BIOINFORMATICS DISCOVERY NOTE Vol. 28 no. 15 2012, pages 1953–1956
doi:10.1093/bioinformatics/bts319

Sequence analysis

Advance Access publication May 29, 2012

A common ancestry for BAP1 and Uch37 regulators

Luis Sanchez-Pulido*, Lesheng Kong and Chris P. Ponting

MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: To reveal how the polycomb repressive–deubiquitinase (PR–DUB) complex controls substrate selection specificity, we undertook a detailed computational sequence analysis of its components: additional sex combs like 1 (ASXL1) and BRCA1-associated protein 1 (BAP1) proteins. This led to the discovery of two previously unrecognized domains in ASXL1: a forkhead (winged-helix) DNA-binding domain and a deubiquitinase adaptor domain shared with two regulators of ubiquitin carboxyl-terminal hydrolase 37 (Uch37), namely adhesion regulating molecule 1 (ADRM1) and nuclear factor related to kappaB (NFRKB). Our analysis demonstrates a common ancestry for BAP1 and Uch37 regulators in PR–DUB, INO80 chromatin remodelling and proteosome complexes.

Contact: luis.sanchezpulido@dpag.ox.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on January 12, 2012; revised on May 18, 2012; accepted on May 23, 2012

1 INTRODUCTION

BRCA1-associated protein 1 (BAP1) and ubiquitin carboxyl-terminal hydrolase 37 (Uch37) (also known as UCHL5) are two deubiquitinating proteases (DUBs) from the same peptidase family that share a high degree of sequence similarity not only in their catalytic ubiquitin carboxy-terminal hydrolase (UCH) domain but also in an additional conserved carboxy-terminal region, previously called the Uch37-like domain (ULD) (Misaghi *et al*., 2009; Yao *et al.*, 2006).

The controlled removal of ubiquitin or ubiquitin-like proteins by DUBs is required in diverse cellular processes. The catalytic domains of DUBs often require adapter domains to control the specificity of substrate selection and thus to avoid cleavage of inappropriate substrates (Komander *et al.*, 2009).

The human PR–DUB complex contains BAP1 and additional sex combs like 1 (ASXL1). The PR–DUB complex represses polycomb group target genes by removing monoubiquitin from histone H2A in nucleosomes. Recently, it was found that the ASXL1 N-terminal region (residues 2–365) is essential for its interaction with BAP1, although the domains responsible for this interaction have been determined in neither ASXL1 nor BAP1 protein sequences (Bott *et al*., 2011; Scheuermann *et al.*, 2010).

The PR–DUB complex has increasingly attracted medical interest as somatic mutations in ASXL1 frequently occur in different myelodysplastic syndromes (Bejar *et al.*, 2011; Boultwood *et al.*, 2010; Carbuccia et *al.*, 2009; Gelsi-Boyer *et al.*, 2009; Hellström-Lindberg, 2010). ASXL1 mutations have been also associated with glioblastoma, ovarian serous cystadenocarcinoma, cervical and colon cancer (Bell *et al.*, 2011; McLendon *et al.*, 2008; Scotto *et al.*, 2008; Williams *et al.*, 2010).

De novo nonsense mutations in human ASXL1 have also been recently associated with Bohring–Opitz syndrome (also known as Oberklaid–Danks or C-like syndromes), characterized by severe intellectual disability, distinctive facial features and multiple congenital malformations (Hoischen *et al.*, 2011).

Human ASXL1 belongs to a family of three vertebrate paralogues (Fisher *et al.*, 2003, 2006; Katoh and Katoh, 2003, 2004) and it is orthologous to Drosophila Asx (additional sex combs), whose product is required for long-term repression of HOX genes during development and whose mutants enhance homeotic phenotypes of both polycomb and trithorax mutations (Breen and Duncan, 1986; Milne *et al.*, 1999; Sinclair *et al.*, 1992). Mouse ASXL1 is required for normal hematopoiesis and its knockout results in simultaneous anterior and posterior transformations of the axial skeleton (Fisher *et al.*, 2010a, b).

