KANSL1 interactions might be mediated be linear motifs.

We will now verify the interaction of KANSL1 with MCRS1 and PHF20 and map the individual interacting regions using FLAG pull-down and co-expression experiments in insect cells that are already established for these proteins in the lab. Baculovirus constructs for all human and *Drosophila* NSL proteins as well as many deletion mutants have already been prepared. Once the interacting

regions of KANSL1 with MCRS1 and PHF20 are identified we will attempt to produce corresponding complexes in bacteria, which will be easier for the follow-up analysis. The interactions will be biophysically characterised using ITC, SEC-MALLS and analytical ultracentrifugation. We will determine the crystal structures of the minimal interacting regions that will be identified by limited proteolysis and mass spectrometry. The key interacting residues will be mutated and the impact of these mutations will be tested *in vivo* and *in vitro* as described above.

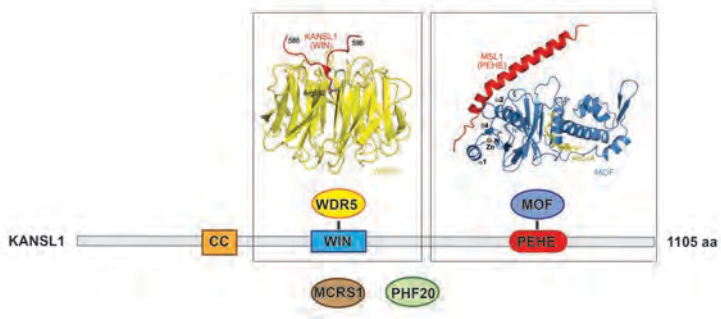


Figure 14. Schematic representation of the domain structure of KANSL1 and its interactions within the NSL complex. KANSL1 interacts with WDR5 and MOF via short conserved motifs. A structure of a complex of MOF with the PEHE domain of MSL1 (MSL complex) is shown, all the key interacting residues are conserved and important for the interactions in KANSL1 (see also Figure 6). CC-coiled coil; WIN- WDR5 interacting motif.

MCRS1

In its C-terminus MCRS1 contains an FHA domain, supposedly involved in binding of phospho-threonine containing peptides (Figure 15). While the N-terminal region of MCRS1 seems rather disordered, the middle region is predicted to be mostly helical. We can produce all three fragments of MCRS1 in bacteria and insect cell. We have initial pull-down data that MCRS1 interacts with KANSL2, PHF20, KANSL1 and MOF. We will define the regions involved in these interactions and eventual complexes will be subject to structural characterization. The interactions with MOF and PHF20 could only be observed with proteins produced in insect cell and it thus possible that some of them could depend on phosphorylation. Hypotheses based upon the structural work will be tested in functional assays as described above.

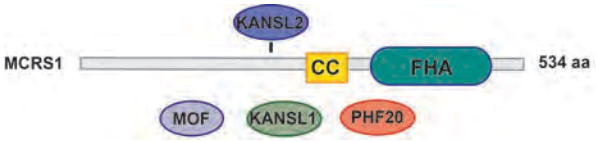


Figure 15. Schematic representation of the domain structure of MCRS1 and its interactions within the NSL complex. CC-coiled coil.

KANSL3 - additional enzymatic activity

KANSL3 is a large protein predicted to contain a hydrolase domain, which might possess another potential enzymatic activity of the complex. Indeed, when tethered to promoters, KANSL3 can stimulate transcription (Raja et al., 2010). Using bioinformatics, we could predict the likely limits of the domain, but so far we were unable to obtain soluble protein. We will now produce the full-length protein in insect cells, identify the limits of the hydrolase domain by limited proteolysis and determine its structure that might the most efficient to understand its possible enzymatic activity. We will also analyse the interactions of KANSL3 with other subunits.

Other NSL subunits

PHF20

PHF20 contains multiple small domains connected by disordered linkers (Figure 16). Similarly to KANSL1, we have performed the ESPRIT analysis to identify soluble fragments. We can now produce four constructs that span the entire sequence and can be expressed in both bacteria and insect cells. The structures of the tudor domains of PHF20, that recognise H3K4me or p53 K120me are known (Cui et al., 2012). It has been suggested that PHF20 functions in p53 regulation as a dimer (Cui et al., 2012). Using limited proteolysis on one of the ESPRIT derived fragments we have recently identified a short Zn-containing dimerization domain of PHF20 and determined its crystal structure (Figure 16). The residues essential for the dimerization will be mutated and the effect of these mutations on the dimerization of full-length PHF20 will be assessed. We will also test the effect of these mutations on the assembly and chromatin role of the NSL complex.

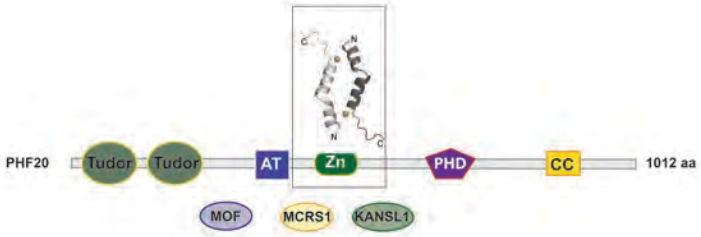


Figure 16. Schematic representation of the domain structure of PHF20 and its interactions within the NSL complex. The crystal structure of its Zn dependent dimerization domain is shown. CC-coiled coil; AT – AT hook.

WDR5 and MOF

We and others have already reported structures of WDR5 and the HAT and chromobarrel domains of MOF (Dias et al., 2014; Han et al., 2006; Kadlec et al., 2011; Nielsen et al., 2005). The KANSL1 and KANSL2 interactions seem to be the only interactions of WDR5 within the complex. We have already characterized the interaction of MOF with KANSL1(Kadlec et al., 2011). Since MOF seems to further interact with PHF20 and MCRS1 it will be very interesting to see whether these interactions are involved in the regulation of its histone acetyltransferase activity and specificity. The HAT activity of these sub-complexes will be tested in fluorescence based acetylation assays on histone tail peptides as well as on entire nucleosomes.

OGT and HCF1

Both OGT and HCF-1 are large but well structurally characterised proteins. Our current understanding is that if involved in the NSL complex function, rather than core members, these would be peripheral/recruited factors.

This objective includes multiple structural targets that are functionally interconnected. Thus, each novel structure greatly enhances our understanding of the complex and facilitates structural analysis of its remaining parts. There is always a risk associated with the need of well-diffracting crystal in such X-ray crystallography projects. Thus, in the case some complexes cannot be crystallized, we will attempt to identify interacting residues by the NMR chemical shift perturbation analysis (in collaboration with the Blackledge group, IBS), which together with the biochemical and biophysical analysis might be sufficient for the follow-up functional work.

2. Structural analysis of the full NSL complex and its interaction with nucleosome

NSL complex

The overall target of the structural analysis is the structure determination of the entire NSL complex. The combination of the previously obtained data on individual NSL sub-complexes and the results generated in Objective 1 will help us rationally produce and engineer the full NSL complex (Figure 17). Indeed, detailed biochemical and structural information on individual interactions, defined interacting regions and their binding affinities will provide very solid basis for this part of the project. Given the substantial proportion of flexible regions within the scaffold proteins (e.g. KANSL1), it is possible that the complex consists of several independently folded modules (e.g. chromatin binding or acetyltransferase) whose structure analysis will also be highly revealing. To be able to obtain high quality material for structural work we will have to overcome following issues:

Complex production and characterization:

Given that we can produce most of the human NSL subunits in bacteria, we will attempt to assemble multi-subunit NSL sub-complexes produced in *E.coli*. In parallel, the *Drosophila* NSL complex will be produced in insect cells using a combination of the MultiBac, polyprotein or co-infection strategies. These methods are well established in the lab. The prior data from Objective 1 will help us define which regions of individual subunits can/should be included to obtain soluble and stable proteins. We will also screen for additional higher order interactions including several subunits using insect cells co-expression and co-infection with the known sub-complexes.

Once sufficient amount of homogenous material is obtained it will be used in multi-angle laser light scattering, analytical ultracentrifugation and native mass spectrometry to analyse the stoichiometry of the complex. Mass spectrometry analysis combined with chemical crosslinking will be used to reveal additional interactions among the NSL subunits.

Complex flexibility:

The subunits will be carefully engineered to reduce their flexibility. Many of the disordered regions do not have the same length and sequence in different species, and it might simply function as flexible linkers. In some species, certain proteins are “compressed” lacking most of the problematic regions. E.g. in *Xenopus laevis* KANSL1, many of the disordered linkers are much shorter. We will systematically delete some of these linker regions and evaluate the impact of such modifications on the NSL complex assembly *in vivo* and *in vitro*. This is a common approach in structural analysis and is sometimes inevitable to obtain high quality protein material. It might also be necessary to use different species such as *Xenopus* to obtain a less flexible complex.

The structure of the resulting complex will be analysed by X-ray crystallography and/or cryo-electron microscopy (EM). Thanks to the recent developments in EM software and detectors it is now possible to obtain reasonable high-resolution structures of large complexes. In that case, the high-resolution structures of the individual sub-complexes will be very useful in generation of the final model revealing the overall architecture of the complex. The IBS has all the necessary equipment and expertise for the EM analysis.

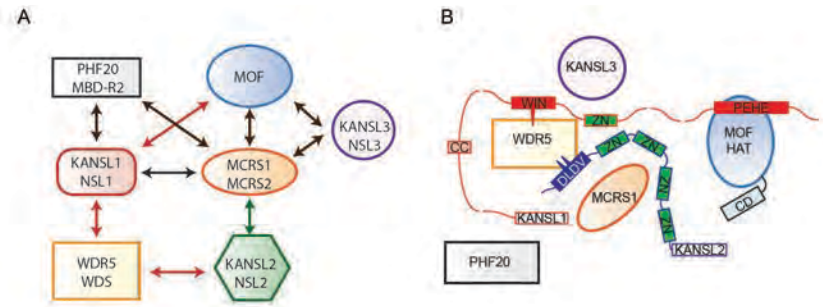


Figure 17. (A) Preliminary interaction map of the NSL complex. Red arrows indicate interactions that we already characterized. The green arrow indicates the interaction that is well defined and is being structurally analysed (B). Our current model of the NSL complex architecture. The interactions of KANSL3 and PHF20 are not yet defined.

The NSL "supercomplexes"

The long-term goal is the analysis of the interaction between the NSL complex and the nucleosome. The interaction will probably require specific histone modifications, which still need to be characterized. We will analyse the interaction of the NSL complex and its modules with the modified histone tail peptide arrays. We will assemble nucleosomes of histones carrying modifications known or susceptible to be bound by the NSL complex. Specific chemical modifications will be introduced to histones using the Expressed Protein Ligation technology. The interactions with the nucleosome might be an efficient way to stabilize such a flexible complex. The eventual "supercomplexes" will be biochemically and biophysically characterised and used for structural analysis as described above. Hypotheses derived from our structural work will then be tested in functional experiments.

Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex

Jorge Dias,[1,2,5] Nhuong Van Nguyen,[3,4,5] Plamen Georgiev,[3] Aline Gaub,[3,4] Janine Brettschneider,[1,2] Stephen Cusack,[1,2] Jan Kadlec,[1,2,6] and Asifa Akhtar[3,6]

[1]Grenoble Outstation, European Molecular Biology Laboratory (EMBL), 38042 Grenoble, France; [2]Unit for Virus Host–Cell Interactions, University Grenoble Alpes-EMBL-CNRS, 38042 Grenoble, France; [3]Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg im Breisgau, Germany; [4]Faculty of Biology, University of Freiburg, 79104 Freiburg im Breisgau, Germany

The subunits of the nonspecific lethal (NSL) complex, which include the histone acetyltransferase MOF (males absent on the first), play important roles in various cellular functions, including transcription regulation and stem cell identity maintenance and reprogramming, and are frequently misregulated in disease. Here, we provide the first biochemical and structural insights into the molecular architecture of this large multiprotein assembly. We identified several direct interactions within the complex and show that KANSL1 acts as a scaffold protein interacting with four other subunits, including WDR5, which in turn binds KANSL2. Structural analysis of the KANSL1/WDR5/KANSL2 subcomplex reveals how WDR5 is recruited into the NSL complex via conserved linear motifs of KANSL1 and KANSL2. Using structure-based KANSL1 mutants in transgenic flies, we show that the KANSL1–WDR5 interaction is required for proper assembly, efficient recruitment of the NSL complex to target promoters, and fly viability. Our data clearly show that the interactions of WDR5 with the MOF-containing NSL complex and MLL/COMPASS histone methyltransferase complexes are mutually exclusive. We propose that rather than being a shared subunit, WDR5 plays an important role in assembling distinct histone-modifying complexes with different epigenetic regulatory roles.

[*Keywords*: chromatin; epigenetics; histone acetylation]

Supplemental material is available for this article.

Received March 12, 2014; revised version accepted March 28, 2014.

