Research

The SWIRM domain: a conserved module found in chromosomal proteins points to novel chromatin-modifying activities
L Aravind and Lakshminarayan M Iyer

Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

Correspondence: L Aravind. E-mail: aravind@ncbi.nlm.nih.gov

Abstract

Background: Eukaryotic chromosomal components, especially histones, are subject to a wide array of covalent modifications and catalytic reorganization. These modifications have an important role in the regulation of chromatin structure and are mediated by large multisubunit complexes that contain modular proteins with several conserved catalytic and noncatalytic adaptor domains.

Results: Using computational sequence-profile analysis methods, we identified a previously uncharacterized, predicted α-helical domain of about 85 residues in chromosomal proteins such as Swi3p, Rsc8p, Moira and several other uncharacterized proteins. This module, termed the SWIRM domain, is predicted to mediate specific protein-protein interactions in the assembly of chromatin-protein complexes. In one group of proteins, which are highly conserved throughout the crown-group eukaryotes, the SWIRM domain is linked to a catalytic domain related to the monoamine and polyamine oxidases. Another human protein has the SWIRM domain linked to a JAB domain that is involved in protein degradation through the ubiquitin pathway.

Conclusions: Identification of the SWIRM domain could help in directed experimental analysis of specific interactions in chromosomal proteins. We predict that the proteins in which it is combined with an amino-oxidase domain define a novel class of chromatin-modifying enzymes, which are likely to oxidize either the amino group of basic residues in histones and other chromosomal proteins or the polyamines in chromatin, and thereby alter the charge distribution. Other forms, such as KIAA1915, may link chromatin modification to ubiquitin-dependent protein degradation.

Background

The structure and dynamics of chromatin clearly differentiate the eukaryotes from the other superkingdoms of life [1,2]. Eukaryotic chromatin undergoes a variety of structural and compositional changes that accompany the progression of the cell cycle and changes in gene expression [3]. These dynamics are regulated both by a variety of protein-protein and protein-DNA interactions and by catalytic modification and reorganization of proteins and nucleic acids that comprise the chromosomes [3]. In recent years it has become clear that covalent modification of histones, transcription factors and other chromosomal proteins plays a major part in the dynamics of chromatin. The covalent modifications include hydroxylation of proline and lysine, methylation of

lysine and arginine, phosphorylation of serine and threonine, and acetylation and ubiquitination of lysine. Other enzymes, such as deacetylases and phosphatases, remove covalent modifications from proteins and thus reverse their effects [4-6]. The covalent modifications may change the local or global charge properties of chromosomal proteins and regulate their interactions with DNA. Additionally, the modifications also seem to form a 'code' that is recognized by specific groups of regulatory proteins [6-8]. Besides covalent modifications, ATP-dependent enzymes such as the Swi2/Snf2 ATPases and other chaperone-like proteins remodel the chromatin by rearranging the binding pattern of histones and other chromosomal proteins [9-12]. These covalent and non-covalent catalytic actions on the chromatin can result in condensation or decondensation of chromatin, either locally or on the chromosomal scale, and thereby regulate access of transcription factors and other proteins to the chromatin.

Many of these chromatin-modifying activities are organized into large multisubunit complexes, with the core enzyme accompanied by several noncatalytic subunits [12]. The subunits of these complexes are characterized by a number of conserved domains that interact with other proteins or DNA. Most of these domains are evolutionarily mobile modules and combine with each other in a very wide range of domain architectures [13]. Well studied examples of these are the bromodomain that interacts with acetylated peptides [14-16], chromodomains that mediate specific interactions with proteins [17,18] and RNA [19,20], the PHD finger mediating protein-protein interactions [21], the Myb and SANT domains [22] that interact with both DNA and proteins [23], and the SAP and AT-hook domains that interact with DNA [24,25]. These domains often occur together in large proteins linked to catalytic domains and may serve to tether these proteins to different components of chromatin and deliver catalytic activities to specific locations.

The computational analysis of chromatin proteins has helped in the identification of a large number of conserved modules in the chromosomal proteins [13]. These studies, followed by biochemical and structural characterization of these modules has thrown considerable light on the biology of chromatin dynamics and the roles these domains have played in course of evolution [14-16,20]. Using computational analysis, we have discovered a previously uncharacterized domain in multiple chromatin proteins. We also show that this domain is found linked to catalytic domains such as oxidoreductase and Jun activating binding protein 1 (JAB) domains and these proteins may define a novel class of chromatin-modifying enzymes.

Results and discussion
Identification of the SWIRM domain
In the course of our systematic survey of eukaryotic chromosomal proteins, we observed a conserved globular module shared by the SWI3p and RSC8p proteins from the yeast *Saccharomyces cerevisiae* that did not map to any previously characterized domain. These homologous proteins are parts of the multisubunit SWI/SNF and RSC complexes that drive chromatin remodeling through the SNF2/SWI2-like ATPase subunit [26-28]. The region of similarity encompassing this segment has been termed 'conserved region 1' in [27]. Both these proteins additionally contain a Myb-related helix-turn-helix domain termed the SANT domain [22] at their carboxyl terminus and a further carboxy-terminal conserved α-helical extension restricted to orthologs of RSC8p and SWI3p. Given that these subunits are key components of the respective chromatin-remodeling complexes, and mediate multiple interactions, we sought to further investigate the provenance of their conserved amino-terminal module using computational methods. A PSI-BLAST search [29] initiated with this region from RSC8p (gi:14318562, region 80-177) not only recovered its paralogs and orthologs from other organisms such as Moira from *Drosophila* [30] and BAF155 and BAF170 from vertebrates [27], but also several other uncharacterized proteins from diverse eukaryotes with statistically significant expected values (*e*-values) at the point of detection. For example, the search recovered the nuclear protein SPAC23E2.02 from *Schizosaccharomyces pombe* [31] in iteration 4 (*e* = 2 x 10^{-4}), the human proteins KIAA1915 (*e* = 10^{-5}) and KIAA0601 [32,33] (*e* = 10^{-7}) in iteration 2 and the *Arabidopsis thaliana* protein At2g47620 in iteration 3 (*e* = 10^{-7}). We prepared a multiple alignment of these regions from all the proteins detected in these searches (Figure 1) and used it to construct a PSI-BLAST position-specific score matrix (PSSM) [34] and a hidden Markov model [35] for further searches against the complete or nearly complete proteome databases of individual eukaryotes. These searches allowed us to detect additional copies of this region in the carboxyl termini of the Asynaptic 1 protein (a Hop1p homolog) from plants [36]. Additionally these searches also recovered a conserved globular region in the extreme carboxyl terminus of the ADA2-like proteins from eukaryotes [37], with moderate significance. While these newly detected proteins have a somewhat distinct conservation pattern from the originally detected set (Figure 1), their secondary structure predicted using the JPred program [38] closely matches that predicted for the originally detected set. This observation, together with the information that the ADA2-like proteins are chromatinic proteins, just like the members of the original set, suggests that these domains are probably distant relatives of each other (Figure 1).