In addition, several lines of evidence suggest that ASXL1 has a genuine transcription factor role by regulating retinoic acid receptors and peroxisome proliferator-activated receptors (Cho *et al.*, 2006; Lee *et al.,* 2010; Park *et al.*, 2011). ASX proteins have been previously reported to contain two evolutionarily conserved regions in metazoa, from fly to humans: a region termed the ASX homology (ASXH) domain and a C-terminal plant homeo domain (PHD) zinc finger (Fisher *et al.*, 2003, 2006).

However, with the exception of its C-terminal PHD domain, significant sequence similarity was not detected with any other known protein domain. In order to trace its evolutionary history and to identify functional domains, we embarked on a computational protein sequence analysis of the ASX protein family.

2 RESULTS AND DISCUSSION

2.1 Computational protein sequence analysis. The DEUBiquitinase ADaptor domain

Our iterative profile-based database searches identified previously unreported homologues of ASXL1 in fungi and land plants. Animal ASX HMMer3 (Finn *et al.*, 2011) profile searches (based on Supplementary Figure S1 multiple sequence alignment corresponding to the ASXH conserved region) revealed ASX homologues in fungi, identifying the *Aspergillus fumigatus* (UniProt:Q4WDB3_ASPFU) sequence with an *E*-value (HMMer3-best first domain) of 3.5×10^{-4}.

*To whom correspondence should be addressed.