Eukaryotic transcription is controlled by a complex interplay between transcription regulators and local chromatin organization. Various post-translational modifications of histones, such as acetylation or methylation, help modify chromatin structure and are important for recruitment of effector proteins such as transcription factors or chromatin remodelers. MOF (males absent on the first), a member of the MYST family of histone acetyltransferases (HATs), is responsible for a large fraction of histone H4 Lys16 acetylation (H4K16ac) in mammalian cells (Taipale et al. 2005). Similar to other chromatin regulators, MOF does not exert its function alone but within multiprotein complexes that regulate its activity and specificity. For a long time, MOF has been known to be a component of the dosage compensation complex (DCC; or male-specific lethal [MSL] complex), which plays a key role in dosage compensation in *Drosophila* males (Conrad and Akhtar 2011). Recently, MOF was also shown to exist within the NSL (nonspecific lethal) complex involved in global transcription regulation (Raja et al. 2010).

The NSL complex is an evolutionarily conserved multiprotein assembly consisting of at least MYST1 (KAT8)/MOF, KANSL1/NSL1, KANSL2/NSL2, KANSL3/NSL3, WDR5/WDS, MCRS1/MCRS2, and PHF20/MBD-R2 proteins in mammals and *Drosophila*, respectively (Mendjan et al. 2006; Cai et al. 2010; Raja et al. 2010). NSL1, NSL2,

[5]These authors contributed equally to this work.
[6]Corresponding authors
E-mail akhtar@immunbio.mpg.de
E-mail kadlec@embl.fr
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.240200.114.

NSL3, and MCRS2 were shown to be essential for *Drosophila* viability (Mendjan et al. 2006). In *Drosophila*, the complex associates with promoters of >4000 genes, and loss of the NSL complex binding severely affects their expression levels (Raja et al. 2010). The majority of the NSL complex-bound targets were shown to belong to housekeeping genes, and the NSL complex was shown to be required for efficient recruitment of RNA polymerase II at their target promoters (Feller et al. 2012; Lam et al. 2012).

In mammals, loss of MOF leads to early embryonic lethality (Thomas et al. 2008), while PHF20 knockout mice die just after birth (Badeaux et al. 2012). MOF and H4K16ac have been shown to regulate autophagy (Fullgrabe et al. 2013). In humans, haploinsufficiency of KANSL1 causes a multisystem disorder called the 17q21.31 microdeletion syndrome, characterized by intellectual disability (Koolen et al. 2012; Zollino et al. 2012). These observations point to an essential but also a more complex function of these proteins in higher eukaryotes. Consistently, MOF was shown to catalyze p53 Lys120 acetylation, which is required for optimal transcription activation of p53 target genes (Sykes et al. 2006; Li et al. 2009), while the tudor domain of PHF20 associates with methylated Lys370 and Lys382 of p53 (Cui et al. 2012). Both MOF and WDR5 are important regulators of the embryonic stem cell core transcription network (Ang et al. 2011; Li et al. 2012; Taylor et al. 2013), and PHF20 is required for somatic cell reprogramming (Zhao et al. 2013a). Despite the emerging global importance of the NSL complex in transcription regulation, the details and specificity of its recruitment to chromatin as well as its mode of action are currently poorly understood. Similarly, the molecular architecture of the complex and individual protein–protein interactions among the subunits remain unknown.

The human KANSL1 subunit consists of 1105 amino acid residues. It is predicted to be mostly unstructured; however, its C terminus contains the PEHE domain, which interacts with the HAT domain of MOF (Kadlec et al. 2011). KANSL2 is not well characterized. In humans, it consists of 492 residues and contains four putative Zn-coordinating motifs possibly involved in DNA binding. WDR5 (WDS in *Drosophila*) is also an established subunit of the human MLL/COMPASS histone H3 Lys4 (H3K4) methyltransferase complexes (Shilatifard 2012). It contains a seven-bladed β-propeller domain that was first shown to interact with an H3R2 peptide (Couture et al. 2006; Han et al. 2006; Ruthenburg et al. 2006; Schuetz et al. 2006). Within the MLL complexes, WDR5 recognizes both a WIN (WDR5-interacting) motif of MLL proteins and a short motif of RbBP5. These interactions stimulate the methyltransferase activity of MLL proteins (Patel et al. 2008b; Odho et al. 2010; Avdic et al. 2011). The NSL and MLL/COMPASS complexes were proposed to form a larger assembly that would possess both H3K4 methylation and H4 acetylation activities (Dou et al. 2005; Li et al. 2009).

In this study, we focus on the role of KANSL1/NSL1 and WDR5/WDS within the NSL complex. We show that KANSL1/NSL1 is a scaffold of the complex. We identified two previously unknown interactions that the WDR5/WDS subunit makes with KANSL1 and KANSL2, mapped the interacting regions, and determined crystal structures of these subcomplexes. We show that the NSL1–WDS interaction is important for efficient targeting of the NSL complex to promoters of target genes in *Drosophila* and that loss of this interaction results in male and female lethality. Our study provides the first insights into the molecular interactions within the NSL complex and reveals a mutually exclusive division of labor for WDR5/WDS between the NSL and MLL/COMPASS complexes.

Results

NSL1 interacts with multiple subunits of the NSL complex

The acetyltransferase MOF/KAT8/MYST1 is recruited into the NSL complex via its interaction with the C-terminal PEHE domain of NSL1, analogous to the interaction between MOF and MSL1 in the DCC or the MSL complex (Kadlec et al. 2011). Both NSL1 and MSL1 are predicted to be mostly intrinsically disordered, and we demonstrated previously that within the MSL complex, MSL1 functions as a scaffold protein interacting independently with MSL2, MSL3, and MOF (Kadlec et al. 2011; Hallacli et al. 2012). To test whether NSL1 has an analogous function within the NSL complex, we expressed individual subunits of the *Drosophila* NSL complex in insect cells and performed pull-down experiments using 3xFlag-tagged NSL1 as the bait. While MOF, WDS, MCRS2, and MBD-R2 copurified with NSL1 (Fig. 1A, lanes 4–7; Supplemental Fig. S1), no binding was observed for NSL2 and NSL3. None of the NSL subunits copurified with a 3xFlag-tagged GFP control, confirming specificity of the interaction (data not shown). In the inverse experiment, N-terminal or C-terminal 3xFlag fusions of MOF, MBD-R2, MCRS2, and WDS copurified untagged NSL1 (Fig. 1B, lanes 1,6–11; Supplemental Fig. S1B). While the interactions with MOF and MCRS2 have already been described (Raja et al. 2010; Kadlec et al. 2011), the additional direct binding of NSL1 to WDS and MBD-R2 was not known and indicates that NSL1 might function as the scaffold subunit of the NSL complex. Since NSL1 is predicted to contain no known globular domains and to be predominantly intrinsically disordered, it is likely to interact with its binding partners via short linear motifs, similar to MSL1. Indeed, in humans, the interaction with MOF requires only a short motif spanning residues 883–952 (Kadlec et al. 2011).

Whereas the role of WDR5 within the MLL/COMPASS methyltransferase complex is well characterized (Patel et al. 2008a; Song and Kingston 2008; Odho et al. 2010; Avdic et al. 2011), nothing is known about the function of WDR5/WDS within the NSL or other chromatin-modifying complexes. Interestingly, it has been proposed that the WDR5 subunit might serve as a physical link between MOF and MLL/COMPASS complexes, forming a complex that would possess both H3K4 methylation and H4 acetylation activities to promote transcription initiation at target genes (Dou et al. 2005; Li et al. 2009). To better understand

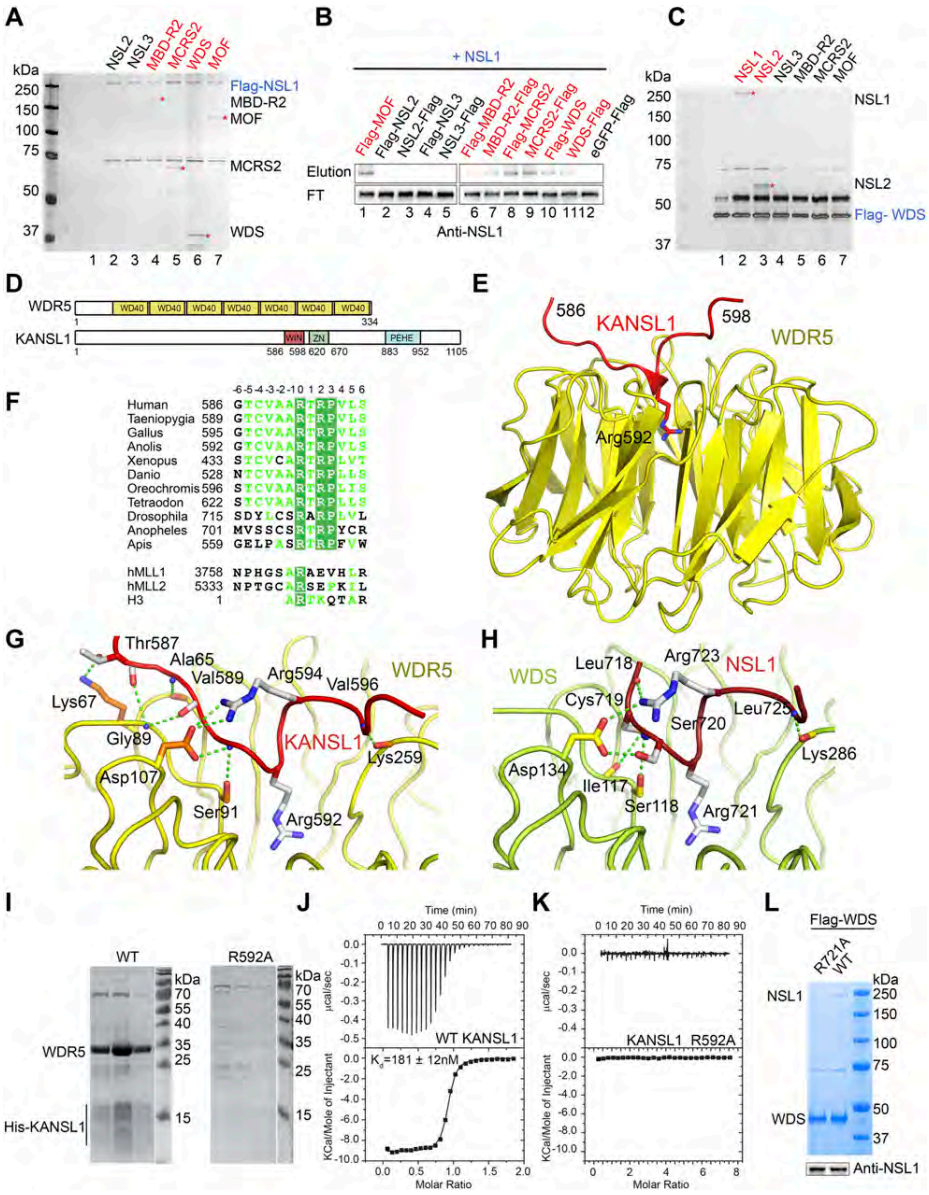


Figure 1. Structural analysis of the KANSL1–WDR5 interaction. (*A*) SDS-PAGE and silver staining analysis of Flag pull-down assays using N-terminally 3xFlag-NSL1 as the bait and other untagged NSLs as prey. All proteins are from baculovirus-expressed protein extracts. Interacting proteins are indicated by asterisks and highlighted in red. (See Supplemental Fig. S1A for the input sample). (*B*) Western blot analysis of the reverse Flag pull-down assays of *A* using untagged NSL1 as the prey and either N-terminal or C-terminal 3xFlag-tagged NSLs as bait. (FT) Flowthrough sample. Interacting proteins are highlighted in red. The eluted prey proteins are shown in Supplemental Figure S1B. (*C*) SDS-PAGE and silver staining analysis of Flag pull-down assays using 3xFlag-WDS as the bait and other untagged NSLs as prey. All proteins are from baculovirus-expressed protein extracts. Interacting proteins are indicated by asterisks and highlighted in red. (See Supplemental Fig. S1C for the input sample). (*D*) Schematic representation of the domain structure of human WDR5 and KANSL1. KANSL1 contains a PEHE domain involved in the binding of MOF (Kadlec et al. 2011) and a WIN motif identified in this study. (*E*) Ribbon diagram of the minimal human WDR5–KANSL1 complex structure. $WDR5_{23–334}$ is shown in yellow, and KANSL1 is in red. The key interacting KANSL1 Arg592 is shown as sticks. (*F*) Sequence alignment of NSL1 proteins. Only the sequence of the fragment involved in the interaction with WDR5 is shown. Identical residues are in green boxes. Numbering of the motif residues centered on Arg592 (in humans) is shown *above* the alignment. Representative sequences of corresponding MLL and H3 motifs are aligned with NSL1. (*G*) Details of the interaction between the β-propeller domain of WDR5 and the WIN motif of KANSL1. Interacting WDR5 residues are shown in orange, and KANSL1 residues are in gray. Residues without side chains represent main chain interactions. Contacts with Arg592 of KANSL1 are similar to the ones of H3 and MLL proteins and are not shown for clarity (see Supplemental Fig. S3B). (*H*) Interaction interface of the complex between the β-propeller domain of WDS (green) and the WIN motif of NSL1 (red). (*I*) SDS-PAGE analysis of the binding of His-tagged $KANSL1_{584–690}$ (wild type [WT] and the R592A mutant) to coexpressed untagged $WDR5_{23–334}$ after purification using Ni^{2+} resin. $KANSL1_{584–690}$ is only detectable (but still degrading) when bound to WDR5. (*J*) ITC measurement of the interaction between $WDR5_{23–334}$ and the KANSL1 WIN motif-containing peptide (585-DGTCVAA*R*TRPVLS-598-Y). The *bottom* panel represents a fit of the calorimetric data to single-site-binding model. Dissociation constant (K_d) derived from the fit is indicated. (*K*) ITC measurement of the interaction between $WDR5_{23–334}$ and the mutated KANSL1 WIN motif-containing peptide (585-DGTCVAA*A*TRPVLS-598-Y). (*L*, *top* panel) 3xFlag-WDS was coexpressed with either wild-type NSL1 (WT) or mutant NSL1 (R721A) in a single baculovirus and purified using Flag-M2 resin. (*Bottom* panel) The Western blot shows that equal amounts of wild-type and mutant NSL1 (R721A) were present in the input extracts, yet only the wild-type NSL1 interaction is detected with WDS in the Coomassie blue gel.