The conserved region described above occurred in various distinct domain-architecture contexts, suggesting that it is an evolutionarily mobile domain (Figure 2). We named this domain the SWIRM domain after the proteins SWI3p, Rsc8p and Moira in which it was first recognized. Secondary-structure predictions, based on the multiple alignment, indicate the presence of four distinct α helices. Hence, it is predicted that the SWIRM domain is likely to form a
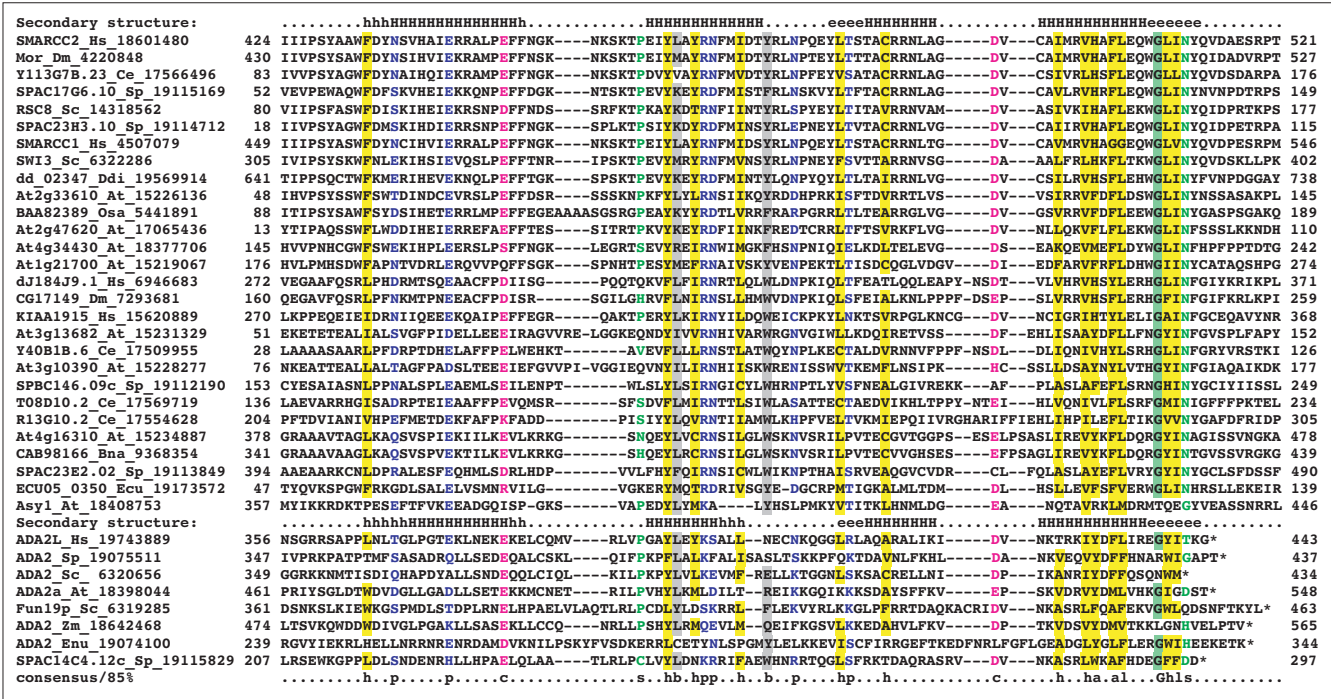
```
Secondary structure:              .........hhhHHHHHHHHHHHHHHHh..............HHHHHHHHHHHHH.......eeeeHHHHHHHH...........HHHHHHHHHHHHHeeeeee.........
SMARCC2_Hs_18601480   424 IIIPSYAAWFDYNSVHAIERRALPEFFNGK----NKSKTPEIYLAYRNFMIDTYRLNPQEYLTSTACRRNLAG-----DV---CAIMRVHAFLEQWGLINYQVDAESRPT 521
Mor_Dm_4220848        430 IIVPSYSAWFDYNSIHVIEKRAMPEFFNSK----NKSKTPEIYMAYRNFMIDTYRLNPQEYLTTACRRNLAG-----DV---CAIMRVHAFLEQWGLINYQIDADVRPT 527
Y113G7B.23_Ce_17566496 83 IVVPSYAGWFDYNAIHQIEKRAMPEFFNGK----NKSKTPDVYVAYRNFMVDTYRLNPFEYVSATACRRNLAG-----DV---CSIVRLHSFLEQWGLLNYQVDSDARPA 176
SPAC17G6.10_Sp_19115169 52 VEVPEWAQWFDFSKVHEIEKKQNPEFFDGK----NTSKTPEVYKEYRDFMISTFRLNSKVYLTFTACRRNLAG-----DV---CAVLRVHRFLEQWGLINYNVNPDTRPS 149
RSC8_Sc_14318562       80 VAIPSFASWFDISKIHEIEKRSNPDFFNDS----SRFKTPKAYKDTRNFIINTYRLSPYEYLTITAVRRNVAM-----DV---ASIVKIHAFLEKWGLINYQIDPRTKPS 177
SPAC23H3.10_Sp_19114712 18 IIVPSYASWFDMSKIHDIERRSNPEFFNGK----SPLKTPSIYKDYRDFMINSYRLEPNEYLTVTACRRNLVG-----DV---CAIIRVHAFLEQWGLINYQIDPETRPA 115
SMARCC1_Hs_4507079    449 IIIPSYASWFDYNCIHVIERRALPEFFNGK----NKSKTPEIYLAYRNFMIDSYRLNPQEYLTSTACRRNLTG-----DV---CAVMRVHAGGEQWGLVNYQVDPESRPM 546
SWI3_Sc_6322286       305 IVIPSYSKWFNLEKIHSIEVQSLPEFFTNR----IPSKTPEVYMRYRNFMVNSYRLNPNEYFSVTTARRNVSG-----DA---AALFRLHKFLTKWGLINYQVDSKLLPK 402
dd_02347_Ddi_19569914 641 TIPPSQCTWFKMERIHEVEKNQLPEFFTGK----SPSKTPEVYKEYRDFMINTYLQNPYQYLTLTAIRRNLVG-----DV---CSILRVHSFLEHWGLINYFVNPDGGAY 738
At2g33610_At_15226136  48 IHVPSYSSWFSWTDINDCEVRSLPEFFDNK----SSSKNPKFYLYLRNSIIKQYRDDHPRKISFTDVRRTLSS-----DV---VSIRRVFDFLDSWGLINYNSSASAKPL 145
BAA82389_Osa_5441891   88 ITIPSYSAWFSYDSIHETERRLMPEFFEGEAAAASGSRGPEAYKYRDTLVRRFRARPGRRLTLTEARRGLVG-----DV---GSVRRVFDFLEEWGLINYGASPSGAKQ 189
At2g47620_At_17065436  13 YTIPAQSSWFLWDDIHEIERREFAEFFTES----SITRTPKVYKEYRDFIINKFREDTCRRLTFTSVRKFLVG-----DV---NLLQKVFLFLEKWGLINFSSSLKKNDH 110
At4g34430_At_18377706 145 HVVPNHCGWFSWEKIHPLEERSLPSFFNGK----LEGRTSEVYREIRNWIMGKFHSNPNIQIELKDLTELEVG-----DS---EAKQEVMEFLDYWGLINFHPFPPTDTG 242
At1g21700_At_15219067 176 HVLPMHSDWFAPNTVDRLERQVVPQFFSGK----SPNHTPESYMEFRNAIVSKYVENPEKTLTISDCQGLVDGV----DI---EDFARVFRFLDHWGIINYCATAQSHPG 274
dJ184J9.1_Hs_6946683  272 VEGAAFQSRLPHDRMTSQEAACFPDIISG------PQQTQKVFLFIRNRTLQLWLDNPKIQLTFEATLQQLEAPY-NSDT---VLVHRVHSYLERHGLINFGIYKRIKPL 371
CG17149_Dm_7293681    160 QEGAVFQSRLPFNKMTPNEEACFPDISR------SGILGHRVFLNIRNSLLHMWVDNPKIQLSFEIALKNLPPPF-DSEP---SLVRRVHSFLERHGFINFGIFKRLKPI 259
KIAA1915_Hs_15620889  270 LKPPEQEIEIDRNIIQEEEKQAIPEFFEGR-----QAKTPERYLKIRNYILDQWEICKPKYLNKTSVRPGLKNCG---DV---NCIGRIHTYLELIGAINFGCEQAVYNR 368
At3g13682_At_15231329  51 EKETETEALIALSVGFPIDELLEEEIRAGVVRE-LGGKEQNDYIVVRNHIVARWRGNVGIWLLKDQIRETVSS-----DV---EF-EHLISAAYDFLLFNGYINYGVSPLFAPY 152
Y40B1B.6_Ce_17509955   28 LAAAASAARLPFDRPTDHELAFFPELWEHKT-------AVEVFLLLRNSLLATWQYNPLKECTALDVRNNVFPPF-NSDL---DLIQNIVHYLSRHGLINFGRYVRSTKI 126
At3g10390_At_15228277  76 NKEATTEALLALTAGFPADSLTEEEIEFGVVPI-VGGIEQVNYILIRNHIISKWRENISSWVTKEMFLNSIPK-----HC--SSLLDSAYNYLVTHGYINFGIAQAIKDK 177
SPBC146.09c_Sp_19112190 153 CYESAIASNLPPNALSPLEAEMLSEILENPT-------WLSLVLSIRNGICYLWHRNPTLYVSFNEALGIVREKK--AF---PLASLAFEFLSRNGHINYGCIYIISSL 249
T08D10.2_Ce_17569719  136 LAEVARRHGISADRPTEIEAAFFPEVQMSR-------SFSDVFLMIRNTTLSIWLASATTECTAEDVIKHLTPPY-NTEI---HLVQNIVLFLSRFGMINIGFFFPKTEL 234
R13G10.2_Ce_17554628  204 PFTDVIANIVHPEFMETDEKFAFPKFADD--------PISIYYLQVRNTIIAMWLKHPFVELTVKMIEPQIIVRGHARIFFIEHLIHPILEFLTIKGVVNYGAFDFRIDP 305
At4g16310_At_15234887 378 GRAAAVTAGLKAQSVSPIEKIILKEVLKRKG-------SNQEYLVCRNSILGLWSKNVSRILPVTECGVTGGPS--ESELPSASLIREVKFLDQRGYINAGISSVNGKA 478
CAB98166_Bna_9368354  341 GRAAAVAAGLKAQSVSPVEKTILKEVLKRKG-------SHQEYLRCRNSILGLWSKNVSRILPVTECVVGHSES---EFPSAGLIREVYKFLDQRGYINTGVSSVRGKG 439
SPAC23E2.02_Sp_19113849 394 AAEAARKCNLDPRALESFEQHMLSDRLHDP--------VVLFHYFQIRNSILCLWIKNPTHAISRVEAQGVCVDR----CL--FQLASLAYEFLVRYGYINVGCLSFDSSF 490
ECU05_0350_Ecu_19173572 47 TYQVKSPGWFRKGDLSALELVSMNRVILG--------VGKERYMQTRDRIVSGYE-DGCRPMTIGKALMLTDM-----DL---HSLLEVFSFVERWGLINHRSLLEKEIR 139
Asy1_At_18408753      357 MYIKKRDKTPESEFTFVKEEADGQISP-GKS------VAPEDYLYMKA----LYHSLPMKYVTITKLHNMLDG-----EA----NQTAVRKLMDRMTQEGYVEASSNRRL 446
Secondary structure:              .........hhhhHHHHHHHHHHHHh...............HHHHHHHHHh............HHHHHHHHhh.......eeeHHHHHHHH...........HHHHHHHHHHHHeeeeee.........
ADA2L_Hs_19743889     356 NSGRRSAPPLNLTGLPGTEKLNEKEKELCQMV----RLVPGAYLEYKSALL--NECNKQGGLRLAQARALIKI-----DV---NKTRKIYDFLIREGYITKG*     443
ADA2_Sp_19075511      347 IVPRKPATPTMFSASADRQLLSEDEQALCSKL----QIFPKPFLALKFALISASLTSKKPFQKTDAVNLFKHL-----DA---NKVEQVYDFFHNARWIGAPT*     437
ADA2_Sc_6320656       349 GGRKKNMTISDIQHAPDYALLSNDEQQLCIQL----KILPKPYLVLKEVMF-RELLKTGGNLSKSACRELLNI-----DP---IKANRIYDFFQSQNWM*        434
ADA2a_At_18398044     461 PRIYSGLDTWDVDGLLGADSLSETEKKMCNET----RILPVHYLKMLDILT--REIKKGQIKKKSDAYSFFKV-----EP---SKVDRVVDMLVHKGIGDST*     548
Fun19p_Sc_6319285     361 DSNKSLKIEWKGSPMDLSTDPLRNELHPAELVLAQTLRLPCDLYLDSKRRL--FLEKVYRLKKGLPFRRTDAQKACRIDV---NKASRLFQAFEKVGWLQDSNFTKYL*  463
ADA2_Zm_18642468      474 LTSVKQWDDWDIVGLPGPGAKLLSASEKLLCCQ----NRLLPSHYLRMQEVLM--QEIFKGSVLKKEDAHVLFKV-----DP---TKVDSVYDMVTKKLGNHVELPTV*  565
ADA2_Enu_19074100     239 RGVYIEKRLHELLNRNRNENRDAMDVKNILPSKYFVSDKERRLCETYNLSPGMYLELKKEVISCFIRRGEFTKEDFNRLFGFLEADGLYGLFLERGWIHEEKETK*  344
SPAC14C4.12c_Sp_19115829 207 LRSEWKGPPLDLSNDENRHLLHPAELQLAA----TLRLPCLVYLDNKRRIFAEWHNRRTQGLSFRKTDAQRASRV---DV---NKASRLWKAFHDEGFFDD*     297
consensus/85%                     .........h..p.....p.....c.................s..hb.hpp..h..b..p.....hp...h............c......h..ha.al...Ghls.........
```

