

*Sequence analysis***Meisetz and the birth of the KRAB motif**

Zoë Birtle and Chris P. Ponting\*

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

Received on August 8, 2006; revised on September 26, 2006; accepted on September 27, 2006

Advance Access publication October 10, 2006

Associate Editor: Alex Bateman

**ABSTRACT**

The largest family of transcription factors in mammals is of Cys<sub>2</sub>His<sub>2</sub> zinc finger-proteins, each with an NH<sub>2</sub>-terminal KRAB motif. Extensive expansions of this family have occurred in separate mammalian lineages, with ~400 such genes known in the human genome. Despite their widespread occurrence, the evolutionary provenance of the KRAB motif is unclear since previously it has not been found outside of the tetrapod vertebrates. Here, we show that homologues of the histone methyltransferase Meisetz are present within the sea urchin (*Strongylocentrotus purpuratus*) genome. Sea urchin and mammalian Meisetz sequences each contain an N-terminal KRAB motif, which thereby establishes an early origin of the KRAB motif prior to the divergence of echinoderm and chordate lineages. Finally, we present evidence that KRAB motifs derive from a novel family of KRI (KRAB Interior) motifs that were present in the last common ancestor of animals, plants and fungi.

**Contact:** chris.ponting@anat.ox.ac.uk**Supplementary information:** Supplementary data for this article are available at *Bioinformatics* online.**1 INTRODUCTION**

Krüppel-associated box (KRAB) motif—and krüppel Cys<sub>2</sub>His<sub>2</sub> zinc finger-encoding transcription factors present three enduring mysteries. The first of these is the molecular basis to their unprecedented family expansions in mammalian genomes. KRAB motifs are encoded in ~either 300–400 genes in the human or mouse genomes (Huntley *et al.*, 2006). Many of these arose either in the human or in the mouse lineage by gene duplication in the last ~80 million years, with dispersal to many distinct chromosomes [(Faisst and Meyer, 1992; Rousseau-Merck *et al.*, 2002; Urrutia, 2003; Huntley *et al.*, 2006) and references therein]. As the vast majority of these genes are of unknown function, it is unclear what selective processes governed these gene family expansions.

The second mystery is the rapid evolution of mammalian KRAB motifs. Among all known families of domains or motifs, KRAB motifs appear, on average, to have evolved the fastest (Waterston *et al.*, 2002). Currently, it is unknown whether such divergent evolution results from reductions in selective constraints or from adaptive evolution.

The final mystery is the evolutionary provenance of the KRAB motif itself. Despite the large numbers of this motif encoded in

mammalian genomes (Bellefroid *et al.*, 1991; Huntley *et al.*, 2006), strangely it has only been found in proteins from tetrapod vertebrates (Looman *et al.*, 2002; Urrutia, 2003). KRAB motifs appear to be absent from more divergent vertebrates, such as fish, from non-vertebrate eukaryotes and from prokaryotes (Lander *et al.*, 2001). Extant domain families, such as that of KRAB, are not thought to have emerged recently from non-coding DNA (Ponting *et al.*, 2002). Rather, it is predicted that essentially all protein families arose from duplication of pre-existing coding sequence, which often diverges sufficiently rapidly to render imperceptible residual similarities in sequence (Copley *et al.*, 2003).

Within KRAB-Cys<sub>2</sub>His<sub>2</sub> transcription factors, KRAB motifs are thought to mediate protein–protein interactions (Bellefroid *et al.*, 1991; Urrutia, 2003) whereas the Cys<sub>2</sub>His<sub>2</sub> zinc fingers bind the major groove of B-DNA by packing their single  $\alpha$ -helices against 3 bp subsites (Pavletich and Pabo, 1991). Typically, these factors' KRAB motifs are divisible into two. All KRAB-containing genes contain the KRAB A motif, which appears often to mediate a transcription repression function (Margolin *et al.*, 1994). KRAB A may occur alone, but commonly it is accompanied by an additional KRAB B modulatory motif (Huntley *et al.*, 2006). KRAB A and B motifs are each encoded by separate exons and these are invariably positioned N-terminal to tandem arrays of Cys<sub>2</sub>His<sub>2</sub> zinc fingers, often all encoded within a single exon (Miller *et al.*, 1985; Shannon *et al.*, 1998). The exception to this KRAB-Cys<sub>2</sub>His<sub>2</sub> zinc finger domain architecture is the family of SSX (Synovial sarcoma, X breakpoint) proteins, which possess an N-terminal and sequence-divergent KRAB motif unaccompanied by zinc fingers (Lim *et al.*, 1998).