the details of the function and the recruitment of WDR5/WDS into the NSL complex, we first assessed how WDS interacts with other subunits. Using pull-down assays with 3xFlag-tagged WDS and untagged NSL subunits, we could show that in addition to NSL1, WDS also directly interacts with NSL2, while it does not interact with MOF (Fig. 1C, lanes 2,3,7; Supplemental Fig. S1C). These results indicate that the interaction between MOF and WDS is mediated by NSL1.

WDR5/WDS–KANSL1 structure

The mapping of the WDS- and NSL1-interacting domains was performed using their human orthologs, as they could be more easily produced in bacteria. To identify the region involved in the interaction with WDR5, KANSL1, predicted to be mostly intrinsically disordered, was divided into four ~25-kDa segments, and their interactions with His-tagged WDR5 (residues 23–334) were tested in a coexpression experiment. When expressed alone, all four fragments were unstable/degraded (data not shown). However, when coexpressed with WDR5, the KANSL1 fragment spanning residues 537–773 could be efficiently copurified (Supplemental Fig. S2A, lane 3), indicating that it possesses a WDR5-binding region. This fragment contains several stretches of well-conserved residues and four highly conserved cysteine and histidine residues, possibly forming a Zn-coordinating structure. When loaded onto an S75 gel filtration columns, the WDR5 and KANSL1$_{537–773}$ coeluted as a complex in the same elution volume, distinct from that of WDR5 alone (Supplemental Fig. S2B,C). KANSL1, however, suffered from significant degradation, consistent with its predicted unstructured nature (Supplemental Fig. S2C). Using mass spectrometry analysis, we mapped the minimal WDR5-binding fragment to a region encompassing a conserved sequence around Arg592. We then determined by X-ray crystallography the structure of the WDR5 β-propeller domain (residues 23–334) in complex with this KANSL1 fragment spanning residues 585–598 at a resolution of 1.5Å (Fig. 1D–F). Interestingly, the structure revealed that the KANSL1 motif centered on Arg592 interacts with WDR5 in a manner similar to MLL proteins and histone H3. Arg592 inserts into the central pore of the β propeller of WDR5, forming hydrogen bonds and stacking interactions with Phe133, Phe263, Ser91, and Cys261 (Fig. 1E; Supplemental Fig. S3A,B). Compared with the known WDR5 structures, Thr587 and Val589 make additional hydrogen bonds with Lys67, Ala65, and Gly89, and the highly conserved Arg594 makes a salt bridge with Asp107 (Fig. 1F,G; Supplemental Fig. S3A). Downstream from Arg592, Val596 interacts with a shallow hydrophobic surface lined by Tyr191, Pro216, and Leu234 in a way similar to MLL2 or MLL3 (Supplemental Fig. S3C).

We also identified a corresponding WDS-interacting motif in *Drosophila* NSL1 that spans residues 714–729 and solved its crystal structure in complex with WDS. Compared with the human structure, the residues in position −4 to −7 relative to the central arginine are not ordered (Fig. 1F,H; Supplemental Fig. S3D). The unusual feature of the NSL1 WIN motif is a serine residue (Ser720) in the −1 position that is invariantly an alanine in all known WIN motif structures (Fig. 1H). This serine, however, fits well into the interaction interface and makes an additional hydrogen bond with Asp134 (Fig. 1H). While Arg592, Arg594, and Pro595 of KANSL1 are well conserved among species, residues mediating additional contacts in these two structures are conserved only among vertebrates or insects, respectively (Fig. 1F).

His-tagged KANSL1$_{584–690}$ could be copurified and stabilized with untagged WDR5 in a coexpression experiment (Fig. 1I). However, when expressed on its own, the fragment became unstable and degraded. To assess the importance of Arg592 for the interaction with WDR5, we mutated this residue to alanine. The R592A mutant could not engage in a stabilizing interaction with WDR5 and was therefore degraded (Fig. 1I). To confirm the role of Arg592 in the interaction, we used isothermal titration calorimetry (ITC). We observed that WDR5 binds the KANSL1 peptide (585–598) with a dissociation constant (K_d) of 180 nM (Fig. 1J). As expected, no binding was observed for the KANSL1 peptide containing the R592A mutation (Fig. 1K). The NSL1–WDR5 dissociation constant is in the same range as the reported values for MLL WIN motifs (Dharmarajan et al. 2012). In addition, we also tested the corresponding R721A mutation within *Drosophila* NSL1, which significantly reduced the interaction between full-length WDS and NSL1 (Fig. 1L).

WDR5/WDS–KANSL2 structure

Within the MLL complexes, the β-propeller domain of WDR5 interacts with the MLL proteins and RbBP5 using two separate interaction surfaces (Odho et al. 2010; Avdic et al. 2011). Having shown that within the NSL complex, WDR5 interacts with both KANSL1 and KANSL2, we wondered whether a similar situation arose (Fig. 1C). To map the WDR5-binding region of KANSL2, we prepared several truncation constructs of KANSL2. While constructs covering the putative Zn-coordinating domains could not be expressed in bacteria (data not shown), a C-terminal construct spanning residues 381–492 yielded soluble protein. When loaded onto an S200 gel filtration column, KANSL2$_{381–492}$ coeluted in complex with WDR5, while WDR5 alone eluted in a separate peak (Fig. 2A; Supplemental Fig. S4), indicating that KANSL2 interacts with WDR5 with its C terminus. This KANSL2 fragment as well as its truncated versions suffered from degradation and resisted crystallization. Eventually, a shorter KANSL2 fragment (residues 406–417) could be cocrystallized with the WDR5/KANSL1 subcomplex, and its structure could be solved at a resolution of 2 Å (Fig. 2B–E). Interestingly, the interaction between KANSL2 and WDR5 indeed resembles that of WDR5 with RbBP5 (Odho et al. 2010; Avdic et al. 2011). KANSL2 binds on the opposite side of the WDR5 β-propeller domain from KANSL1 (Fig. 2B). Three hydrophobic residues (Leu411, Val413, and Val414) of KANSL2 point into a large hydrophobic cavity in a cleft between blades 5 and 6 of the WDR5 domain (Fig. 2D). These interactions are stabilized by several hydrogen bonds that KANSL2 forms with Asn225, Lys250, and

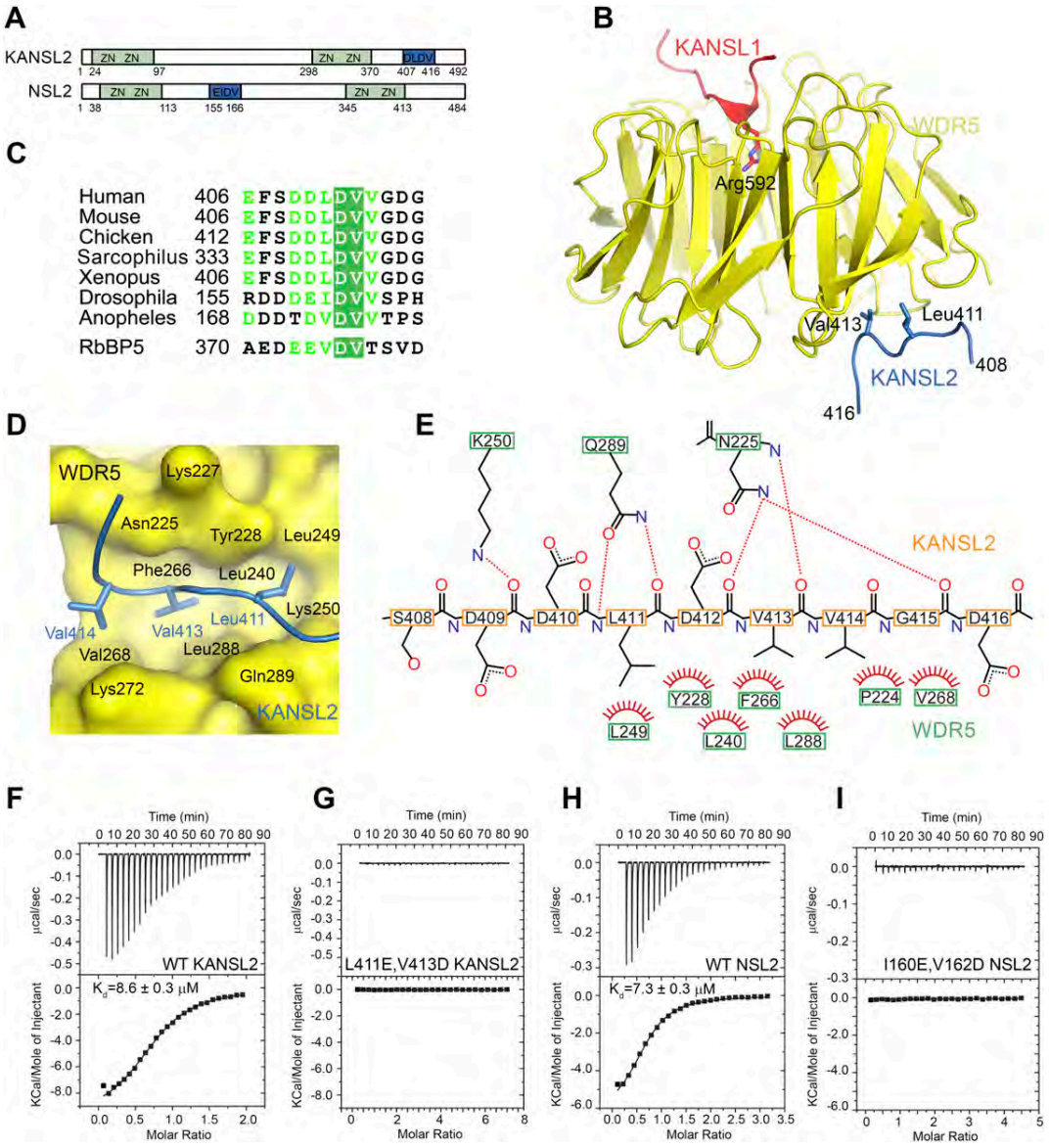


Figure 2. Structural basis for the WDR5–KANSL2 interaction. (*A*) Schematic representation of the domain structure of human and *Drosophila* KANSL2/NSL2. KANSL2 contains four putative Zn-coordinating motifs and an additional WDR5-binding motif (shown in blue) identified in this study. (*B*) Ribbon diagram of the minimal human KANSL1–WDR5–KANSL2 complex structure. $WDR5_{23–334}$ is shown in yellow, KANSL1 is in red, and KANSL2 is in blue. The key interacting residues are shown as sticks. (*C*) Sequence alignment of NSL2 proteins. Only the sequence WDR5-binding fragment is shown. The corresponding sequence of the RbBP5- and WDR5-binding motif is aligned with KANSL2. (*D*) Details of the WDR5–KANSL2 interaction interface. KANSL2 Leu411, Val413, and Val414 insert into a wide hydrophobic pocket on WDR5. (*E*) Schematic representation of the interactions between KANSL2 and WDR5. Hydrogen bonds are denoted with dotted lines. Hydrogens are not shown for clarity. (*F*,*G*) ITC measurement of the interaction between $WDR5_{23–334}$ and KANSL2 WDR5-binding peptide (Y-406-EFSDD*LDV*VGDG-417) (*F*) and its mutated version (Y-406-EFSDD*EDD*VGDG-417) (*G*). (*H*,*I*) ITC measurement of the interaction between $WDS_{50–361}$ and the NSL2 WDS-binding peptide (Y-155-RDDDE*IDV*VSPH-166) (*H*) and its I160E/V162D mutant (Y-155-RDDDE*EDD*VSPH-166) (*I*).