Figure 1

Multiple sequence alignment of the SWIRM domain. Proteins are designated by their gene names followed by the species abbreviations and GenBank (gi) numbers. The coloring represents the conservation profile of amino-acid residues at 90% consensus distinguished by the following amino-acid classes: h, hydrophobic residues (LIYFMWACV (in the single-letter amino-acid code)), a, aromatic residues (FHYW) and l, aliphatic (LIAV) residues, all shaded yellow; c, charged (KERDH) residues (basic (KRH) residues and acidic (DE) residues) colored magenta; p, polar (STEDRKHNQC) residues colored blue; s, small (SACGDNPVT) residues colored green; u, tiny (GAS) residues shaded green (in positions that are always glycine, this is indicated with a G); b, big (LIYERFMWQ) residues shaded gray. The predicted secondary structure is shown above the alignment: H or h, α helix; E or e, β strand. Species abbreviations: At, *Arabidopsis thaliana*; Bna, *Brassica napus*; Ce, *Caenorhabditis elegans*; Ddi, *Dictyostelium discoideum*; Dm, *Drosophila melanogaster*; Ecu, *Encephalitozoon cuniculi*; Hs, *Homo sapiens*; Osa, *Oryza sativa*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Zm, *Zea mays*. A subset of Rsc8p orthologs is represented in the automatically generated, uncurated PFAM-B entry 3680.

globular structure in the form of a tetrahelical bundle (Figure 1). However, a direct comparison showed no specific relationship between the sequence conservation pattern of the SWIRM domain and those of other α-helical bundles that are commonly encountered in chromosomal proteins such as the bromodomain [16], the Myb or SANT domains and other helical DNA-binding domains [22]. The differences in conservation pattern between the distantly related ADA2p carboxy-terminal domains and the canonical SWIRM domains suggest that the former are likely to interact with a very distinct set of partners.

Functional implications and domain architectures of the SWIRM domains

Rsc8p (Swh3p) has been shown to mediate multiple interactions in the RSC complex: it undergoes dimerization via the carboxy-terminal coiled-coil segment, associates with the SWI2/SNF2 ATPase Sth1p by forming two distinct contacts, and it forms a complex with RSC6p subunit. Deletion analyses have shown that the region of the protein that includes the SWIRM domain is probably required for at least one of the contacts with Sth1p and perhaps even those with Rsc6p