Recently, a putative transcription factor gene named *Meisetz* was characterized, which appears to possess a role in the progression of early meiotic prophase. *Meisetz*<sup>−/−</sup> mice are sterile and exhibit deficits in meiosis (Hayashi *et al.*, 2005). The Meisetz protein contains both a central SET (Suvar3–9, Enhancer-of-zeste, Trithorax) domain, which catalyses the trimethylation of lysine 4 in histone H3, and a C-terminal array of tandem Cys<sub>2</sub>His<sub>2</sub> zinc fingers all encoded in a single exon.

Our interest in Meisetz arose initially from an observation that it contains a hitherto unrecognized N-terminal KRAB motif, similar to those of SSX proteins. Subsequently, we realized that Meisetz is not specific to mammals since the purple sea urchin (*Strongylocentrotus purpuratus*) possesses Meisetz orthologues. Furthermore, single N-terminal KRAB domains are apparent within these sequences, and similar homologues from amphioxus

\*To whom correspondence should be addressed.

(*Branchiostoma floridae*), and from sea squirts (*Ciona intestinalis* and *Ciona savignyi*). This pushes back the origins of this enigmatic domain to at least the common deuterostome ancestor of vertebrates and echinoderms. Eventually, by exploiting the divergence of these early-branching KRAB motifs, we obtained evidence for the common ancestry of the KRAB A motif with a novel KRI (KRAB Interior) motif present in molecules from among the crown group of eukaryotes, the fungi, plants and animals. KRI occurs within proteins possessing chromatin-related functions, and also alongside other domains predicted to be involved in transcriptional regulation. The cellular function of KRAB may thus have been inherited from its pre-existing KRI ancestor.

## 2 RESULTS AND DISCUSSION

### 2.1 Meisetz contains an N-terminal KRAB motif

Mouse (*Mus musculus*) Meisetz, and its mammalian orthologues, were each found to possess an N-terminal KRAB motif. Searches of the SMART database (Ponting *et al.*, 1999) and the conserved domain database (Marchler-Bauer and Bryant, 2004) revealed significant similarities ( $E = 5.5 \times 10^{-8}$  and  $1 \times 10^{-9}$ , respectively) to known KRAB motifs. In addition, using a hidden Markov model of the putative KRAB motif from seven vertebrate and non-vertebrate chordates (see below), Meisetz KRAB motifs were found to be substantially more similar to those in mammalian SSX proteins ( $7 \times 10^{-17} < E < 4.5 \times 10^{-3}$ ) than those in KRAB-Cys<sub>2</sub>His<sub>2</sub> zinc finger proteins ( $E > 0.14$ ), in a search of all known protein sequences (Fig. 1).

### 2.2 MEISETZ homologues outside of mammals

Unexpectedly, we found evidence for Meisetz homologues in three diverse non-vertebrate chordates (Delsuc *et al.*, 2006), an echinoderm, a cephalochordate and a tunicate. Each of these homologues contains a single N-terminal KRAB motif, which represent the first recorded members of this family from outside of tetrapod vertebrates.

First, by comparing mouse Meisetz with current protein sequences using BLASTP (Altschul *et al.*, 1997), we identified a Meisetz homologue (XP\_799022.1;  $E = 7 \times 10^{-47}$ ) from an echinoderm, the California purple sea urchin (*S.purpuratus*). Indeed, further investigation of the draft sea urchin genome revealed a total of three complete, and also three incomplete, sequence-similar Meisetz homologues. Given the increased levels of heterozygosity of the sea urchin genome sequenced, some of the multiple Meisetz homologues in the assembly may instead represent allelic variants. A Meisetz-like KRAB motif was then found in expressed sequence tags (ESTs) from amphioxus (*B.floridae*, GenBank accession nos BW709598 and BW709302). Finally, a divergent Meisetz homologue was also apparent in the genome assembly of a tunicate, *C.intestinalis*, whose predicted amino acid sequence is significantly similar to a human Meisetz orthologue, PRDM9 ( $E = 7 \times 10^{-8}$ ), including within its KRAB motif.

Next, using BLASTP (Altschul *et al.*, 1997), we detected a Meisetz-like gene encoding SET and 14 C-terminal Cys<sub>2</sub>His<sub>2</sub> zinc fingers in zebrafish (*Danio rerio*; RefSeq NP\_957196.1). This gene's orthologues in two pufferfish (*Tetraodon nigroviridis* and *Takifugu rubripes*) genomes, and in ESTs from the minnow (*Pimephales promela*), were then also identified using BLAT (Kent 2002) and TBLASTn, respectively. A KRAB motif was apparent

neither from the zebrafish cDNA, nor from three minnow ESTs in their short regions up to the 5'-most in-frame stop codon, nor from orthologous pufferfish gene predictions, and nor from upstream genomic sequence.