Gnl289 (Fig. 2E). Similar to RbBP5, the central conserved Asp412 stabilizes the conformation of the peptide by multiple hydrogen bonds with +2 and +3 residues (Val414 and Gly415). To confirm the importance of this motif for the interaction between KANSL2 and WDR5, we mutated Val413 within $KANSL2_{381–492}$ to aspartate and showed that the corresponding protein no longer coeluted with WDR5 on the S200 gel filtration column (Supplemental Fig. S5A). Using ITC, we showed that WDR5 binds the KANSL2 peptide (406–417) with a K_d of 8.6 μM (Fig. 2F), which is comparable with the WDR5–RbBP5 interaction (Odho et al. 2010; Avdic et al. 2011). No binding was

observed for an L411E, V413D KANSL2 mutant even at four times higher ligand concentration (Fig. 2G). While this motif is very well conserved in vertebrate KANSL2 proteins (Fig. 2C), it is absent in the C terminus of *Drosophila* or *Anopheles*. Surprisingly, in the *Drosophila* NSL2 sequence, a similar motif is located between the two sets of the Zn-coordinating motifs (Fig. 2A,C). The corresponding peptide (residues 155–166) interacted with WDS with a K_d of 7.3 μM, while no binding was detectable for an I160E, V162D NSL2 double mutant (Fig. 2H,I). We also determined the crystal structure of the *Drosophila* NSL1/WDS/NSL2 minimal complex, which showed the WDS/NSL2 interaction to be virtually identical to that in the human complex (Supplemental Fig. S5B,C).

Overall, these biochemical and structural analyses revealed that KANSL1 and KANSL2 subunits contain WDR5-interacting motifs that are similar to the motifs found in the MLL proteins and RbBP5. These interactions are mediated by the same WDR5 residues, indicating that the KANSL1 and MLL as well as KANSL2 and RbBP5 interactions with WDR5 are mutually exclusive (Fig. 3A,B). To further confirm the mutually exclusive nature of the NSL1/MLL binding to WDR5, we performed ITC measurements of the interaction of WDR5 with the MLL4 WIN motif in the absence or presence of saturating amounts of KANSL1. While MLL4 normally interacts with WDR5 with a K_d of 32 nM, no additional binding was observed when WDR5 was saturated with KANSL1 (Supplemental Fig. S6A,B). WDR5 also uses the same surface to interact with the histone H3 N terminus (H3R2). While the H3R2 peptide interacts with WDR5 with a K_d of 46 μM (comparable with the reported value) (Schuetz et al. 2006), no additional binding could be detected in the presence of the KANSL1 WIN motif (Supplemental Fig. S6C,D). These data clearly show that WDR5 cannot interact with KANSL1, MLL, or H3 at the same time.

The NSL complex assembly in vivo

To study the impact of NSL1 mutations on the incorporation of WDS into the NSL complex in vivo, we first transiently expressed 3xFlag-tagged wild-type or mutated (R721A) NSL1 proteins in *Drosophila* SL-2 cells and immunoprecipitated the corresponding NSL complexes using anti-Flag resin (Fig. 3C). While the wild-type full-length NSL1 protein could copurify all NSL subunits, the R721A mutant failed to interact with WDS, whereas the interaction of MOF, MBD-R2, MCRS2, and NSL2 remained unaffected in vivo. Interestingly, R721A also displayed reduced interaction with NSL3, indicating that the NSL1–WDS interaction plays an important role in the integration of this protein into the NSL complex.

To better understand the role of the NSL2–WDS interaction within the NSL complex, we prepared SL-2 cell lines stably expressing 3xFlag-tagged wild-type or mutated (I160E, V162D) NSL2 (Fig. 3D). In contrast to the NSL1 mutant, both wild-type and mutated NSL2 efficiently copurified all NSL subunits but not MSL3 or Trx/MLL, indicating that the NSL2–WDS interaction is not absolutely required for WDS or NSL2 to remain in the complex. The fact that Trx/MLL does not copurify with NSL2 also provides additional evidence for the separate roles of WDS/WDR5 within the NSL and MLL complexes. To test whether another subunit is involved in the recruitment of NSL2, we performed pull-down assays of untagged NSL2 with other subunits fused with a 3xFlag produced in insect cells and showed that in addition to WDS, NSL2 can also directly interact with MCRS2 (Fig. 3E, lanes 8,9; Supplemental Fig. S1D). Since MCRS2 also interacts with NSL1 (Fig. 1A), it is possible that NSL2 is tethered to NSL1 via MCRS2. Thus, the NSL2–WDS interaction seems to have rather a stabilizing role within the complex, being in agreement with its K_d of 8.6 μM, which is relatively weak compared with the 180 nM K_d for the KANSL1–WDR5 interaction.

The NSL1–WDS interaction is important for viability in flies

In order to further assess the functional relevance of the NSL1–WDS interaction in *Drosophila* in vivo, we generated transgenic flies expressing either the wild-type *nsl1* or a mutant *nsl1* variant carrying a single point mutation (R721A) (Fig. 3F) that strongly reduces the interaction of NSL1 with WDS (Figs. 1I–L, 3C). The two transgenes were inserted in the same genomic location (*VK33 attP*-docking site at the polytene location 65B2 on the left arm of chromosome 3) by φC31 integrase-mediated transformation to minimize positional effects upon phenotypic comparison (Groth et al. 2004; Venken et al. 2006). Both transgenes carried the Gal4 recognition site UAS upstream of the *nsl1* ORF to allow ectopic expression in a spatiotemporally regulated manner (Brand and Perrimon 1993). In addition, a 3xFlag tag sequence was fused to the N terminus to allow easy detection and purification. Strong and ubiquitous ectopic expression of the two transgenes in a wild-type background using the *αTub84-Gal4* driver (Lee and Luo 1999) did not affect the viability of the flies (data not shown). To assess the functionality of the ectopically expressed proteins, we used another strong and ubiquitous driver, *Act5C-Gal4*, to induce expression in the absence of endogenous NSL1 ($nsl1^{S009413}/nsl1^{j2E5}$-null mutant background) (Fig. 3G). Lack of NSL1 causes 100% lethality in both males and females (Mendjan et al. 2006; Yu et al. 2010). As expected, the ectopically expressed wild-type NSL1 rescued the lethality of the loss-of-function mutants in both males and females (Fig. 3H). Interestingly, ectopic expression of $nsl1^{R721A}$ in the null mutant background could only partially rescue the lethality in males (~40% survival). However, $nsl1^{R721A}$ caused complete lethality in females (0% of the female animals reached the adult stage) (Fig. 3H). These results clearly showed that the Arg721-mediated interaction between NSL1 and WDS identified in our structural analysis is important for fly viability. Partial rescue in males suggests an attractive possibility that there may be a compensatory function provided by the MOF-containing MSL complex in males, which would be a very interesting avenue to explore in the future.

We next studied the mechanism underlying the strong phenotype observed in the $nsl1^{R721A}$ mutants. We first

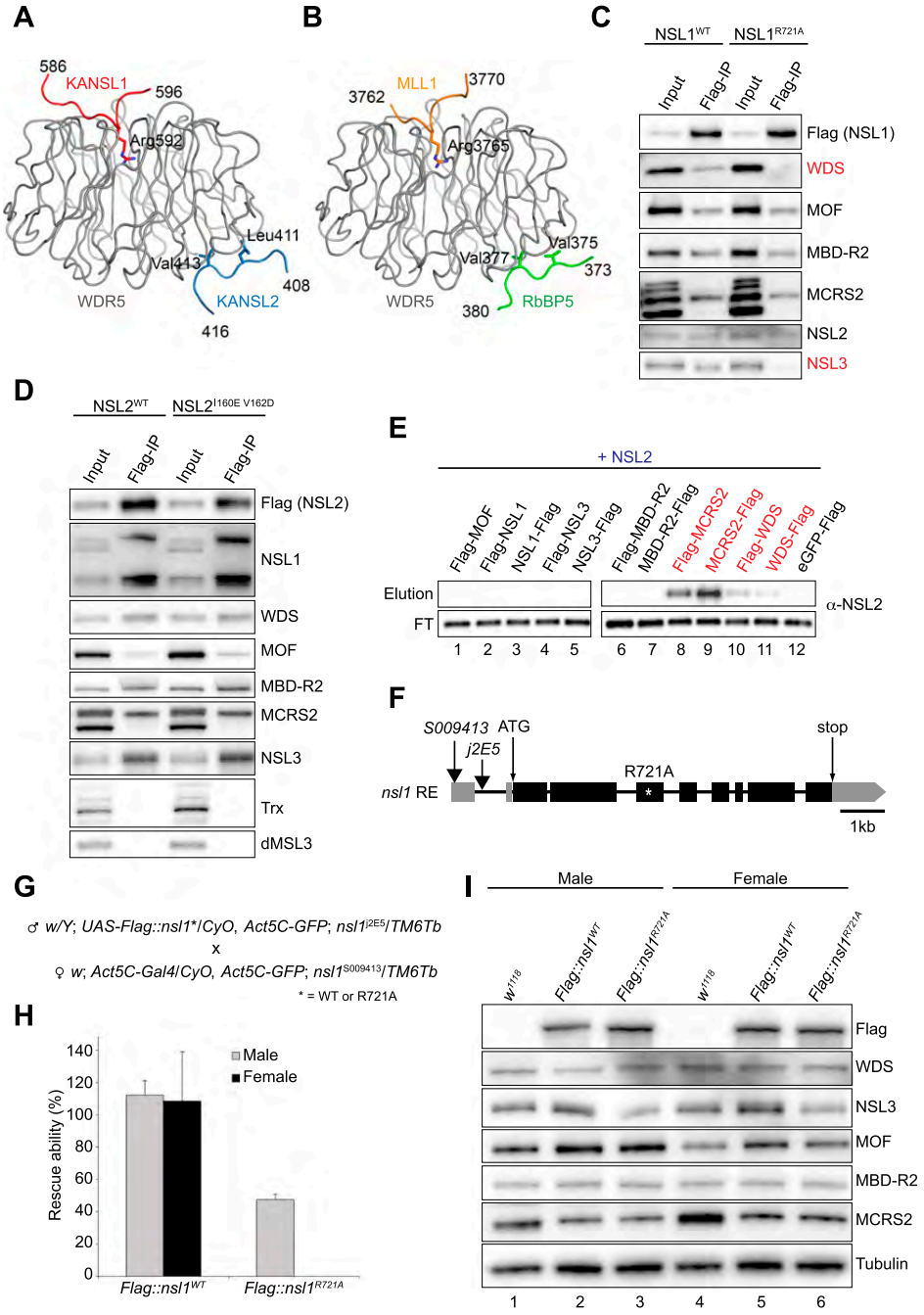


Figure 3. The in vivo function of the NSL1–WDS–NSL2 interactions. (*A*,*B*) Comparison of the crystal structures of human ternary KANSL1–WDR5–KANSL2 (*A*) and MLL1–WDR5–RbBP5 (Protein Data Bank [PDB] code 3P4F) (*B*) complexes. The two complexes were superimposed using WDR5. (*C*) Flag immunoprecipitation of wild-type (WT) NSL1 and NSL1^{R721A} in SL-2 cells. Both wild-type *nsl1* and *nsl1*^{R721A} were transiently transfected into SL-2 cells, and immunoprecipitation was performed using Flag-M2 resin. Antibodies used for Western blot analysis are indicated. (*D*) Flag immunoprecipitation of wild-type (WT) NSL2 and NSL2$^{I160E–V162D}$ in stably expressing SL-2 cells. Both wild-type *nsl2* and *nsl2*$^{I160E–V162D}$ were transfected into SL-2 cells, and stably expressing cells were selected. Immunoprecipitation was performed using Flag-M2 resin. Antibodies used for Western blot analysis are indicated. (*E*) Western blot analysis of Flag pull-down assays using untagged NSL2 as the prey and either N-terminal or C-terminal 3xFlag-tagged NSLs as bait. (FT) Flowthrough sample. Interacting proteins are highlighted in red. The eluted prey proteins are shown in Supplemental Figure S1D. (*F*) A schematic representation of the structure of the RE transcript, encoded by *nsl1*/CG4699, used in this study. Exons (black bars), introns (black lines), and untranslated regions (UTRs) (gray bars) are shown. The ATG start and the STOP codon are indicated as well as the positions of the transposon insertions (black arrows). The site of the introduced mutation disrupting the NSL1 binding to WDS is marked by a white asterisk. The protein alteration is given at the *top*. (*G*) Details of the crosses used in the rescue experiments. (*H*) Relative percentage of adult male (gray bars) and adult female (black bars) viability upon ectopic expression of wild-type *nsl1* and mutant *nsl1*^{R721A} in the absence of endogenous NSL1 at 25°C. The non-*CyO; Tb* siblings in which endogenous *nsl1* is expressed were used as internal controls and scored as 100% viable. The detail of the fly crosses is given in *G*. The error bars represent standard deviations of three independent crosses. (*I*) Western blot analysis of proteins extracted from third instar male and female larvae expressing either wild-type or mutant (R721A) NSL1 in the absence of endogenous NSL1. w^{1118} larvae were used as a wild-type control. Antibodies used for the Western blot analysis are indicated. Tubulin was used as a loading control.