[39,40]. This suggests that the SWIRM domain is most likely to mediate protein-protein interactions. The Rsc8p- and SWI3p-related proteins have fairly complex architectures, implying that the different modules may mediate distinct interactions. Rsc8p has an additional ZZ domain [41] between the SWIRM and Myb domains, whereas the animal and *Dictyostelium* versions, like Moira, have an additional amino-terminal chromodomain (Figure 2). However, the common denominator in all these proteins are the SWIRM and Myb domains, suggesting that they mediate the evolutionarily conserved interactions of these proteins, such as those with the Swi2/Snf2 ATPase. The presence of a divergent SWIRM domain, linked to a HORMA domain, in Asy1, the plant ortholog of the yeast Hop1p [36], is consistent with the role of these proteins in mediating protein-protein interactions in the process of chromosome synapsis [42]. The ADA2p-like proteins [37] (Figure 2) are chromatin-associated proteins and are known to mediate protein-protein interactions [23]. They contain a distinct carboxy-terminal globular domain that is distantly related to the SWIRM domain, suggesting that this domain could mediate a subset of their interactions with the transcriptional machinery.
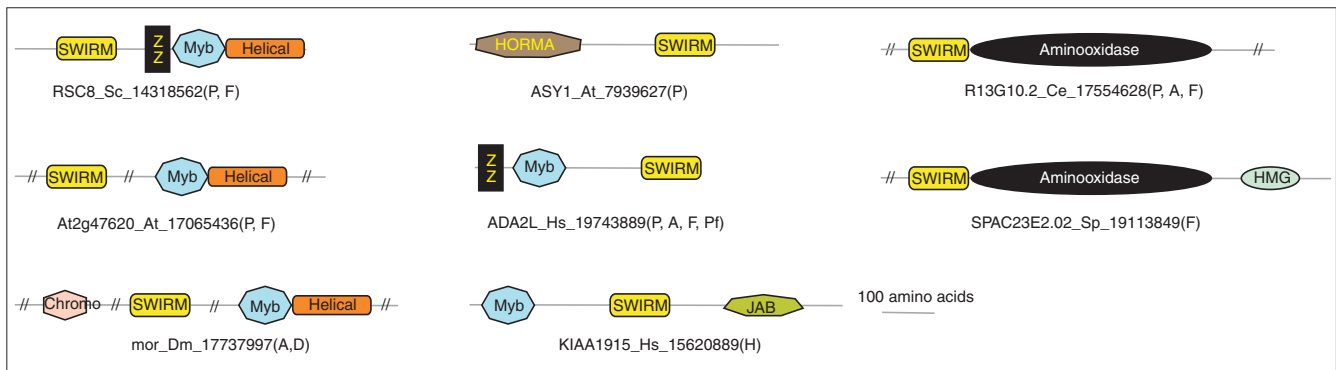
Figure 2

Domain architectures of proteins predicted to have the SWIRM domain. Proteins are designated by their gene names, species abbreviations and gi numbers. The phyletic distribution of a particular domain architecture is additionally given in brackets, where A represents animals; P, plants; F, fungi; D, *D. discoideum* and Pf, *Plasmodium falciparum*. Domains are typically represented by their standard names or abbreviations. ZZ represents the ZZ-type of zinc finger, and 'helical' designates the conserved α-helical domain found at the carboxyl terminus of proteins with a Myb (SANT) and an RSC8-like SWIRM domain.

A striking group of proteins with the SWIRM domain is the one typified by SPAC23E2.02 and SPBC146.09c from *S. pombe*. Between one and four orthologs of this protein are encoded by different crown-group eukaryotes (Figures 1,3) and contain the SWIRM domain linked to a predicted FAD-dependent oxidoreductase domain [43]. A similar observation regarding the presence of a FAD-binding domain has been reported for KIAA0601, a vertebrate protein of this family [32,33]. KIAA0601, along with the co-repressor CoREST, is stably associated with the histone deacetylase complex. Furthermore, SPAC23E2.02 has been shown to be a nuclear protein [31] and along with its paralog SPBC146.09c, contains at its extreme carboxyl terminus an HMG1 domain, which is a common DNA-binding module found in diverse chromosomal proteins [44] (Figures 2,3). The closest relatives of the predicted oxidoreductase domains linked to the SWIRM domain are the polyamine oxidases and the monoamine oxidases (Figure 3). These enzymes are involved in the oxidation of amino groups of polyamines such as spermine and spermidine and monoamines such as dopamine and serotonin [45]. The availability of the crystal structure for the polyamine oxidase from maize allowed us to evaluate the sequence conservation of the oxidoreductase domain fused to the SWIRM domains [45,46]. The alignment of this domain from the SPAC23E2.02-like proteins spans the entire length of the amino oxidase and contains the hallmark residues required to bind the FAD cofactor (Figure 3). These include the

glycine-rich loop bounded by a β-strand and an α helix that is typical of Rossmann fold nucleotide-binding proteins, a conserved glutamate specific to the amino oxidase family, that interacts with ribose in FAD and other residues of the substrate and cofactor-binding site. Thus, these oxidoreductase domains are likely to function as amino oxidases; this, together with their linkage within the same polypeptide to domains typical of chromatin proteins, and evidence for nuclear localization, suggests that they are novel chromatin-modifying enzymes. Consistent with this, the complex containing KIAA0601 has been shown to contain FAD and was suggested to be a chromatin-modifying enzyme [32]. Building on these observations, we suggest two functional possibilities for these proteins: first, they could act as novel protein-modifying enzymes that oxidize the amino groups of lysines or arginines present on histones or transcription factors; alternatively they could affect chromatin structure by oxidizing basic polyamines present in the chromatin and thereby reducing the charge balance.

It is also conceivable that the FAD cofactor of these enzymes functions analogously to NAD, in a deacetylation reaction of acetyl-lysines similar to that carried out by Sir2 enzymes [47]. Whereas both these enzymes are derived versions of the Rossmann fold and bind a dinucleotide cofactor, the SWIRM amino oxidase proteins do not possess equivalents of the unique inserts with residues that allow the SIR2-like proteins to catalyze the deacetylation reaction. On the