Phylogenetic analysis of these sequences' SET domains (Supplementary Fig. 1) indicates that the mammalian, fish and sea urchin Meisetz genes are orthologous, whereas the sea squirt homologues may have acquired the KRAB-SET-Cys<sub>2</sub>His<sub>2</sub> zinc fingers' domain architecture of Meisetz independently. We were unable to identify Meisetz homologues in the chicken (*Gallus gallus*) or frog (*Xenopus tropicalis*) genomes, which are the only genomes from outside the mammals previously known to encode KRAB motifs.

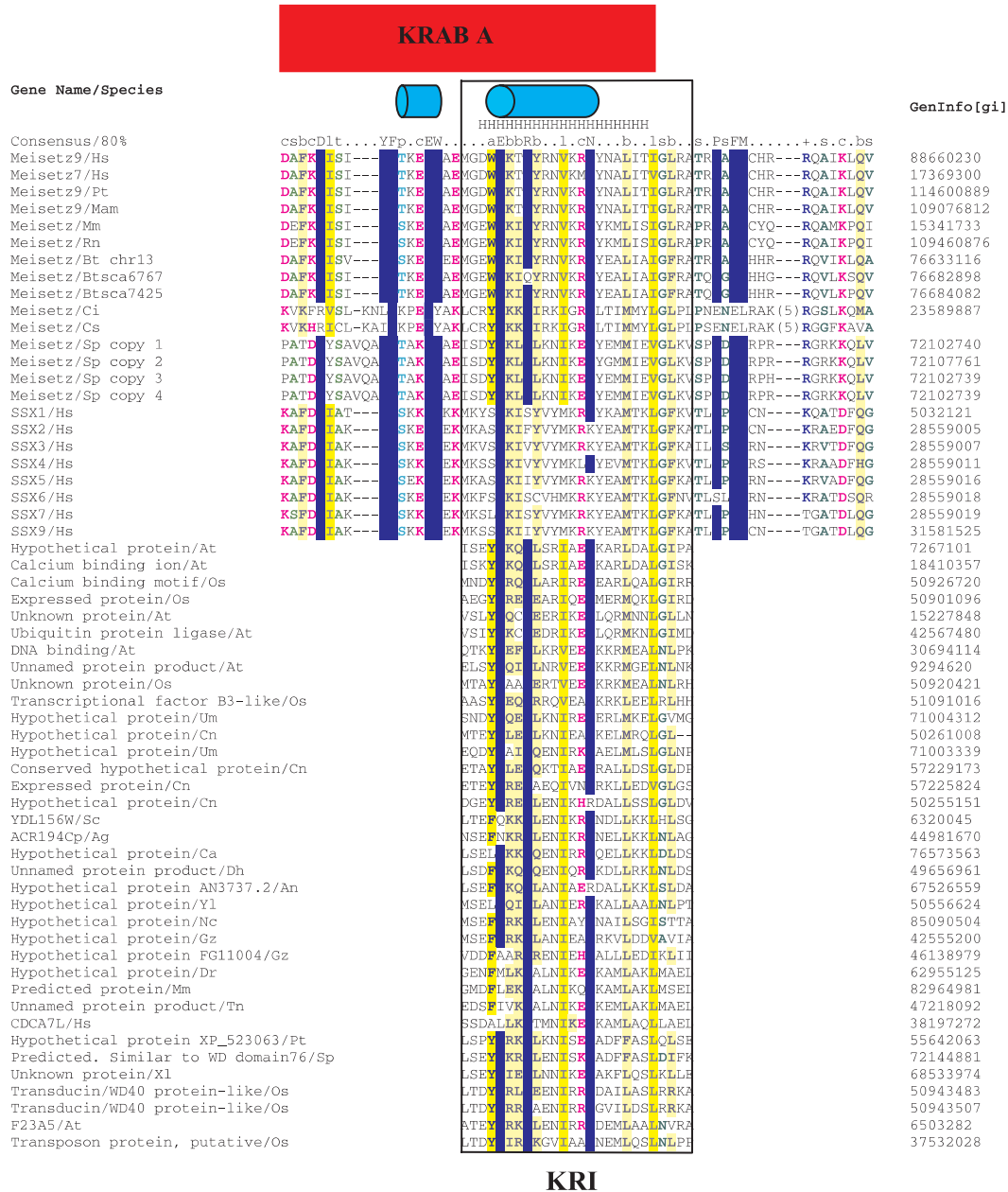
The *S.purpuratus* Meisetz homologues possess domain architectures equivalent to those of mammals and thus they each possess a KRAB motif. For example, using HMMER (Eddy 1998) we find that a sea urchin Meisetz sequence (XP\_799022.1) is significantly similar ( $E = 8 \times 10^{-3}$ ) to an alignment of known KRAB motif sequences (Marchler-Bauer and Bryant, 2004). *S.purpuratus*, *C.intestinalis* and mammalian Meisetz genes were also found to share equivalent exon boundaries about their KRAB motifs.