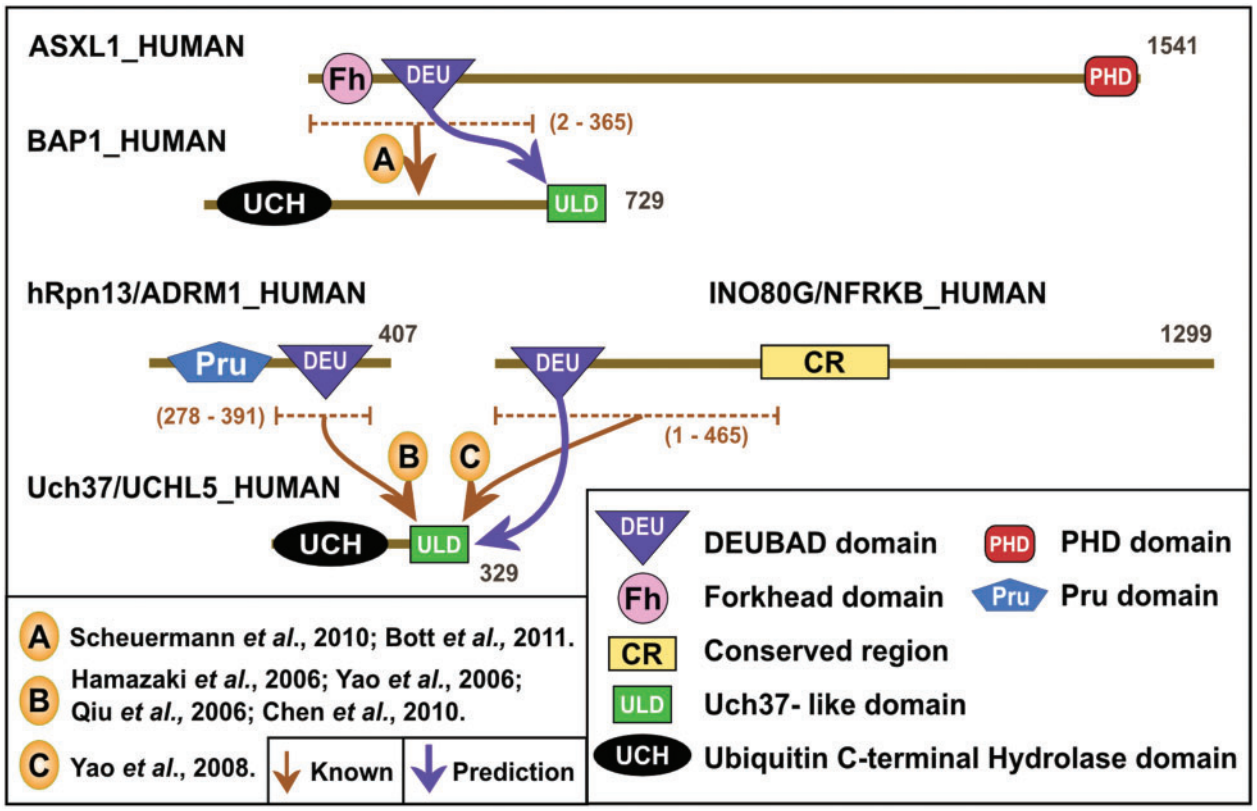

Fig. 1. Domain architecture of DEUBAD domain-containing proteins and their known and predicted interactions with BAP1 and Uch37 DUBs. Experimentally determined interactions are shown with brown arrows corresponding to: A) An ASXLI N-terminal region (amino acids 2 to 365) interacts with full length BAPI (Scheuermann *et al*., 2010; Bott *et al*., 2011); B) The hRpnl3/ADRM l ULD-interacting region was localised to its second domain (amino acids 278-391), here termed its DEUBAD domain (Hamazaki *et al*., 2006; Yao *et al*., 2006; Qiu *et al*., 2006; Chen *et al*., 2010); and C) The Uch37 ULD-interacting region of NFRKB was localised to its first 465 residues (Yao *et al*., 2008). Summary of the experimental evidences supporting each interaction: Scheuermann *et al*., 2010: Reconstituted complex in humans and Affinity Capture-Western for D. *melanogaster* orthologs (ASX and Calypso). Bott *et al*., 2011: Coimmunoprecipitation. Hamazaki *et al*., 2006: Affinity Capture-Western and reconstituted complexes. Yao *et al*., 2006: Yeast two-hybrid confirmed by Affinity Capture-Western and reconstituted complexes. Qiu *et al*., 2006: Affinity Capture-Western. Chen *et al*., 2010: Intermolecular NOE (Nuclear Overhauser Effect) interactions. Yao *et al*., 2008: Anti tag coimmunoprecipitation and reconstituted complexes. Purple arrows indicate predicted interactions between DEUBAD domains and C-terminal ULDs in BAPI and Uch37 DUBs.

These approaches, however, failed to identify potential ASX orthologues in the nematode worm *Caenorhabditis elegans* or in yeasts (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*). A profile derived from animal and fungal ASX sequences identified the *Arabidopsis thaliana* subfamily IV of GATA zinc fingers (Manfield *et al*., 2007): GATA27 (Tair:At5g47140, UniProt: Q5PP38) and GATA26 (Tair:At4g17570, UniProt: Q8W4H1) with *E*-values of 1.6×10^{-4} and 2.3×10^{-4}, respectively.

Consistent with the conservation of the Polycomb H2A ubiquitination machinery in plants (Bratzel *et al.*, 2010; Sanchez-Pulido *et al*., 2008), the finding of ASX and BAP1 homologues in *Arabidopsis* suggests that the histone H2A deubiquitination machinery (PR–DUB complex) could also be conserved between plants and animals.

During a search for more remote evolutionary relationships, we were surprised to find statistically significant similarities with ASXL1 for two further human proteins. HMMer3 profile searches using the ASX family profile (corresponding to Supplementary Figure S1 multiple sequence alignment marked with the red bar) revealed the arthropod *Pediculus humanus* (UniProt: E0VM25_PEDHC) sequence (which is a known orthologue of the human NFRKB protein) with an *E*-value of 0.0073. The ASX/NFRKB profile (based on Supplementary Figure S1 multiple sequence alignment marked with the red and yellow bars) localized the *Trichoplax adhaerens* (UniProt: B3S476_TRIAD) sequence (a known orthologue of the human ADRM1 protein) with an *E*-value of 0.0067. In addition, profile-versus-profile comparisons (using HHpred) (Soding *et al*., 2005) confirm the previously described evidences of sequence similarity among these families: the profile of the ASX family identified the NFRKB and ADRM1 families with *E*-values of 2×10^{-7} and 5×10^{-6}, respectively. Statistically significant *E*-values of sequence similarity connected these human ASXL1,2,3, NFRKB and ADRM1 proteins and reciprocal searches produced convergent results; no further human homologues were identified.