Figure 3 *(see the figure on the next page)*

Multiple sequence alignment of the polyamine oxidase domain. Amino-acid residues are colored at 100% consensus. The protein designations, coloring scheme and consensus abbreviations are as described in Figure 1; +, acidic residues. The crystal-structure-derived secondary-structure assignments are shown above the alignment with positions involved in cofactor or substrate binding marked by an asterisk. Inserts in individual sequences are given as brown numbers. Species abbreviations: At, *A. thaliana*; Ce, *C. elegans*; Dm, *D. melanogaster*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Mtu, *Mycobacterium tuberculosis*; Rn, *Rattus norvegicus*; Sp, *S. pombe*; Zma, *Z. mays*.

```
Secondary structure     ...EEEEE..HHHHHHHHHHH......EEEEEE.......EEEEE.......EEEEEE.........HHHHHHH.........................
                                ******                    *       *

PAO_Zma_13096271          3 VGPRVIVVGAGMSGISAAKRLSEAGIT--DLLILEATDHIGRMHKTNF----AGINVELGANWVEGVNGGKMNPIWPIVNSTLKLRNFRSDFDY--LAQNVYKEDGGVY-
C24G6.6_Ce_17558184      30 IKPSIAIVGAGISGLSTARRLIELGID--DFDIYEGLDRIGGRIHAIPY----KDGFLQMGAQFINGA----QNPLYKIANR-LGLLADVVSDTA--HVDNAHFAFGNQ--
PAO1_Hs_14860862         23 GQPRVVVIGAGLAGLAAAKALLEQGFT--DVTVLEASSHIGGRVQSVKL----GHATFELGATWIHGSH---GNPIYHLAEA-NGLLEETTDGER11GVACYLTNHGRRIP
MAOB_Hs_4505093           3 NKCDVVVVGGGISGMAAAKLLHDSGL---NVVVLEARDRVGGRTYTLRN----QKVKYVDLGGSYVGPT----QNRILRLAEK-LGLETYKVNEVE2-IHHVKGKSYPFRGP
At5g13700_At_15240690     2 STASVIIIGAGISGLAAKVLVENGVE--DVLILEATDRIGRHKQNF----GDVPVELGAGWIAGVGGKESNPVWELASR-FNLRTCFSDYTN--ARFNIYDRSGKIFP
Y0B1B.6_Ce_17509955     133 DRRSVIVIGAGAAGISAATQLESFGF---DVIVLEARNCIGGRIHSFKS---KSGEIMETGGDTLRKIED---SPMATLLHQ-VNFEEHGVF-----DFTSVFVE-GRPL-
MaoB_Rn_6981180           3 NKCDVVVVGGGISGMAAAKLLHDCGL---SVVVLEARDCVGGRTYTIR--NKNVKYVDLGGSYVGPT----QNRILRLAEK-LGLETYKVNEV-----ERLIHFVKGKSYA
Fig1_Mm_6753872          57 KPQKVVVVGAGVAGLVAAKMLSDAGH---KVTILEADNRIGGRIFTFRD--EKTGWIGELGAMRMPSS-----HRILHKLCRT-LGLNLTQFTQYD1-NTWTEVHNVK-----
MT3259_Mtu_15842747      12 WTVDVVVVGAGFAGLSAARELTRQGH---EVLVFEARDRVGGRSLTGR----VAGVPADMGGSFIGPT----QDAVLALATE-LGIPTTPTHRD---GRNVIQWRGSARS
R13G10.2_Ce_17554628    309 MRPKIAIIGAGISGISTARHLKHLGI--DAVLFEAKDRFGGRMMDDQS----LGVSVGKGAQIIVGNI---NNPITLLCEQ-IGIKYRNSN-----FFCPLIDENGRCFT
KIAA0601_Hs_3043726     311 KTGKVIIIGSGVSGLAAARQLQSFGM---DVTLLEARDRVGGRVATFRK---GNYVADLGAMVVTGLG---GNPMAVVSKQ-VNMELAKIK-----QKCPLYEANGQAVP
CG17149_Dm_7293681      263 KLGKVVIIGAGISGLAVAHQLQQFGM--DVIVLEARDRVGGRISTFRK----NSYIADVGAMVVTGVY---GNPMTILSKQ-IGMDLVPIQ-----QTCPLYGPDGKPVP
At3g13682_At_15231329   158 TEGSVIVVGAGLAGLAAAARQLLSFGF---KVLVLEGRSRPGGRVYTQKMGGKDRFAAVELGGSVITGLH---GNPLAIQARQ-LSIPLHKVR-----DNCPLYNSEGVLVD
At3g10390_At_15228277   183 SKSSVIIVGAGLSGLAAAARQLMRFGF---KVTVLEGRKRPGGRVYTTKKMEANRVGAAADLGGSVLTGTL---GNPLGIIANA-LGSSLYKVR-----DKCPLYRVDGKPVD
At2g43020_At_15224204    26 RSPSVVIIGGGFGGISAARTLQDASF---QVMVLEGRHTDYS----FGFPVDLGASWLHGVCK--ENPLAPVIGR-LGLPLYRTSGDN7-ESYALFDMDGNQVP
At1g65840_At_15218830    27 GAPSVVVIGSGISGLAAARNLSEASF---KVTVLEGRHTDYS----FGCPVDMGASWLHGVSD--ENPLAPIIRR-LGLTLYRTSGDD7-ESYGLFDMHGNKIP
SPBC146.09c_Sp_19112190 256 SQKTVAIIGAGMAGISCARQLTNLFAQ13RIVIYEASERLGGHIYTHMVP-LSDNEVSEKSSLATTVNAT--NECMVNLLTDSLIGMPTLDSD-----PLYIISSQQL---
consensus/100%              ...plhllGuGhuGhssAp.L.p......ph.lbEu.pp.GG+........uu......p.h..l..........................................


PAO_Zma_13096271          --------DEDYVQKRIELADSVEEMGEKLSATLHASGRDDMSILAMQRLNEH5--ATPVDMVVDYYKFDYEF--AEPP--RVTSLQNTVPLATF--------------S
C24G6.6_Ce_17558184       ----------NVQEKDIKTFLDFTSKLDPKYRSIAKHDEKTARRYTFKEIFTLD14-TNVFDSLARSFRSYWEF-----EWAADWSTLSVHVLKE---------------WND
PAO1_Hs_14860862          ------KDVVEEFSDLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTREEVRNRI5--DPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGE------------WTE
MAOB_Hs_4505093           FP----PVWNPITYLDHNNFWRTMDDMGREIPSDAPWKAPLAEEWDNMTMKELL3--CWTESAKQLATLFVNLC------VTAETHEVSALWFLWYVKQCGGTTRIISTTN
At5g13700_At_15240690     ------------TGIASDSYKKAVDSAILKKLSLEAQCSGQVAEEAPSSP-------KTPIELAIDFILHDFEM-----------AEVEPISTYV--------------
Y0B1B.6_Ce_17509955       ------NEEKIHLFLDHYKSAHGALNYQAHQCEHRDDQGSFISRQQAYENLLSM97-KQYMHPGSEFATFNFML---GFEEYLVGAQLEKVQFSC----------DSMQNK
MaoB_Rn_6981180           FRGPFPPVVWNPITYLDYNNLWRTMDEMGQEIPSDAPWKAPLAEEWDYMTMKELL3-CWTNSTKQIATLFVNLC------VTAETHEVSALWFLWYVKQCGGTTRIISTTN
Fig1_Mm_6753872           ----------LRNYVVEKMPEKLGYLNNRERGHSPEDIYQMALNKAFKDLKAL20-EGNLSRPAVQLLGDVMS----EEGFFYLSFAEALRAH-------------S
MT3259_Mtu_15842747       YRGTI-PKLSLTGLIDIGRLRWQFERIARGVPVAAPWDARRARELDDVSLGEWL51-SSSRNLMAIMTRVTW--------GCEPDDVSMLHAARYVRAAGGLDRLLDV-K
R13G10.2_Ce_17554628      LER---KELDDQVDLHYNNVLDAIRRNKYQSDRNFPDVPLEVTNFRHFTEMFSKM9--LDSLYTPEFEKLLDFHL--GNLEF-SCGTHVSNLSAKDYDHNEKFG-------N
KIAA0601_Hs_3043726       -------KEKDEMVEQEFNRLLEATSYLSHQL-DFNVLNNKPVSLGQALEVVIQL98-SDVYLSSRDRQILDWHF--ANLEF-ANATPLSTLSLKHWDQDDDF--------E
CG17149_Dm_7293681        -------KEKDDVIEREFNRLLESASYLSHRL-DFNYAGDCPVSLGDALEWIISM105SDVYLSSRDRLILDWHF--ANLEF-ANATRLNNLSLKHWDQDDDF--------E
At3g13682_At_15231329     -------KVADSNVEFGFNRLLDKVTE-----VREMMEGAAKKISLGEVLETLRVL---VGVAKDSEERKLFDWHL--ANLEY-ANAGCLSNLSAAYWDQDDPY--------E
At3g10390_At_15228277     -------PDVDIKVEVAFNQLLDKASK-----LRQLMGDVSMDVSLGAALETFRQV---SGNDVATEEMGLFNWHL--ANLEY-ANAGLVSKLSLAFWDQDDPY--------D
At2g43020_At_15224204     -------QELVTQIGVTFERILEEINL-----VRDEQDADISISQAFSIVFSRK------PERLLEGLAHNVLQNY--CRMEG-WFAADAEISKWCDQDE----------L
At1g65840_At_15218830     -------PQLVTKVGDAFKRILEETEK-----IRDETANDMSVLQGISIVLDR1--PELRQEGMAYEVLQWYL--CRMEA-WFAVDANLISLKCWDQDE----------C
SPBC146.09c_Sp_19112190   ------------SLDAVHTRNREFILHDIENGRIDTEHVQRIFRLFDALLFYFN55-SLLHLTVLDLKIFEWF-----KEYLSQSLSVSLENVYPGSIPNLN--------L
consensus/100%            ...........................................................................................................


Secondary structure     .HHHHHHHH...EEEE..EEEE.............EEEEE.........EEEE..EEEEE..HHHHHHH....EEEE..HHHHHHHH..EEEEEEEEEEEE...
                                                                                                 *                    *

PAO_Zma_13096271          YEAVVYYLA11KIVDP-RLQLNKVVREIKYSPGG-----VTVKT------EDNSVYSADYVMVSASLGVLQSD---LIQFKPKLPTWKVRAIYQFDMAVYTKIFLKFPRKF
C24G6.6_Ce_17558184       FKAILDDIA--APIPRKAFNFNSRVENINLDSNT--GKIKLTVS------DRAVPTEYDYIIVTSSLGVLKKY---HHKMFTPPLPRQKIEAIEKIGFGGSCKVFFEWETPF
PAO1_Hs_14860862          FMRVVELLA--EGIPAHVIQLGKPVRCIHWDQAS40WSVVVEC------EDRELIPADHVIVTVSLGVLKRQ---YTSFFRPGLPTEKVAAIHRLGIGTTDKIFLEFEEPF