examined the total protein levels of NSLs in rescued larvae from the crosses in Figure 3G. These larvae all had a *nsl1*-null background (−/−) and carried either *Act5C-Gal4*-driven wild-type *Flag::nsl1* (*Act5C>Flag::nsl1*^{WT}) or *Flag::nsl1*^{R721A} (*Act5C>Flag::nsl1*^{R721A}). The Western blot results showed that the expression levels of *Flag::nsl1* transgenes were relatively equal between *nsl1*^{WT} (*Act5C>Flag::nsl1*^{WT};−/−) and *nsl*^{R721A} (*Act5C>Flag::nsl1*^{R721A};−/−) animals as well as between males and females (Fig. 3I, anti-Flag row), suggesting that both transgenes were well expressed, and the lethal phenotype observed in the *nsl1*^{R721A} was not caused by changes in the NSL1 levels. Interestingly, while total protein levels of MOF, MBD-R2, and MCRS2 were also very similar between male and female rescued larvae, the NSL3 protein level appeared partially reduced in the *nsl1*^{R721A} larvae compared with the *nsl1*^{WT} larvae (Fig. 3I, lanes 3,6). This is consistent with the result of the Flag immunoprecipitation experiments in SL-2 cells when NSL1^{R721A} showed reduced coimmunoprecipitation of NSL3 (Fig. 3C). As a consequence, NSL3 might not be fully integrated into the complex and might therefore be degraded.

The NSL1–WDS interaction is required for the proper targeting of NSL1 to target promoters

To address the role of the NSL1–WDS interaction in the targeting of NSL1, we performed chromatin immunoprecipitation (ChIP) experiments on chromatins prepared from whole rescued male larvae using both anti-Flag antibody (Supplemental Fig. S7A) and endogenous anti-NSL1 antibody (Fig. 4A). Histone H3 ChIP served as a control for the quality of the chromatin preparation (Fig. 4B). It is important to note that despite overexpression of *Flag::nsl1* (*nsl1*^{WT} and *nsl1*^{R721A}) (Supplemental Fig. S7B), NSL1 ChIP signal on promoters of target genes in the *nsl1*^{WT} larvae was partially reduced compared with the real wild-type (*w*^{1118}) larvae (Fig. 4A). However, this reduction has no consequence on fly viability, as *nsl1*^{WT} animals are fully functional and viable (Fig. 3H). In contrast to *nsl1*^{WT}, binding of NSL1 to the promoters of target genes was further reduced in the mutant *nsl1*^{R721A} larvae, suggesting that the NSL1–WDS is important for the targeting of NSL1 (Fig. 4A; Supplemental Fig. S7A).

Additionally, we wanted to test whether reduction in NSL1 chromatin targeting also affects chromatin targeting of other complex members in vivo. We therefore performed ChIP using the MBD-R2 antibody since there were no differences in total protein levels of endogenous MBD-R2 in the *nsl1*^{WT} or *nsl1*^{R721A} larvae (Fig. 3I). Indeed, consistent with NSL1 reduction (Fig. 4A), we also observed a similar reduction in targeting of MBD-R2 to the promoters of target genes (Fig. 4C). To address whether this targeting defect was also observed globally in the *Drosophila* genome, we performed polytene chromosomal staining in *nsl1*^{WT} versus *nsl1*^{R721A} larvae. To be able to directly compare the staining of *nsl1*^{WT} (wild-type) and mutant larvae, the chromosomal squashes were prepared on the same slide, and images of wild-type and mutant polytene chromosomes were taken in the same tile scan. The male sample was distinguished from the female sample using MSL1 as an X-chromosomal marker (Supplemental Fig. S7C). Consistent with the reduction of MBD-R2 in ChIP analyses, we observed a striking reduction of MBD-R2 staining on the mutant chromosomes in comparison with wild type in male and female larvae (Fig. 4D; Supplemental Fig. S7C). These data clearly show that the NSL1–WDS interaction is essential for the viability of *Drosophila* and the proper targeting of the NSL complex to the promoters of target genes in vivo.

Discussion

The subunits of the NSL complex play important roles in various cellular processes, including transcription regulation and stem cell identity maintenance or reprogramming (Raja et al. 2010; Li et al. 2012; Zhao et al. 2013a), and are misregulated in various diseases, including cancer (Fraga et al. 2005; Gupta et al. 2008; Yoshida et al. 2013). The composition of the NSL complex has been analyzed by mass spectrometry (Mendjan et al. 2006; Cai et al. 2010), but essentially nothing is known about its biochemistry, molecular structure, and mode of action. In this study, we report the first detailed biochemical and structural analyses of the NSL complex architecture. We show that NSL1 acts as the scaffold for the complex assembly interacting with MOF, WDR5/WDS, MBD-R2, and MCRS2. The high-resolution crystal structures of the NSL1/WDS/NSL2 complexes revealed that NSL1 interacts via a short linear motif around Arg721 with WDS, which also recognizes another short motif of NSL2. Previously, we showed that another interacting motif in the NSL1 C terminus is involved in the interaction with MOF (Kadlec et al. 2011). The interacting regions for MBD-R2 and MCRS2 remain unknown, but given the high predicted disorder content of NSL1, it is possible that, similar to the MSL1 subunit of the MSL complex (Kadlec et al. 2011; Hallacli et al. 2012), NSL1 uses short conserved motifs separated by long flexible regions to interact with its partners, providing the NSL complex with a certain level of plasticity. Together, our data suggest that NSL1 bridges the MOF acetyltransferase with WDR5/WDS, which in turn brings together the putative Zn finger of NSL1 with four Zn-coordinating motifs of NSL2. We speculate that these putative Zn-binding modules might be involved in the interaction with the target DNA. Additionally, NSL1 helps NSL2 recruitment via the interaction with MCRS2 (Fig. 5).

Mutating NSL1 Arg721 is sufficient to eliminate WDS from the complex, indicating that its interaction with NSL1 is absolutely required for its recruitment into the NSL complex. Interestingly, removal of WDS from the complex is accompanied by the loss of NSL3. Additionally, the overall NSL3 protein level was also partially reduced in the *nsl1*^{R721A} larvae compared with the *nsl1*^{WT}. As we did not observe a direct interaction between NSL3 and WDS (data not shown), NSL3 might require possible WDS-induced conformational changes within the complex or the presence of a composite binding site, including WDS and other complex subunits. Substitutions in the lower-

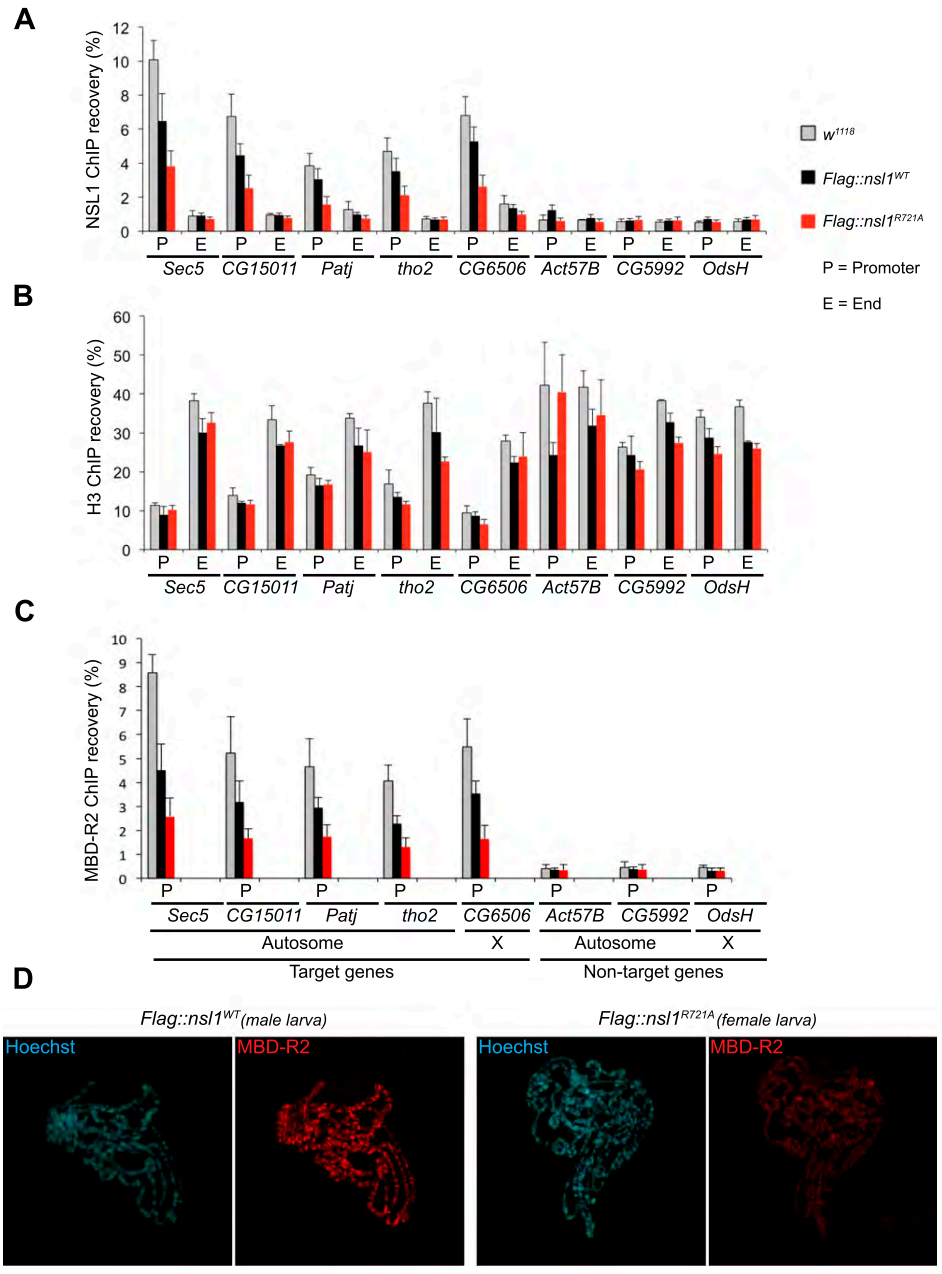


Figure 4. The NSL1–WDS interaction is required for proper targeting of NSL1 to the promoters of target genes. (*A*) NSL1 ChIP from whole male larvae. qPCR was performed using primers corresponding to the promoter (P) and end (E) regions of genes. The list of target and nontarget genes was chosen based on the genome-wide data (Raja et al. 2010; Lam et al. 2012). ChIP recovery was calculated over the input DNA (shown as percentage [%]). The end of target genes and both the promoter and the end of nontarget genes were used as negative controls. The error bars represent standard deviations of three biological replicates. (*B*) As in *A*, but this time Histone H3 ChIP from whole male larvae. This provides a control for the quality of the chromatin preparations. (*C*) As in *A*, but this time MBD-R2 ChIP from whole male larvae. (*D*) Polytene chromosome immunostaining of *Flag::nsl1*^{WT} male and *Flag::nsl1*^{R721A} female third instar larvae. Chromosomal squashes were performed on the same slide; shown images of single nuclei were cropped from the same tile scan image. Immunofluorescence staining of MBD-R2 (red) is shown; DNA is counterstained with Hoechst (cyan).

affinity WDS-binding motif of NSL2 seem to have no effect on WDS levels, consistent with the major role of NSL1 in WDS recruitment.