NFRKB is a subunit of the INO80 chromatin-remodeling complex (Conaway and Conaway, 2009; Papamichos-Chronakis *et al*., 2011; Yao *et al*., 2008). The region of NFRKB that we observe to

be homologous to ASXL1 is located within a region previously described as binding the C-terminal Uch37 ULD (Fig. 1). The second human protein homologous to ASXL1 is ADRM1, whose tertiary structure shows it to consist of a bundle of eight alpha-helices (Chen *et al*., 2010). ADRM1 is a proteasome subunit that serves as a receptor for both ubiquitin and Uch37 (Chen *et al*., 2010; Hamazaki *et al*., 2006; Qiu *et al*., 2006; Yao *et al*., 2006). Again, this ADRM1 region is known to bind the C-terminal Uch37 ULD and hence was initially called the Uch37-binding domain (Yao *et al*., 2006). However, owing to its affinity for a specific portion (the ULD of Uch37) and its presence as a DEUBiquitinase ADaptor (DEUBAD) domain in other proteins, we term these DEUBAD domains. The presence of the DEUBAD domain in ASXL1 and its presence within a region known to interact with BAP1 (Scheuermann *et al*., 2010) strongly implicates the ASXL1 DEUBAD domain in binding to the C-terminal BAP1 ULD (Fig. 1).

2.2 Identifying a putative DNA-BINDING domain

Although these observations help to explain how the ASXL1 DEUBAD domain regulates BAP1 deubiquitinase activity within PR–DUB, they fail to explain the proposed transcription factor function of ASXL1 (Cho *et al.*, 2006; Lee *et al*., 2010; Park *et al.*, 2011). To do this, we turned to the region N-terminal of the ASXL1 DEUBAD domain whose sequence is well conserved among diverse metazoa from vertebrates to cnidarians such as *Nematostella vectensis*, although not in *Drosophila melanogaster* (Supplementary Fig. S2).

Using the HHpred server (Soding *et al*., 2005), the profile generated for the conserved ASXL1 N-terminal region matched the forkhead domain of FOXO1 protein (PDB-ID:3coa) (Brent *et al.*, 2008) with a significant *E*-value of 0.028 (estimated error rate <8%). Moreover, in support of the first match, the next five most statistically significant matches are members of the forkhead superfamily of transcription factors (AFX, PDB-ID:1e17, *E*-value = 0.089; FOXK1a, PDB-ID:2c6y, *E*-value = 0.16; FOXO4, PDB-ID:3l2c, *E*-value = 0.35 and; NF3G, PDB-ID:1vtn, *E*-value = 0.68) (Boura *et al.*, 2010; Clark *et al*., 1993; Tsai *et al*., 2006). The consistency of secondary structure predictions and corroboration by profile-to-profile comparison (HHpred) methods, taken together provide strong evidence that animal ASX proteins contain a forkhead domain located at their N-termini.

By inspecting previously characterized forkhead domain structures in interaction with DNA (Brent *et al.*, 2008; Boura *et al.*, 2010; Clark *et al.,* 1993; Stroud *et al.*, 2006; Tsai *et al.,* 2006), we identified two types of positions whose conservation would be in agreement with a DNA-binding role for the ASXL1 forkhead domain (Supplementary Fig. S2): (i) residues that have been implicated in non-specific DNA binding (DNA backbone contacts) are indicated by blue triangles above the alignment and (ii) indicated with black triangles and boxes are two residues located in alpha-helix 3 (Asn51 and His61; human ASXL1) which correspond to Asn165 and His169 of HNF3G (Clark *et al.*, 1993) that are predicted to confer DNA-binding sequence specificity (major groove DNA contacts).