MAOB_Hs_4505093           SGQVSERIM---DLLGDRVKLERPVIYIDQTREN----VLVET------LNHEMYEAKYVISAIPPTLGMKI------FHFNPPLPMMRNQMITRVPLGSVIKCIVIYKEPF
At5g13700_At_15240690     YECLLYKMA8-GNILDYRLKLNQVVREVQQSRNG----VVVKT------EDGSVYEANYVISASIGVLQSD-----LLSFQPLLPRWKTEAIQKCDVMVYTKIFLKFPQCF
Y0B1B.6_Ce_17509955       IAELLTQLS2-RKLD----IRLKHRVLDIDYSGFE---HVLLKVQREN--GDIEEMKAAFVVSTLPIGVLKKTII6RAPTFTPSLPDKKVEAINGIGCGSVNKCILEFDRVF
MaoB_Rn_6981180           SGQVSERI---KDILGDRVKLERPVIHIDQTGEN----VVVKT------LNHEIYEAKYVISAIPPVLGMKI------HHSPSPLPILRNQLITRVPLGSVIKCMVYYKEPF
Fig1_Mm_6753872           WDLLPRALL--SSLSG-ALLLNAPVVVSITQGRND----VRVHIATSL--HSEKTLTADVVLLTASGPALQRI-------TFSPPLTRKRQEALRALHYVAASKVFLSFRRPF
MT3259_Mtu_15842747       TQQIAQAAA--AQLGA-RVLLNAAVRRIDRHGAG----VTVT------SDQGQAEAGYIVVAIPPAHRVAI-----CFDPPLPPEYQQLAHHWPQGRLSKAYAYSPTPF
R13G10.2_Ce_17554628      AQRIIDFLA--TGLD----IRLNCPVKCIDWGRDD--RRKVKIFFEN----AEQAAEEFDKVVITTSLVLKSN----HSKMFVPPLPIEKQKAIDDLGAGLIEKIAVKFDRRF
KIAA0601_Hs_3043726       YSCVPVALA--EGLD----IRLKHRVIRIDQNGS----VTIVAIPPTV--LNSGDVIVTGVLKK----CEVQRMVPPLPEWKTSAVQRMGFGNLNKVVLCFDRVF
CG17149_Dm_7293681        YSCVPVALT--ENLD----IRVNSAVKEIKYGTKG---VEVVAENLKTSNSQMTYKADLVVCTLTLGVLKVAVA7NTVKFDPPLPDWKQQAIKRLGFGNLNKVVLCFDRIF
At3g13682_At_15231329     NWRLINALA--EGLP---IIYGKSVQTIKYGDDG----VEVI--------SGSQIFQADMILCTVPLGVLKKR---SIKFEPSLPRRKQAAIDRLGFGLLNKVAMLFPSVF
At3g10390_At_15228277     NGRLVQALA--ENVP---ILYEKTVQTIRYGSNG----VVVTA------GNQVYEGDMVLCTVPLGVLKNG----SIKFVPELPQRKLDCIKRLGFGLLNKVAMLFPVVF
At2g43020_At_15224204     YRPVINTLA--KGLD----IRVGHRVTKIVRRYNG---VKVTTE------NGQTFVADAAVIAVGVGLKSG----TIKFEPKLPEWKQEAINDLGVGIENKIILHFEKVF
At1g65840_At_15218830     YEPVIRTIA--KGLD----IRVGHRVSYACGYFLNLH-KVIVAVE------GGTNFVADAVIITVPIGVLKAN--LIQFEPEQWENKTSAISGLGVGNENKIALRFDRAF
SPBC146.09c_Sp_19112190   MADMLNSLA3-SPLP---ILFDQCVHTVKLEDNT----VNLSF------VNETTVSVDKVVICIPMDKLNTH----LITFEPPLEEKKLKAIDRCHFTNVKVILIFKTQF
consensus/100%            ...h...h.....l...h.h...V..l.b............h.l............hh.s.s.s......a.P.L...b..hh..h.......Khhh.a...F


Secondary structure     ..........EEEEE....EEEE...............EEEEEEEhhhhhh...HHHHHHHHHHHHHH.............EEEEEE.......
                         *                                                                                  *

PAO_Zma_13096271          WPEGKGR----EFFLYASSRRGYYGVWQEFE-------------KQYPDANVLLVTVTDEESRRIEQQSDEQTKAEIMQVLRKMFPGK--DVPDATDILVPRWWSDRFYK
C24G6.6_Ce_17558184       WSNNTYSIAPLPVKGMIRDKLDAFEDETTIL---------QVVDWAPNVLSAWYAGRGHQLVDNMSEEELKQRITKLMRDMYNDK--SIPEPSKIIRTQLTKNELLL
PAO1_Hs_14860862          WGPECNS------LQFVWEDEAESHTLTYPPELWYRKICGFDVLYPPERYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFTGNP--NIPKPRRILRSAWGSNPYFR
MAOB_Hs_4505093           WRKKDYC-----GTMIIDGEEAPVAYTL------------DDTKPENYAAIMGFILAHKARKLARLTKEERLKKLCELYAKVLGSL--EALEPVHYEEKNWCEEQYSG
At5g13700_At_15240690     WPCGPGQ-----EFFIYAHEQRGYFTFWQH--------------MENAYPGSNILVVTLTNEQSKRVEAQSDQETMKEAMSVLRDMFGA---TIPYATDILVPRWWNNRFQR
Y0B1B.6_Ce_17509955       WTANGGR----NQFVTVSGNSMNI---------------WSSVPGSKVLCTYIVGEEA--MLELPDDVIIQNAMINLQKAFGNN--CPRAPISAHITRWHDDELAF
MaoB_Rn_6981180           WRKKDFC-----GTMVIEGEEAPIAYTL------------DDTKPDAGCAAIMGFILAHKARKLVRLTKEERLKKLCELYAKVLNSQ--EALQPVHYEEKNWCEEQYSG
Fig1_Mm_6753872           WHEEHIE------GGHSNTDRPSRLIFY---------------PARGEGSLLLASYTWSDAAAPFAGLSTDQTLRLVLQDVAHLQD--FGPVVFFRL-WDGRGVVKRWAEDPHSQ
MT3259_Mtu_15842747       WRAS----------GYSGQALSDEAPVFITF---------DVSPHADGPGILMGFVDAR----GFDSLPIEERRRDALRCFASLFGD----EALDPLDYVDYRWGTEEFAP
R13G10.2_Ce_17554628      WDTVDADGLRTEYFGKVSDCKTDRSLFNIFYDFS--------GKDPNGEDTFVLMSYVTAEHVNLVNVLTESEVADKFCATLRKMFPS----AVINPLGHMMSHWGDARFVG
KIAA0601_Hs_3043726       WDPSVNL-------FGHVGSTTASRGELFLFW---------------LYKAPILLALVAGEAAGIMENISDDVIVGRCLAILKGIFGSS--AVPQPKETVVSRWRADPWAR
CG17149_Dm_7293681        WDPNANL-------FGHVGSTTASRGEMFLFW---------------SISSSPVLLALVAGMAANLVESVTDDIIIGRCMSVLKGIFGST--SVPQPKETVVSRWRSDPWAR
At3g13682_At_15231329     WGDELDT------FGCLNESSINRGEFFLFYA-----------YHTVSGGPALVALVAGEAAQRFECTEPSVLLHRVLKKLRGIYGPKGVVVTPDIQTVCTRWGSDPFSL
At3g10390_At_15228277     WSTDLDT------FGHLTEDPNYRGEFFLFYSY---------APVAGGALIALVAGEAAHKFETMPPTDAVTRVLHILRGIYEPQGINVPDPLQTVCTRWGGDPFSL
At2g43020_At_15224204     WPKV-------EFLGVVAETSYGCSYFLNLH--------------KATGHPVLVYMPAGQLAKDIEKMSDEAAANFAVLQLQRILP----DALPPVQYLVSRWGSDVNSM
At1g65840_At_15218830     WPNV--------EFLGMVAPTSYACGYFLNLH---------------KATGHPVLVYMAAGNLAQDLEKLSDEATANFVMLQKMFFP----DAPDPAQYLVTRWGTDPNTL
SPBC146.09c_Sp_19112190   WEPNISI------FGSLPQDSGRNFIFNDC---------------TRFYEHPTLSVFVKVE---GIDFMKDDDIVNGIVSQLKKVKPKS-EAINPIRTIISNWENNSYTN
consensus/100%            W.........................................l...s.......h....p.....h..h..h........s......h..p....


Secondary structure     ..............hhhh.HHHHHH...EEEEE.........HHHHHHHHHHHHHHHHHHHH.
                                                     *

PAO_Zma_13096271          GTFSNWPV--GVNR-YEYD-QLRAPV--GRVYFTGEHTSEHYNGYVHGAYLSGIDSAEILINCA 458  \Poly- and mono amine oxidases
C24G6.6_Ce_17558184       GSYSYMTQVQALSH-ISHS-QLAIPVK4-PKVLFAGEATHHRLFQTTIGGYLSGREEDRAVNDW 493  |
PAO1_Hs_14860862          GSYSYTQV--GSSG-ADVE-KLAKPLP9-MQVLFSGEATHRKYYSTTHGALLSGQREAARLIEMY 547  |
MAOB_Hs_4505093           GCYTTYFP--PGIL-TQYGRVLRQPV---DRIFFAGTETATHWSGYMEGAVEAGERAAREILNAL 454  |
At5g13700_At_15240690     GSYSNYPM--ISDN-QLLQ-NIKAPV--GRIFFTGEHTSEKFSGYVHGGVLAGIDTSKSLLEEM 433  |
Y0B1B.6_Ce_17509955       GSGAFMSL--RTET-TSFD-DVMEPLK5-SRVYFAGEHTCSSYTSTIQGAWMSGARAAADISNDH 680  |
MaoB_Rn_6981180           GCYTAYFP--PGIL-TQYGRVLRQPV---DRIFFAGTETATHWSGYMEGAVEAGERAAREILHAI 454  |
Fig1_Mm_6753872           GGFVVQPPLYGREA-EDYD--WSAPF---GRIYFAGEHTALPH-GWVETAVKSGLRAAVRINNNY 506  |
MT3259_Mtu_15842747       GGPTAAVP--PGSW-TKYGIWAEYM----GPIHWASTEDATWTGYFDGAVNSGQRAAAEVAALL 454  /
R13G10.2_Ce_17554628      MSYTFVPF--GSDGDATYN-QLKKSID--EKLYFAGEHTIAAEPQTMAGAYISGLREAGQIVMSL 775  \SWIRM-associated oxidase domains
KIAA0601_Hs_3043726       GSYSYVAA--GSSG-NDYD-LMAQPIT12PRLFFAGEHTIRNYPATVHGALLSGLREAGRIADQF 863  |
CG17149_Dm_7293681        GSYSYVSV--GSSG-SDYD-LLAAPVI10PRLFFAGEHTIRNYPATVHGAYLSGLREAGRIADYY 828  |
At3g13682_At_15231329     GSYSHVRV--GSSG-VDYD-ILAESVS--NRLFFAGEATTRQHPATMHGAYLSGLREASKLLHVA 597  |
At3g10390_At_15228277     GSYSNVAV--GASG-DSPYD-ILAESVG1-GRLFFAGEATTHLAPTIHGAYLSGLREAANMAQSA 623  |
At2g43020_At_15224204     GSYSYDIV--GKPH-DLYE-RLRVPVD---NLFFAGEATSSSFPGSVHGAYSTGLMAAEDCRMRV 461  |
At1g65840_At_15218830     GCYAYDVV--GMPE-DLYP-RLGVPVD---NIFFAGEAVTHGLPALFAGVSAGSNCQRYI 463  |
SPBC146.09c_Sp_19112190   HSSYQISN--LFLE-EDYA-ILSEPID--NTVFFASEAISQKNSGSIRGAFDSGILAARDVLASL 758  /
consensus/100%            .s...................h..sh.....l.asup..........suh.sG...u..h....
```