We show that the NSL1–WDS interface mutagenesis triggered a decrease of NSL1 binding to promoters of the NSL target genes, resulting in fly lethality. In addition, ChIP and polytene chromosome immunostaining revealed a similar reduction of the MBD-R2 subunit on mutant chromosomes compared with wild type. Interestingly, while the binding of the *Flag::nsl1*^{WT} transgene was also

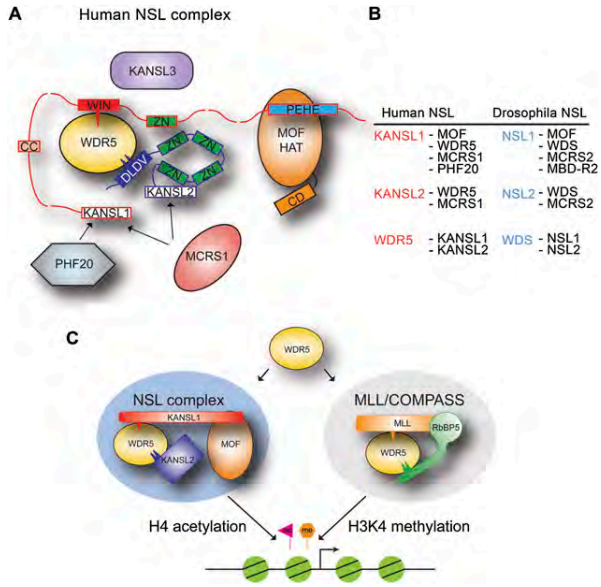


Figure 5. A summary model. (*A*) Schematic model of the NSL complex architecture based on our structural and protein interaction studies. KANSL1 is likely to be a scaffold molecule of the complex, interacting with MOF, WDR5, MCRS1, and PHF20. KANSL1 connects the MOF acetyltransferase (Kadlec et al. 2011) with WDR5 using conserved, short binding motifs. Similarly, WDR5 recognizes short motifs of KANSL1 and KANSL2 bringing in proximity the putative Zn finger of KANSL1 with four Zn-coordinating motifs of KANSL2, possibly involved in the interaction with the target DNA. KANSL2 is incorporated into the complex by additional interaction with MCRS1. (CD) Chromo-barrel domain; (CC) coiled coil; (WIN) WDS-interacting motif; (ZN) Zn finger. The arrows indicate interactions of MCRS1 and PHF20 with KANSL1. (*B*) Summary of known interactions within the NSL complex in humans and *Drosophila*. (*C*) Summary model. WDR5 serves a binding platform mediating mutually exclusive interactions essential for efficient promoter targeting of the NSL HAT complex and stimulating the H3K4 methyltransferase activity of MLL proteins within the MLL/COMPASS complexes (Patel et al. 2008b; Odho et al. 2010; Avdic et al. 2011; Zhang et al. 2012).

partially reduced, it could still efficiently rescue *nsl1*-deficient flies. The observed phenotype thus seems critically dependent on the amount of the chromatin-bound NSL complex. This observation is supported by the fact the KANSL1 haploinsufficiency results in the 17q21.31 microdeletion multisystem syndrome (Koolen et al. 2012; Zollino et al. 2012).

The reduced NSL complex binding to promoters might be a result of an altered NSL2 configuration within the complex (in the wild type, it is connected to NSL1 via WDS), since NSL2 possesses four Zn-coordinating motifs possibly involved in the recognition of target DNA sequences. The strong lethal phenotype could also (at least partially) originate from the lack of NSL3 within the mutated complex, induced by the absence of WDS. NSL3 possesses a putative enzymatic domain of the α/β hydrolase family, but its function remains unknown. Nevertheless, a severe reduction in transcript levels of most NSL target genes was observed upon depletion of NSL3 (Raja et al. 2010). In addition NSL3 is able to activate transcription in a luciferase reporter-tethering assay (Raja et al. 2010). It is thus possible that the lack of NSL3 within the complex directly affects the established role of the NSL complex in transcription regulation.

WDR5 was originally shown to bind H3R2 peptides and proposed to present H3K4 for methylation by MLL proteins (Couture et al. 2006; Han et al. 2006; Ruthenburg et al. 2006; Schuetz et al. 2006). Later, the crystal structure of the MLL1-SET domain in complex with the H3K4 peptide revealed that H3 is unlikely to be bound by WDR5 (H3R2) and MLL1 (H3K4) at the same time (Southall et al. 2009; Trievel and Shilatifard 2009). In accord with this structure, WDR5 was shown to recognize a similar sequence (the WIN motif) of the MLL proteins themselves as well as RbBP5 (Patel et al. 2008a; Song and Kingston 2008; Odho et al. 2010; Avdic et al. 2011). However, while the structure of the WDR5 β-propeller domain and its ability to bind H3, the WIN motif, and RbBP5 within MLL complexes are well documented, it remains unclear how WDR5 presents itself within other complexes, such as NSL or ATAC. WDR5 was proposed to be a bridging or cross-shared subunit between the NSL and MLL/COMPASS complexes, forming complexes with both histone methyltransferase and acetyltransferase activities (Dou et al. 2005; Li et al. 2009; Zhao et al. 2013b). However, we and others could not confirm the existence of such a complex (Cai et al. 2010; Raja et al. 2010). Another possibility is that WDR5 targets its complexes to histone H3 (via H3R2 binding). It could also interact within these complexes in a manner similar to or different from that of the MLL/COMPASS complexes. Since other WDR5-binding partners as well as the molecular details of their interactions were not known, the role of WDR5 within could not be investigated. We now, for the first time, identified proteins directly interacting with WDR5 within the NSL complex and mapped the interacting regions. Our crystal structures clearly show that WDR5 regions interacting with KANSL1 and KANSL2 are the same as within the MLL/COMPASS complexes and that these interactions are mutually exclusive. This study thus provides compelling evidence that WDR5 cannot be a cross-shared subunit between MLL and NSL complexes and explains how WDR5 can serve as a binding platform in distinct complexes such as NSL, MLL/COMPASS, and possibly ATAC. Analogously, the NSL complex cannot use WDR5 to recognize H3R2, since its binding site is occupied by KANSL1. The interaction between WDR5 and the MLL WIN motif has recently been proposed to be a promising new cancer therapeutic target (Karatas et al. 2013; Senisterra et al. 2013; Cao et al. 2014). While the developed small molecule inhibitors interfere efficiently with MLL binding to WDR5, possible side effects on the essential WDR5 interactions within other complexes, such as NSL, should be considered.

Interestingly, WDR5 has also very recently been shown to interact with hundreds of long noncoding RNAs (lncRNAs), including HOTTIP and NeST (Wang et al. 2011; Gomez et al. 2013; Yang et al. 2014), that pre-

sumably target the MLL complex to its target genes in embryonic stem cells. The RNA recognition site on WDR5 was mapped by random mutagenesis to a site overlapping with that responsible for the interaction with RbBP5 and NSL2 (Yang et al. 2014). Indeed, the presence of RbBP5 prevented HOTTIP lncRNA binding to WDR5 (Yang et al. 2014). While the role of lncRNA in the targeting of the NSL has not yet been studied, it should be noted that the way WDR5 is recruited into the NSL complex is compatible with the RNA binding as reported for the MLL complex. It is tempting to speculate that the recruitment of the NSL and MLL complexes to their target genes via WDR5-bound lncRNAs could possibly explain observed functional cross-talk between them that was originally assigned to WDR5 itself.

Given the mutually exclusive binding of WDR5 to KANSL1/MLL and KANSL2/RbBP5, it will be important to understand how the proper pairing of the interactions in distinct complexes is achieved. For example, we could show that NSL2 pulls down the entire NSL complex but not Trx, indicating that such mixed complexes (e.g., Trx/WDS/NSL2), which might be inactive or have aberrant activity, do not exist in vivo. However, the reported minimal H3/WDR5/RbBP5 structure indicates that such complexes can indeed be formed in vitro (Odho et al. 2010). We show that both MLLs and KANSL1 as well as KANSL2 and RbBP5 interact with WDR5 with similar affinities. Interestingly, in both the NSL and MLL complexes, the two WDR5-binding proteins are linked by other interactions/proteins. While RbBP5 was proposed to directly interact with MLL1 (Odho et al. 2010), NSL2 is linked to NSL1 via MCRS2. It is thus likely that the proper combinations of WDR5 binders connected together by additional interactions will have a higher affinity for WDR5 than the incorrect combination of WDR5-binding proteins that cannot be physically linked. Future structural studies will be very important in unraveling how other components of the NSL complex shape its molecular interaction network and cross-talk with other chromatin-modifying complexes.

Materials and methods

Flag pull-down assays

All protein extracts used for binding assays were produced by baculoviruses in the Sf21 cells (see the Supplemental Material for details of generation of baculoviruses and protein extraction). For each binding assay, 1.5 mL of the bait protein extract was mixed with 5 mL of the prey protein extract. The mixture was incubated for 2 h at 4°C while rotating. About 50 µL of equilibrated Flag resin (Sigma) was added to each mixture and then incubated for 3 h at 4°C while rotating. After five washes with HEMGT250 buffer (25 mM HEPES at pH 7.6, 12.5 mM $MgCl_2$, 10% glycerol, 0.2% Tween 20, 250 mM KCl), the beads were either boiled in 125 µL of 2× Roti-Load buffer (Carl-Roth) or eluted overnight in 500 µL of HMGT250 buffer containing 250 µg/mL 3xFlag peptide (Sigma).

Protein expression and purification for structural studies

6xHis-$WDR5_{23–334}$ was produced by expression in *Escherichia coli* BL21Star (DE3, Invitrogen) from the pProEXHTb (Invitrogen) expression vector and affinity-purified on Ni^{2+} chelating Sepharose (GE Healthcare). After His tag cleavage with TEV protease, the protein was further purified by a second Ni^{2+} column and subsequent size exclusion chromatography on a Superdex 75 (GE Healthcare) gel filtration column pre-equilibrated with 20 mM Tris (pH 7.0), 200 mM NaCl, and 5 mM β-mercaptoethanol. 6xHis-WDS 50-361 was produced by expression in Hi5 insect cells from the pFastBacHTb (Invitrogen) vector and purified as for WDR5.

KANSL1–WDR5 and KANSL2–WDR5 interaction analysis

Four constructs of untagged KANSL1 (residues 1–233, 262–537, 537–773, and 777–1105) were each coexpressed with 6xHis-WDR5 (residues 23–334) in *E. coli* BL21Star (DE3, Invitrogen) from pRSFDuet-1 (Novagen) and pProEXHTb (Invitrogen) expression vectors, respectively. 6xHis-WDR5 was affinity-purified on Ni^{2+} chelating Sepharose (GE Healthcare) in 20 mM Tris (pH 7.0), 200 mM NaCl, and 5 mM β-mercaptoethanol. Proteins were eluted with an increasing concentration of imidazole. Samples were analyzed on SDS-PAGE. A His-tagged minimal interacting region of KANSL1 (residues 584–690; wild type or R592A mutant) was coexpressed with untagged WDR5 (residues 23–334) in *E. coli* BL21Star (DE3, Invitrogen) from pProEXHTb (Invitrogen) and pRSFDuet-1 (Novagen) expression vectors, respectively. Pull-down experiments were performed as above.

Constructs of KANSL2 were produced as 6xHis-GST fusion proteins in pETM30 expression vector. $KANSL2_{381–492}$ and mutated $KANSL2_{381–492}$ L411E,V413D were expressed in *E. coli* BL21Star (DE3, Invitrogen) and purified as for WDR5. Pure wild-type $KANSL2_{381–492}$ or its V413D mutant were mixed with pure $WDR5_{23–334}$ and loaded onto a Superdex 200 gel filtration column. Fractions containing protein were analyzed on SDS-PAGE.

Crystallization

All of the crystals of the WDR5/WDS-containing complexes were grown using pure WDR5 or WDS concentrated to 12–16 mg/mL supplemented with a threefold molar excess of peptide of individual peptides. The best diffracting crystals of the binary complex of $WDR5_{23–334}$ with KANSL1 peptide (585-DGTCVAARTRPVLS-598-Y; Y was added to facilitate quantification) were obtained in a condition containing 26% (w/v) PEG 3350, 0.2 M ammonium sulphate, and 0.1 M Bis-Tris (pH 6.5). Crystals of the binary complex of $WDS_{50–361}$ with NSL1 peptide (714-GSDYLCSRARPLVLSE-729) grew in 20% (w/v) PEG 3350 and 0.2 M potassium sodium tartrate (pH 7.2). Crystals of the ternary complex of $WDR5_{23–334}$ with KANSL1 peptide (585-DGTCVAARTRPVLSY-598) and KANSL2 peptide (Y-406-EFSDDLDVVGDG-417) grew in 20% (w/v) PEG 3350 and 0.2 M potassium nitrate (pH 6.9). The best crystals of the ternary complex of $WDS_{50–361}$ with NSL1 peptide (714-GSDYLCS RARPLVLSE-729) and NSL2 peptide (Y-155-RDDDEIDVVSPH-166) were obtained in 20% (w/v) PEG 3350 and 0.2 M di-ammonium tartrate (pH 6.6). For data collection at 100 K, crystals were snap-frozen in liquid nitrogen with a solution containing mother liquor and 30% (v/v) glycerol.