In agreement with a DNA-binding role for the ASXL1 forkhead domain, candidate ASX orthologues from plants contain, rather than a forkhead domain, a different, GATA zinc finger DNA-binding domain at their N-termini (Supplementary Fig. S2) (subfamily IV of *Arabidopsis thaliana* GATA zinc fingers; GATA26 and GATA27) (Manfield *et al.*, 2007).

The human ASXL1 protein, therefore, likely binds DNA through its N-terminal forkhead domain thereby conferring DNA sequence specificity to the PR–DUB complex activity of histone H2A deubiquitination. In a similar manner, a forkhead domain that has recently been identified in the trithorax protein ASH2L appears to link sequence-specific DNA binding to histone H3K4 methylation (Chen *et al*., 2011; Sarvan *et al.*, 2011).

Although we were finalizing this article, a bioinformatic analysis of ASXL1 was published (Aravind and Iyer, 2012), mainly focused on the identification of the N-terminal winged helix-turn-helix/forkhead domain. Our article is consistent with this report, but further extends the ASX family analysis by showing the structural and functional predictions for the DEUBAD domain.

3 CONCLUSION

In summary, the computational identification of the DEUBAD and forkhead domains in the human ASXL1 protein should now greatly assist in experimentally delineating its critical function in regulating the histone code under non-pathological and pathological conditions.

Funding: Medical Research Council UK (to L.S.-P. and C.P.P.) and European Research Council (to L.K.).

Conflict of Interest: none declared.

REFERENCES

Aravind,L. and Iyer,L.M. (2012) The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle*, **11**, 119–131.

Bejar,R. *et al.* (2011) Clinical effect of point mutations in myelodysplastic syndromes. *N. Engl. J. Med.,* **364**, 2496–2506.

Bell,D. *et al.* (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.

Bott,M. *et al.* (2011) The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. *Nat. Genet.*, **43**, 668–672.

Boultwood,J. *et al.* (2010) Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia*, **24**, 1062–1065.

Boura,E. *et al.* (2010) Structure of the human FOXO4–DBD–DNA complex at 1.9 A resolution reveals new details of FOXO binding to the DNA. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 1351–1357.

Bratzel,F. *et al.* (2010) Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr. Biol.*, **20**, 1853–1859.

Breen,T.R. and Duncan,I.M. (1986) Maternal expression of genes that regulate the bithorax complex of *Drosophila melanogaster*. *Dev. Biol.*, **118**, 442–456.

Brent,M.M. *et al.* (2008) Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure*, **16**, 1407–1416.

Carbuccia,N. *et al.* (2009) Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia*, **23**, 2183–2186.

Chen,X. *et al.* (2010) Structure of proteasome ubiquitin receptor hRpn13 and its activation by the scaffolding protein hRpn2. *Mol. Cell*, **38**, 404–415.

Chen,Y. *et al.* (2011) Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep.*, **12**, 797–803.

Cho,Y.S. *et al.* (2006) Additional sex comb-like 1 (ASXL1), in cooperation with SRC-1, acts as a ligand-dependent coactivator for retinoic acid receptor. *J. Biol. Chem.*, **281**, 17588–17598.

Clark,K.L. *et al.* (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, **364**, 412–420.

Conaway,R.C. and Conaway,J.W. (2009) The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem. Sci.*, **34**, 71–77.

Finn,R.D. *et al.* (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.*, **39**, W29–W37.

Fisher,C.L. *et al.* (2003) A human homolog of additional sex combs, ADDITIONAL SEX COMBS-LIKE 1, maps to chromosome 20q11. *Gene*, **306**, 115–126.

Fisher,C.L. *et al.* (2006) Characterization of Asxl1, a murine homolog of additional sex combs, and analysis of the Asx-like gene family. *Gene*, **369**, 109–118.