Figure 3 *(see the legend on the previous page)*

contrary, the conserved residues in the SWIRM amino oxidases are the same as other amino oxidases, suggesting that they share similar catalytic activities, as proposed above.

The combination with the SWIRM domain, which is predicted to be a protein-protein interaction module, suggests that these enzymes may be part of a larger complex, like other chromatin-modifying enzymes. Whereas in vertebrates KIAA0601 has been shown to interact with the histone-deacetylase-containing complexes [32,33], the presence of multiple proteins in this family could point to the formation of other distinct complexes. Interestingly, the SWIRM amino-oxidase-type proteins are absent in *S. cerevisiae* but present in *S. pombe* and other crown-group eukaryotes. Hence, it is likely to have been present in the ancestral crown-group eukaryote and secondarily lost in *S. cerevisiae*. We have shown previously that several functionally linked genes, encoding proteins involved in chromatin-structure dynamics, which are conserved in *S. pombe* and other crown-group eukaryotes, have been lost as a group in *S. cerevisiae* [43]. This implies that the SWIRM amino-oxidase-type proteins may be functionally linked to these other genes that were co-eliminated along with it in *S. cerevisiae*. These include genes for proteins such as the SET domain methyltransferase Clr4p, the chromodomain protein Swi6p, the PHD finger protein Mlo2p, the chromosomal actin-like protein SPAC23D3.09, and the predicted prolyl-hydroxylase with the double-stranded β-helix domain SPAC343.11c [43]. Some of these proteins may be part of a multiprotein chromatin-modifying catalytic complex of which the SWIRM amino oxidase proteins are a part.

In vertebrates, the SWIRM domain is found fused along with a Myb (SANT) domain to a JAB1/PAD1 domain [48,49] (for example, in the human protein KIAA1915, Figure 2). This latter domain is commonly associated with several proteasomal and signalosomal proteins and is involved in ubiquitin-mediated protein degradation [50]. In this case, the SWIRM domain may serve to recruit proteasomal complexes to specific chromatin proteins to effect their degradation.

Conclusions
We define a conserved domain of about 85 residues, predicted to participate in protein-protein interactions, in different eukaryotic chromatin proteins such as Swi3p and Rsc8p. Homologs of these molecules, with the SWIRM domain, are found in all eukaryotes belonging to the crown group, as well as earlier-branching protists such as the apicomplexans. A version of the SWIRM domain is found linked to an amino-oxidase domain in a class of nuclear proteins that are represented in most crown-group eukaryotes. These proteins are present in multiple copies in plant proteomes and are entirely absent in the yeast *S. cerevisiae*. We predict that these proteins define a new class of chromatin-modifying enzymes that are likely to oxidize the amino groups of histones or other

nuclear proteins. Alternatively, they may oxidize polyamines in chromatin to alter the charge balance in the chromatin. In humans the SWIRM domain is found linked to a JAB1/PAD1 catalytic domain, suggesting that this protein may serve as a link in the regulation of chromatinic proteins through proteasomal degradation.

Materials and methods
The nonredundant (NR) database of protein sequences (National Center for Biotechnology Information, NIH, Bethesda) was searched using the BLASTP program. Profile searches were carried out using the PSI-BLAST program [29] with either a single sequence or an alignment used as the query, with a profile inclusion expectation (E) value threshold of 0.01, and were iterated until convergence. Previously known conserved protein domains were detected using the corresponding PSI-BLAST-derived PSSMs [34]. The PSSMs were prepared by choosing one or more starting queries (seeds) for a set of most frequently encountered domains (see [34] for details) and run against the NR database until convergence with the -C option of PSI-BLAST to save the PSSM. It was ensured that at convergence no false positives were included in the profiles. This profile database can be downloaded from [51] or used on the Internet via the RPS-BLAST program [52]. All globular segments of proteins that did not map to domains with previously constructed PSSMs were searched individually using PSI-BLAST to detect any additional domains that may have been overlooked. Hidden Markov model-based searches were run using the HMMER2 package [35].

Multiple alignments were constructed using the T_Coffee program [53], followed by manual correction based on the PSI-BLAST results. Protein secondary structure was predicted using a multiple alignment as the input for the JPRED program [38].