Data collection and structure determination

Crystals of the binary complex $WDR5_{23–334}$ with KANSL1 peptide (585-DGTCVAARTRPVLSY-598) belong to the space group $P2_12_12_1$ with unit cell dimensions of a = 39.8 Å, b = 92.5 Å, c = 81.2 Å, and β = 90.2°. The asymmetric unit contained two complexes and had a solvent content of 45%. A complete data set was collected to a resolution of 1.5 Å on beamline ID14-EH4

Table 1. *Data collection and refinement statistics*

	KANSL1/WDR5	NSL1/WDS	KANSL1/WDR5/KANSL2	NSL1/WDS/NSL2
Data collection				
Space group	$P2_1$	$C222_1$	$P2_12_12$	$P2_12_12$
Cell dimensions				
a, *b*, *c*	39.8 Å, 92.5 Å, 81.2 Å	78.7 Å, 98.8 Å, 80 Å	81.6 Å, 86.6 Å, 45.1 Å	81.5 Å, 86.2 Å, 47.6 Å
α, β, γ	90°, 90.2°, 90°	90°, 90°, 90°	90°, 90°, 90°	90°, 90°, 90°
Resolution	100 Å–1.5 Å (1.55 Å–1.5 Å)[a]	100 Å–1.4 Å (1.46 Å–1.4 Å)	100 Å–2.0 Å (2.1 Å–2.0 Å)	100 Å–2.3 Å (2.4 Å–2.3 Å)
R_{merge}	7.6 (63.8)	10.6 (86.8)	15 (83.7)	8.6 (73.7)
I/σ(I)	13.6 (2.4)	13.8 (2.5)	10.7 (2.6)	15.3 (2.3)
Completeness	96.7% (89%)	98.9% (91.2%)	100% (99.9%)	99.4% (98.9%)
Redundancy	3.8 (3.8)	7.3 (7.1)	7.3 (7.4)	4.4 (4.4)
Refinement				
Resolution	40 Å–1.5 Å	48 Å–1.4 Å	43 Å–2 Å	37 Å–2.3 Å
Number of reflections	86,154	57,889	21,314	14,628
R_{work}/R_{free}	17.2 (18.4)	18.0 (19.0)	20.23 (22.2)	20.8 (24.6)
Average *B*-factors	11.8	10.3	15.1	42
RMSDs				
Bond lengths	0.004 Å	0.004 Å	0.006 Å	0.006 Å
Bond angles	1.011°	1.018°	1.071°	1.152°

[a]Values in parentheses are for highest-resolution shell.

at the European Synchrotron Radiation Facility (ESRF). Crystals of the binary complex $WDS_{50–361}$ with NSL1 peptide (714-GSDYLCSRARPLVLSE-729) belong to the space group $C222_1$, with unit cell dimensions a = 78.7 Å, b = 98.6 Å, and c = 80 Å. A complete data set was collected to a resolution of 1.4 Å on the ESRF beamline ID23-EH1. Crystals of the ternary complex $WDR5_{23–334}$ with KANSL1 peptide (585-DGTCVAARTRPVLSY-598) and KANSL2 peptide (406-YEFSDDLDVVGDG-417) belong to the space group $P2_12_12$, with unit cell dimensions a = 81.6 Å, b = 86.6 Å, and c = 45.1 Å. A complete data set was collected to a resolution of 2 Å on the ESRF beamline ID23.2. Crystals of the ternary complex $WDS_{50–361}$ with NSL1 peptide (714-GSDYLCS RARPLVLSE-729) and NSL2 peptide (155-YRDDDEIDVVSPH-166) belong to the space group $P2_12_12_1$ with unit cell dimensions a = 81.5 Å, b = 86.2 Å, and c = 47.6 Å. A complete data set was collected to a resolution of 2.3 Å on beamline BM14 at the ESRF. The data were processed using XDS (Kabsch 2010). The structures were solved by molecular replacement with Phaser (McCoy et al. 2005) using the structure of WDR5 (Protein Data Bank [PDB] code 2G99) as a search model. The structures were manually rebuilt in Coot (Emsley and Cowtan 2004) and refined in REFMAC5 (Murshudov et al. 1997) to the final R factors indicated in Table 1 with all residues in allowed (94%–96% in favored) regions of the Ramachandran plot, as analyzed by MolProbity (Davis et al. 2004). Representative parts of the F_o–F_c electron density maps calculated using the refined models are shown in Supplemental Figure S8.

ITC

ITC experiments were performed at 25°C using an ITC200 microcalorimeter (MicroCal). Experiments included 26 injections of 1.5 µL of 0.5 mM wild-type or 2 mM mutant peptide solution into the sample cell containing 50 µM $WDR5_{23–334}$ or $WDS_{50–361}$ in 20 mM Tris (pH 7.0) and 200 mM NaCl. Peptides used were KANSL1 wild type (585-DGTCVAARTRPVLS-598-Y), KANSL1 R592A (585-DGTCVAAATRPVLSY-598), KANSL2 wild type (406-YEFSDDLDVVGDG-417), KANSL2 L411E,V413D (406-YEFSD DEDDVGDG-417), NSL2 wild type (155-YRDDDEIDVVSPH-166), and NSL2 I160E V162D (155-YRDDDEEDDVSPH-166). The initial data point was deleted from the data sets. Binding isotherms were fitted with a one-site binding model by nonlinear regression using Origin software version 7.0 (MicroCal).

To confirm that KANSL1 and MLL4 bind to the same binding site on WDR5, 0.56 and 0.35 mM MLL4 peptide (2504-LNPHGAARAEVYLRK-2518) was injected into the sample cell containing 50–60 µM $WDR5_{23–334}$ with or without 200 µM KANSL1 peptide, respectively. To test the mutually exclusive binding of KANSL1 and histone H3, 1.9 and 1.8 mM H3 peptide (1-ARTKQTARK-9-Y) was injected into the sample cell containing 120 µM $WDR5_{23–334}$ with or without 400 µM KANSL1 peptide, respectively.

Flag immunoprecipitation from SL-2 cells

Flag immunoprecipitation from transiently or stably expressing SL-2 cells were performed as previously described (Hallacli et al. 2012).

Analysis of lethality rescue and rate of adult eclosion

To determine the relative viability upon ectopic expression of *Flag::nsl1* (wild type or R721A), *w; Act5C-Gal4/CyO, Act5C-GFP; nsl1*^{S009413}*/TM6BTb* virgin females were crossed to either *w/Y; UAS-Flag::nsl1*^{WT}*; nsl1*^{j2E5}*/TM6BTb* or *w/Y; UAS-Flag::nsl1*^{R721A}*; NLS1*^{j2E5}*/TM6BTb* males. Male and female adult flies from at least three independent crosses were counted every other day for a period of 10 d from the start of eclosion. The total number of non-CyO, non-Tb males and females was divided by the total number of non-CyO, Tb males and females, respectively, which were used as an internal control with 100% viability. Details of fly genetics are available in the Supplemental Material.

ChIP

ChIP from male whole larvae was carried out as previously described (Raja et al. 2010) using primers listed in Supplemental Table S1.

Coordinates

The atomic coordinates and structure factors of the human KANSL1/WDR5 and KANSL1/WDR5/KANSL2 complexes and

Drosophila NSL1/WDS and NSL1/WDS/NSL2 complexes have been deposited under the PDB accession codes 4CY1, 4CY2, 4CY3, and 4CY5, respectively.

Acknowledgments

We thank the Grenoble European Molecular Biology Laboratory (EMBL)–European Synchrotron Radiation Facility (ESRF)–Institut Laue-Langevin (ILL)–Institut de Biologie Structurale (IBS) Partnership for Structural Biology (PSB) for access to structural biology instrumentation—notably L. Signor of the IBS for his help with mass spectrometry, and the High-Throughput Crystallisation Laboratory (HTX) group of EMBL Grenoble for performing initial screening crystallization trials. We thank A. Palencia for help with the ITC, and K. Dzeyk of EMBL Heidelberg for help with mass spectrometry. We thank M. Shvedunova and H. Holz for the initial help with baculoviruses. We thank S. Chlamydas and T. Lingg for technical support, and other members of both laboratories and Ramesh Pillai for useful advice and fruitful discussions. We thank R. Sawarkar and R. Paro for providing the Trx antibody. We thank the ESRF–EMBL Joint Structural Biology Group for access to and assistance on the ESRF synchrotron beamlines. This work was supported by the EU FP7-funded Network of Excellence EpiGeneSys awarded to S.C. and A.A., and SFB746 and SFB992 awarded to A.A. A.A. is also part of the BIOSS II excellence initiative.

References

Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, et al. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145:** 183–197.

Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, Couture JF. 2011. Structural and biochemical insights into MLL1 core complex assembly. *Structure* **19:** 101–108.

Badeaux AI, Yang Y, Cardenas K, Vemulapalli V, Chen K, Kusewitt D, Richie E, Li W, Bedford MT. 2012. Loss of the methyl lysine effector protein PHF20 impacts the expression of genes regulated by the lysine acetyltransferase MOF. *J Biol Chem* **287:** 429–437.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118:** 401–415.

Cai Y, Jin J, Swanson SK, Cole MD, Choi SH, Florens L, Washburn MP, Conaway JW, Conaway RC. 2010. Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem* **285:** 4268–4272.

Cao F, Townsend EC, Karatas H, Xu J, Li L, Lee S, Liu L, Chen Y, Ouillette P, Zhu J, et al. 2014. Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol Cell* **53:** 247–261.

Conrad T, Akhtar A. 2011. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet* **13:** 123–134.

Couture JF, Collazo E, Trievel RC. 2006. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol* **13:** 698–703.

Cui G, Park S, Badeaux AI, Kim D, Lee J, Thompson JR, Yan F, Kaneko S, Yuan Z, Botuyan MV, et al. 2012. PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nat Struct Mol Biol* **19:** 916–924.

Davis IW, Murray LW, Richardson JS, Richardson DC. 2004. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32:** W615–W619.

Dharmarajan V, Lee JH, Patel A, Skalnik DG, Cosgrove MS. 2012. Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *J Biol Chem* **287:** 27275–27289.

Dou Y, Milne TA, Tackett AJ, Smith ER, Fukuda A, Wysocka J, Allis CD, Chait BT, Hess JL, Roeder RG. 2005. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121:** 873–885.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132.

Feller C, Prestel M, Hartmann H, Straub T, Soding J, Becker PB. 2012. The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. *Nucleic Acids Res* **40:** 1509–1522.

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, et al. 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* **37:** 391–400.

Fullgrabe J, Lynch-Day MA, Heldring N, Li W, Struijk RB, Ma Q, Hermanson O, Rosenfeld MG, Klionsky DJ, Joseph B. 2013. The histone H4 lysine 16 acetyltransferase hMOF regulates the outcome of autophagy. *Nature* **500:** 468–471.

Gomez JA, Wapinski OL, Yang YW, Bureau JF, Gopinath S, Monack DM, Chang HY, Brahic M, Kirkegaard K. 2013. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. *Cell* **152:** 743–754.

Groth AC, Fish M, Nusse R, Calos MP. 2004. Construction of transgenic *Drosophila* by using the site-specific integrase from phage φC31. *Genetics* **166:** 1775–1782.

Gupta A, Guerin-Peyrou TG, Sharma GG, Park C, Agarwal M, Ganju RK, Pandita S, Choi K, Sukumar S, Pandita RK, et al. 2008. The mammalian ortholog of *Drosophila* MOF that acetylates histone H4 lysine 16 is essential for embryogenesis and oncogenesis. *Mol Cell Biol* **28:** 397–409.

Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A, Kadlec J. 2012. Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Mol Cell* **48:** 587–600.

Han Z, Guo L, Wang H, Shen Y, Deng XW, Chai J. 2006. Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol Cell* **22:** 137–144.

Kabsch W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125–132.

Kadlec J, Hallacli E, Lipp M, Holz H, Sanchez-Weatherby J, Cusack S, Akhtar A. 2011. Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol* **18:** 142–149.

Karatas H, Townsend EC, Cao F, Chen Y, Bernard D, Liu L, Lei M, Dou Y, Wang S. 2013. High-affinity, small-molecule peptidomimetic inhibitors of MLL1/WDR5 protein-protein interaction. *J Am Chem Soc* **135:** 669–682.

Koolen DA, Kramer JM, Neveling K, Nillesen WM, Moore-Barton HL, Elmslie FV, Toutain A, Amiel J, Malan V, Tsai AC, et al. 2012. Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat Genet* **44:** 639–641.

Lam KC, Muhlpfordt F, Vaquerizas JM, Raja SJ, Holz H, Luscombe NM, Manke T, Akhtar A. 2012. The NSL complex regulates housekeeping genes in *Drosophila*. *PLoS Genet* **8:** e1002736.

Lee T, Luo L. 1999. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22:** 451–461.

Li X, Wu L, Corsa CA, Kunkel S, Dou Y. 2009. Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. *Mol Cell* **36:** 290–301.

Li X, Li L, Pandey R, Byun JS, Gardner K, Qin Z, Dou Y. 2012. The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. *Cell Stem Cell* **11:** 163–178.

McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. 2005. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61:** 458–464.

Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J, et al. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol Cell* **21:** 811–823.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255.

Odho Z, Southall SM, Wilson JR. 2010. Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *J Biol Chem* **285:** 32967–32976.

Patel A, Dharmarajan V, Cosgrove MS. 2008a. Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem* **283:** 32158–32161.