Fisher,C.L. *et al.* (2010a) Loss-of-function additional sex combs like 1 mutations disrupt hematopoiesis but do not cause severe myelodysplasia or leukemia. *Blood*, **115**, 38–46.

Fisher,C.L. *et al.* (2010b) Additional sex combs-like 1 belongs to the enhancer of trithorax and polycomb group and genetically interacts with Cbx2 in mice. *Dev. Biol.*, **337**, 9–15.

Gelsi-Boyer,V. *et al.* (2009) Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br. J. Haematol.*, **145**, 788–800.

Hamazaki,J. *et al.* (2006) A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO J.*, **25**, 4524–4536.

Hellström-Lindberg,E. (2010) Significance of JAK2 and TET2 mutations in myelodysplastic syndromes. *Blood Rev.*, **24**, 83–90.

Hoischen,A. *et al.* (2011) *De novo* nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat. Genet.*, **43**, 729–731.

Katoh,M. and Katoh,M. (2003) Identification and characterization of ASXL2 gene *in silico*. *Int. J. Oncol.*, **23**, 845–850.

Katoh,M. and Katoh,M. (2004) Identification and characterization of ASXL3 gene *in silico*. *Int. J. Oncol.*, **24**, 1617–1622.

Komander,D. *et al.* (2009) Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.*, **10**, 550–563.

Lee,S.W. *et al.* (2010) ASXL1 represses retinoic acid receptor-mediated transcription through associating with HP1 and LSD1. *J. Biol. Chem.*, **285**, 18–29.

Manfield,I.W. *et al.* (2007) Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family. *Plant Physiol.*, **143**, 941–958.

McLendon,R. *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.

Milne,T.A. *et al.* (1999) The additional sex combs gene of *Drosophila* is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. *Mol. Gen. Genet.*, **261**, 753–761.

Misaghi,S. *et al.* (2009) Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol. Cell Biol.*, **29**, 2181–2192.

Papamichos-Chronakis,M. *et al.* (2011) Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell*, **144**, 200–213.

Park,U.H. *et al.* (2011) Additional sex comb-like (ASXL) proteins 1 and 2 play opposite roles in adipogenesis via reciprocal regulation of peroxisome proliferator-activated receptor {gamma}. *J. Biol. Chem.*, **286**, 1354–1363.

Qiu,X.B. *et al.* (2006) hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. *EMBO J.*, **25**, 5742–5753.

Sanchez-Pulido,L. *et al.* (2008) RAWUL: a new ubiquitin-like domain in PRC1 ring finger proteins that unveils putative plant and worm PRC1 orthologs. *BMC Genomics*, **9**, 308.

Sarvan,S. *et al.* (2011) Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat. Struct. Mol. Biol.*, **18**, 857–859.

Scheuermann,J.C. *et al.* (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature,* **465**, 243–247.

Scotto,L. *et al.* (2008) Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. *Genes Chromosomes Cancer,* **47**, 755–765.

Sinclair,D.A. *et al.* (1992) Genetic analysis of the additional sex combs locus of *Drosophila melanogaster*. *Genetics*, **130**, 817–825.

Soding,J. *et al.* (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, 244–248.

Stroud,J.C. *et al.* (2006) Structure of the forkhead domain of FOXP2 bound to DNA. *Structure,* **14**, 159–166.

Tsai,K.L. *et al.* (2006) Crystal structure of the human FOXK1a–DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. *J. Biol. Chem.*, **281**, 17400–17409.

Williams,D.S. *et al.* (2010) Nonsense mediated decay resistant mutations are a source of expressed mutant proteins in colon cancer cell lines with microsatellite instability. *PLoS One*, **5**, e16012.

Yao,T. *et al.* (2006) Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nat. Cell. Biol.*, **8**, 994–1002.

Yao,T. *et al.* (2008) Distinct modes of regulation of the Uch37 deubiquitinating enzyme in the proteasome and in the Ino80 chromatin-remodeling complex. *Mol. Cell*, **31**, 909–917.