References
1. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD: In *Molecular Biology of the Cell*. New York and London: Garland Publishing; 1994.
2. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell JE: In *Molecular Cell Biology*. New York: WH Freeman; 1999.
3. Lamond AI, Earnshaw WC: **Structure and function in the nucleus.** *Science* 1998, **280:**547-553.
4. Kouzarides T: **Acetylation: a regulatory modification to rival phosphorylation?** *EMBO J* 2000, **19:**1176-1179.
5. Grant PA: **A tale of histone modifications.** *Genome Biol* 2001, **2:**reviews0003.1-0003-6.
6. Jenuwein T, Allis CD: **Translating the histone code.** *Science* 2001, **293:**1074-1080.
7. Strahl BD, Allis CD: **The language of covalent histone modifications.** *Nature* 2000, **403:**41-45.
8. Turner BM: **Histone acetylation and an epigenetic code.** *BioEssays* 2000, **22:**836-845.
9. Fry CJ, Peterson CL: **Chromatin remodeling enzymes: who's on first?** *Curr Biol* 2001, **11:**R185-R197.
10. Sudarsanam P, Winston F: **The Swi/Snf family nucleosome-remodeling complexes and transcriptional control.** *Trends Genet* 2000, **16:**345-351.

11. Wade PA, Wolffe AP: **Transcriptional regulation: SWItching circuitry.** *Curr Biol* 1999, **9:**R221-R224.

12. Muchardt C, Yaniv M: **ATP-dependent chromatin remodelling: SWI/SNF and Co. are on the job.** *J Mol Biol* 1999, **293:**187-198.

13. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.

14. Jacobson RH, Ladurner AG, King DS, Tjian R: **Structure and function of a human TAFII250 double bromodomain module.** *Science* 2000, **288:**1422-1425.

15. Syntichaki P, Topalidou I, Thireos G: **The Gcn5 bromodomain co-ordinates nucleosome remodelling.** *Nature* 2000, **404:**414-417.

16. Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM: **Structure and ligand of a histone acetyltransferase bromodomain.** *Nature* 1999, **399:**491-496.

17. Jacobs SA, Khorasanizadeh S: **Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail.** *Science* 2002, **295:**2080-2083.

18. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI: **Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.** *Science* 2001, **292:**110-113.

19. Koonin EV, Zhou S, Lucchesi JC: **The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin.** *Nucleic Acids Res* 1995, **23:**4229-4233.

20. Akhtar A, Zink D, Becker PB: **Chromodomains are protein-RNA interaction modules.** *Nature* 2000, **407:**405-409.

21. Aasland R, Gibson TJ, Stewart AF: **The PHD finger: implications for chromatin-mediated transcriptional regulation.** *Trends Biochem Sci* 1995, **20:**56-59.

22. Aasland R, Stewart AF, Gibson T: **The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB.** *Trends Biochem Sci* 1996, **21:**87-88.

23. Sterner DE, Wang X, Bloom MH, Simon GM, Berger SL: **The SANT domain of Ada2 is required for normal acetylation of histones by the yeast SAGA complex.** *J Biol Chem* 2002, **277:**8178-8186.

24. Aravind L, Koonin EV: **SAP - a putative DNA-binding motif involved in chromosomal organization.** *Trends Biochem Sci* 2000, **25:**112-114.

25. Aravind L, Landsman D: **AT-hook motifs identified in a wide variety of DNA-binding proteins.** *Nucleic Acids Res* 1998, **26:**4413-4421.

26. Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, Erdjument-Bromage H, Tempst P, Du J, Laurent B, Kornberg RD: **RSC, an essential, abundant chromatin-remodeling complex.** *Cell* 1996 **87:**1249-1260.

27. Wang W, Xue Y, Zhou S, Kuo A, Cairns BR, Crabtree GR: **Diversity and specialization of mammalian SWI/SNF complexes.** *Genes Dev* 1996, **10:**2117-2130.

28. Peterson CL, Dingwall A, Scott MP: **Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement.** *Proc Natl Acad Sci USA* 1994, **91:**2905-2908.

29. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.

30. Crosby MA, Miller C, Alon T, Watson KL, Verrijzer CP, Goldman-Levi R, Zak NB: **The trithorax group gene moira encodes a brahma-associated putative chromatin-remodeling factor in *Drosophila melanogaster*.** *Mol Cell Biol* 1999, **19:**1159-1170.

31. Ding DQ, Tomita Y, Yamamoto A, Chikashige Y, Haraguchi T, Hiraoka Y: **Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library.** *Genes Cells* 2000, **5:**169-190.

32. Humphrey GW, Wang Y, Russanova VR, Hirai T, Qin J, Nakatani Y, Howard BH: **Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1.** *J Biol Chem* 2001, **276:**6817-6824.

33. Ballas N, Battaglioli E, Atouf F, Andres M, Chenoweth J, Anderson M, Burger C, Moniwa M, Davie JR, Bowers WJ, *et al.*: **Regulation of neuronal traits by a novel transcriptional complex.** *Neuron* 2001, **31:**353-365.

34. Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L, Altschul SF: **IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices.** *Bioinformatics* 1999, **15:**1000-1011.

35. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14:**755-763.

36. Caryl AP, Armstrong SJ, Jones GH, Franklin FC: **A homologue of the yeast *HOP1* gene is inactivated in the *Arabidopsis* meiotic mutant *asy1*.** *Chromosoma* 2000, **109:**62-71.

37. Candau R, Moore PA, Wang L, Barlev N, Ying CY, Rosen CA, Berger SL: **Identification of human proteins functionally conserved with the yeast putative adaptors ADA2 and GCN5.** *Mol Cell Biol* 1996, **16:**593-602.

38. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ: **JPred: a consensus secondary structure prediction server.** *Bioinformatics* 1998, **14:**892-893.

39. Treich I, Carlson M: **Interaction of a Swi3 homolog with Sth1 provides evidence for a Swi/Snf-related complex with an essential function in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1997, **17:**1768-1775.

40. Treich I, Ho L, Carlson M: **Direct interaction between Rsc6 and Rsc8/Swh3, two proteins that are conserved in SWI/SNF-related complexes.** *Nucleic Acids Res* 1998, **26:**3739-3745.

41. Ponting CP, Blake DJ, Davies KE, Kendrick-Jones J, Winder SJ: **ZZ and TAZ: new putative zinc fingers in dystrophin and other proteins.** *Trends Biochem Sci* 1996, **21:**11-13.

42. Aravind L, Koonin EV: **The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair.** *Trends Biochem Sci* 1998, **23:**284-286.

43. Aravind L, Watanabe H, Lipman DJ, Koonin EV: **Lineage-specific loss and divergence of functionally linked genes in eukaryotes.** *Proc Natl Acad Sci USA* 2000, **97:**11319-11324.

44. Baxevanis AD, Landsman D: **The HMG-1 box protein family: classification and functional relationships.** *Nucleic Acids Res* 1995, **23:**1604-1613.

45. Binda C, Coda A, Angelini R, Federico R, Ascenzi P, Mattevi A: **A 30-angstrom-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase.** *Structure Fold Des* 1999, **7:**265-276.

46. Binda C, Angelini R, Federico R, Ascenzi P, Mattevi A: **Structural bases for inhibitor binding and catalysis in polyamine oxidase.** *Biochemistry* 2001, **40:**2766-2776.

47. Sauve AA, Celic I, Avalos J, Deng H, Boeke JD, Schramm VL: **Chemistry of gene silencing: the mechanism of NAD⁺-dependent deacetylation reactions.** *Biochemistry* 2001, **40:**15456-15463.

48. Aravind L, Ponting CP: **Homologues of 26S proteasome subunits are regulators of transcription and translation.** *Protein Sci* 1998, **7:**1250-1254.

49. Hofmann K, Bucher P: **The PCI domain: a common theme in three multiprotein complexes.** *Trends Biochem Sci* 1998, **23:**204-205.

50. Kim T, Hofmann K, von Arnim AG, Chamovitz DA: **PCI complexes: pretty complex interactions in diverse signaling pathways.** *Trends Plant Sci* 2001, **6:**379-386.

51. **Impala (National Center for Biotechnology Information)** [http://ftp.ncbi.nih.gov/pub/impala/]

52. **CD-search** [http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi]

53. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302:**205-217.