Patel A, Vought VE, Dharmarajan V, Cosgrove MS. 2008b. A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem* **283:** 32162–32175.

Raja SJ, Charapitsa I, Conrad T, Vaquerizas JM, Gebhardt P, Holz H, Kadlec J, Fraterman S, Luscombe NM, Akhtar A. 2010. The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol Cell* **38:** 827–841.

Ruthenburg AJ, Wang W, Graybosch DM, Li H, Allis CD, Patel DJ, Verdine GL. 2006. Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol* **13:** 704–712.

Schuetz A, Allali-Hassani A, Martin F, Loppnau P, Vedadi M, Bochkarev A, Plotnikov AN, Arrowsmith CH, Min J. 2006. Structural basis for molecular recognition and presentation of histone H3 by WDR5. *EMBO J* **25:** 4245–4252.

Senisterra G, Wu H, Allali-Hassani A, Wasney GA, Barsyte-Lovejoy D, Dombrovski L, Dong A, Nguyen KT, Smil D, Bolshan Y, et al. 2013. Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. *Biochem J* **449:** 151–159.

Shilatifard A. 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* **81:** 65–95.

Song JJ, Kingston RE. 2008. WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* **283:** 35258–35264.

Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR. 2009. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell* **33:** 181–191.

Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, McMahon SB. 2006. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Mol Cell* **24:** 841–851.

Taipale M, Rea S, Richter K, Vilar A, Lichter P, Imhof A, Akhtar A. 2005. hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol Cell Biol* **25:** 6798–6810.

Taylor G, Eskeland R, Hekimoglu-Balkan B, Pradeepa M, Bickmore WA. 2013. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res* **23:** 2053–2065.

Thomas T, Dixon MP, Kueh AJ, Voss AK. 2008. Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. *Mol Cell Biol* **28:** 5093–5105.

Trievel RC, Shilatifard A. 2009. WDR5, a complexed protein. *Nat Struct Mol Biol* **16:** 678–680.

Venken KJ, He Y, Hoskins RA, Bellen HJ. 2006. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314:** 1747–1751.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472:** 120–124.

Yang YW, Flynn RA, Chen Y, Qu K, Wan B, Wang KC, Lei M, Chang HY. 2014. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *eLife* **3:** e02046.

Yoshida K, Toki T, Okuno Y, Kanezaki R, Shiraishi Y, Sato-Otsubo A, Sanada M, Park MJ, Terui K, Suzuki H, et al. 2013. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat Genet* **44:** 1293–1299.

Yu L, Song Y, Wharton RP. 2010. E(nos)/CG4699 required for nanos function in the female germ line of *Drosophila*. *Genesis* **48:** 161–170.

Zhang P, Lee H, Brunzelle JS, Couture JF. 2012. The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res* **40:** 4237–4246.

Zhao W, Li Q, Ayers S, Gu Y, Shi Z, Zhu Q, Chen Y, Wang HY, Wang RF. 2013a. Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination. *Cell* **152:** 1037–1050.

Zhao X, Su J, Wang F, Liu D, Ding J, Yang Y, Conaway JW, Conaway RC, Cao L, Wu D, et al. 2013b. Crosstalk between NSL histone acetyltransferase and MLL/SET complexes: NSL complex functions in promoting histone H3K4 dimethylation activity by MLL/SET complexes. *PLoS Genet* **9:** e1003940.

Zollino M, Orteschi D, Murdolo M, Lattante S, Battaglia D, Stefanini C, Mercuri E, Chiurazzi P, Neri G, Marangi G. 2012. Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat Genet* **44:** 636–638.

References

1. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
2. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* **20,** 259–66 (2013).
3. Schapira, M. Structural Chemistry of Human SET Domain Protein Methyltransferases. *Curr. Chem. Genomics* **5,** 85–94 (2011).
4. Hohenauer, T. & Moore, A. W. The Prdm family: expanding roles in stem cells and development. *Development* **139,** 2267–2282 (2012).
5. Nakaki, F. *et al.* Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501,** 222–6 (2013).
6. Pinheiro, I. *et al.* Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. *Cell* **150,** 948–960 (2012).
7. Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327,** 836–40 (2010).
8. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327,** 876–9 (2010).
9. Hayashi, K., Yoshida, K. & Matsui, Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438,** 374–378 (2005).
10. Grey, C. *et al.* Mouse Prdm9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.* **9,** (2011).
11. Ponting, C. P. What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics* **27,** 165–171 (2011).
12. Patel, A., Horton, J. R., Wilson, G. G., Zhang, X. & Cheng, X. Structural basis for human PRDM9 action at recombination hot spots. *Genes Dev.* **30,** 257–265 (2016).
13. Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase. *Science (80-.).* **323,** 373–5 (2009).
14. Wu, H. *et al.* Molecular Basis for the Regulation of the H3K4 Methyltransferase Activity of PRDM9. *Cell Rep.* **5,** 13–20 (2013).
15. Avvakumov, N. & Côté, J. The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene* **26,** 5395–5407 (2007).
16. Conrad, T. & Akhtar, A. Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet* **13,** 123–34 (2011).
17. Keller, C. I. & Akhtar, A. The MSL complex: Juggling RNA-protein interactions for dosage compensation and beyond. *Current Opinion in Genetics and Development* **31,** 1–11 (2015).
18. Taipale, M. *et al.* hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell. Biol.* **25,** 6798–6810 (2005).
19. Li, X. *et al.* The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. *Cell Stem Cell* **11,** 163–178 (2012).
20. Thomas, T., Dixon, M. P., Kueh, A. J. & Voss, A. K. Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. *Mol. Cell. Biol.* **28,** 5093–105 (2008).
21. Füllgrabe, J. *et al.* The histone H4 lysine 16 acetyltransferase hMOF regulates the outcome of autophagy. *Nature* **500,** 468–71 (2013).
22. Li, X. *et al.* MOF and H4 K16 acetylation play important roles in DNA damage repair by modulating recruitment of DNA damage repair protein Mdc1. *Mol. Cell. Biol.* **30,** 5335–47 (2010).
23. Sharma, G. G. *et al.* MOF and histone H4 acetylation at lysine 16 are critical for DNA damage response and double-strand break repair. *Mol. Cell. Biol.* **30,** 3582–95 (2010).
24. Li, X., Wu, L., Corsa, C. A. S., Kunkel, S. & Dou, Y. Two Mammalian MOF Complexes Regulate Transcription Activation by Distinct Mechanisms. *Mol. Cell* **36,** 290–301 (2009).
25. Su, J., Wang, F., Cai, Y. & Jin, J. The functional analysis of histone acetyltransferase MOF in tumorigenesis. *International Journal of Molecular Sciences* **17,** (2016).
26. Raja, S. J. *et al.* The nonspecific lethal complex is a transcriptional regulator in Drosophila. *Mol. Cell* **38,** 827–841 (2010).
27. Ilik, I. *et al.* Tandem stem-loops in roX RNAs act together to mediate X Chromosome dosage compensation in Drosophila. *Mol. Cell* **51,** 156–173 (2013).
28. Maenner, S., Müller, M., Fröhlich, J., Langer, D. & Becker, P. ATP-dependent roX RNA remodeling by the Helicase maleless enables specific association of MSL proteins. *Mol. Cell* **51,** 174–184 (2013).

29. Alekseyenko, A. A. *et al.* A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the Drosophila X Chromosome. *Cell* **134,** 599–609 (2008).
30. Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A. & Becker, P. B. The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. *PLoS Genet.* **4,** (2008).
31. Villa, R., Schauer, T., Smialowski, P., Straub, T. & Becker, P. B. PionX sites mark the X chromosome for dosage compensation. *Nature* **537,** 244–248 (2016).
32. Larschan, E. *et al.* X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* **471,** 115–118 (2011).
33. Kim, D. *et al.* Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat. Struct. Mol. Biol.* **17,** 1027–9 (2010).
34. Nielsen, P. R. *et al.* Structure of the chromo barrel domain from the MOF acetyltransferase. *J. Biol. Chem.* **280,** 32326–32331 (2005).
35. Zheng, S. *et al.* Structural basis of X chromosome DNA recognition by the MSL2 CXC domain during Drosophila dosage compensation. *Genes Dev.* **28,** 2652–2662 (2014).
36. Kadlec, J. *et al.* Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat. Struct. Mol. Biol.* **18,** 142–149 (2011).
37. Sanchez-Weatherby, J. *et al.* Improving diffraction by humidity control: A novel device compatible with X-ray beamlines. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65,** 1237–1246 (2009).
38. Hallacli, E. *et al.* Msl1-Mediated Dimerization of the Dosage Compensation Complex Is Essential for Male X-Chromosome Regulation in Drosophila. *Mol. Cell* **48,** 587–600 (2012).
39. Prabu, J. R. *et al.* Structure of the RNA Helicase MLE Reveals the Molecular Mechanisms for Uridine Specificity and RNA-ATP Coupling. *Mol. Cell* **60,** 487–499 (2015).
40. Mendjan, S. *et al.* Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. *Mol. Cell* **21,** 811–823 (2006).
41. Cai, Y. *et al.* Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the Male-specific Lethal (MSL) complex. *J. Biol. Chem.* **285,** 4268–4272 (2010).
42. Badeaux, A. I. *et al.* Loss of the methyl lysine effector protein PHF20 impacts the expression of genes regulated by the lysine acetyltransferase MOF. *J. Biol. Chem.* **287,** 429–437 (2012).
43. Koolen, D. a *et al.* Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat. Genet.* **44,** 639–641 (2012).
44. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511,** 344–347 (2014).
45. Ang, Y.-S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145,** 183–97 (2011).
46. Zhao, W. *et al.* Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination. *Cell* **152,** 1037–50 (2013).
47. Lam, K. C. *et al.* The NSL complex regulates housekeeping genes in Drosophila. *PLoS Genet.* **8,** e1002736 (2012).
48. Chelmicki, T. *et al.* MOF-associated complexes ensure stem cell identity and Xist repression. *Elife* **2014,** 1–31 (2014).
49. Feller, C. *et al.* The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. *Nucleic Acids Res.* **40,** 1509–1522 (2012).
50. Meunier, S. *et al.* An epigenetic regulator emerges as microtubule minus-end binding and stabilizing factor in mitosis. *Nat. Struct. Mol. Biol.* **6,** doi: 10.1038/ncomms8889.37 (2015).
51. Fraga, M. F. *et al.* Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* **37,** 391–400 (2005).
52. Pfister, S. *et al.* The histone acetyltransferase hMOF is frequently downregulated in primary breast carcinoma and medulloblastoma and constitutes a biomarker for clinical outcome in medulloblastoma. *Int. J. Cancer* **122,** 1207–1213 (2008).
53. Gupta, A. *et al.* The mammalian ortholog of Drosophila MOF that acetylates histone H4 lysine 16 is essential for embryogenesis and oncogenesis. *Mol. Cell. Biol.* **28,** 397–409 (2008).
54. Zhao, L., Wang, D. L., Liu, Y., Chen, S. & Sun, F. L. Histone acetyltransferase hMOF promotes S phase entry and tumorigenesis in lung cancer. *Cell. Signal.* **25,** 1689–1698 (2013).
55. Chen, Z. *et al.* The histone acetylranseferase hMOF acetylates Nrf2 and regulates anti-drug responses in human non-small cell lung cancer. *Br. J. Pharmacol.* **171,** 3196–211 (2014).
56. Yoshida, K. *et al.* The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat. Genet.* **45,** 1293–9 (2013).

57. Okumura, K., Zhao, M., Depinho, R. a, Furnari, F. B. & Cavenee, W. K. Cellular transformation by the MSP58 oncogene is inhibited by its physical interaction with the PTEN tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2703–2706 (2005).
58. Jiao, L. & Liu, X. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science (80-.).* **350,** 291 (2015).
59. Dias, J. *et al.* Structural analysis of the KANSL1/WDR5/ KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. *Genes Dev.* **28,** 929–942 (2014).
60. Dou, Y. *et al.* Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121,** 873–885 (2005).
61. Odho, Z., Southall, S. M. & Wilson, J. R. Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *J. Biol. Chem.* **285,** 32967–32976 (2010).
62. Avdic, V. *et al.* Structural and biochemical insights into MLL1 core complex assembly. *Structure* **19,** 101–108 (2011).
63. Thomas, L. R. *et al.* Interaction with WDR5 promotes target gene recognition and tumorigenesis by MYC. *Mol. Cell* **58,** 440–452 (2015).
64. Cui, G. *et al.* PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nat. Struct. Mol. Biol.* **19,** 916–924 (2012).
65. Nielsen, P. R. *et al.* Structure of the chromo barrel domain from the MOF acetyltransferase. *J. Biol. Chem.* **280,** 32326–32331 (2005).
66. Han, Z. *et al.* Structural Basis for the Specific Recognition of Methylated Histone H3 Lysine 4 by the WD-40 Protein WDR5. *Mol. Cell* **22,** 137–144 (2006).