We conclude that Meisetz originated prior to the last common deuterostome ancestor of chordates and echinoderms at least 520 million years ago (MYA) (Blair and Hedges, 2005). Moreover, as sea urchin Meisetz genes encode KRAB motifs we have pushed back the origins of KRAB motifs from 370 MYA (the last common ancestor of frogs and mammals (Blair and Hedges, 2005)) by >140 MY. As no other, more distantly-related, KRAB motifs are recognizable, using exhaustive BLAST (Altschul *et al.*, 1997) and HMMER (Eddy, 1998) searches, within current invertebrate (including sea urchin) protein sequences, the Meisetz KRAB motif is thus a good candidate for being the progenitor motif from which all other KRAB family members were derived.

The antiquity of the KRAB motif is consistent with the phyletic distribution of KAP1 (also called Krip-1 or TIF1 $\beta$ ) orthologues (Friedman *et al.*, 1996), to which, in mammalian proteins, it binds [with the notable exception of these motifs in SSX proteins (Lim *et al.*, 1998)]. However, despite the apparent absence of KRAB motifs in arthropods, a single orthologue of the three TIF1 $\alpha$ /TIF1 $\beta$ /TIF1 $\gamma$  molecules is found in the fruitfly, where, as in mammals, it possesses a role in chromatin-mediated transcriptional repression (Beckstead *et al.*, 2001).

### 2.3 KRI—a KRAB progenitor

Our unexpected observation of KRAB motifs in invertebrate chordate proteins prompted an investigation seeking to identify KRAB motif homologues from earlier-branching, non-metazoan eukaryotes, such as plants and fungi. Indeed, significant sequence similarity ( $E = 2 \times 10^{-4}$ , fourth iteration) was detected between the N-terminal region of *Saccharomyces cerevisiae* YDL156W, a nuclear protein (Huh *et al.*, 2003), and the KRAB motif region of sea urchin Meisetz. This search used PSI-BLAST (Altschul *et al.*, 1997) and an *E*-value inclusion threshold of  $E < 2 \times 10^{-3}$  in a comparison of the first 180 amino acids of the *S.cerevisiae* protein with the NCBI's non-redundant protein sequence database (nr). This search was suggested because of an initial, albeit non-significant, similarity found between the *Saccharomyces kluyveri* YDL156W gene (contig AACE02000536.1; 725–973 nt) and the



**Fig. 1.** A multiple sequence alignment of identified KRI motifs together with KRAB motifs encoded in *Meisetz* and *SSX* genes. The span of the KRAB A exon is indicated in red whereas the extent of the KRI motif is indicated by a box. The secondary structure of the KRAB motif consists of two  $\alpha$ -helices, which are illustrated by cylinders [see Protein Data Bank file PDB:1V65 (Berman *et al.*, 2000)]. Below this is the secondary structure of KRI motifs (H =  $\alpha$ -helix) predicted by Jpred (Cuff *et al.*, 1998; Cuff *et al.*, 2000). The alignment was presented using CHROMA (Goodstadt and Ponting, 2001) and an 80% consensus. Protein names and species (using two letter abbreviations, see below) are indicated to the left of the alignment, whereas GenInfo Identifier (gi) numbers, when available, are on the right of the alignment. For the three cattle *Meisetz* gene predictions, chromosome and scaffold numbers are provided. Six *Meisetz* sequences are apparent from within the sea urchin (*S.purpuratus*) genome [obtained from three full length sequences and one partial sequence (labelled Meisetz/Sp copy 4), which contains complete KRAB and SET motifs but no Cys<sub>2</sub>His<sub>2</sub> zinc fingers]. The *C.savignyi* Meisetz homologue was predicted from two EST sequences, BW589804 and BW534495, which were corrected by comparisons with the current genome assembly. We provide predicted gene sequences in Supplementary material. Consensus abbreviations (amino acids): a, aromatic (FHWHY, blue lettering on a dark yellow background); b, big (EFHIKLMQRWY, blue on light yellow); h, hydrophobic (ACFGHILMTVWY, black on dark yellow); l, aliphatic (ILV, grey on dark yellow); p, polar (CDEHKNQRST, blue on white); s, small (ACDGNPSTV, dark green on white); plus sign, positively charged (HKR, dark blue on white); c, charged (DHKER, pink on white). Species abbreviations: Ag, *Ashbya gossypii*; An, *Aspergillus nidulans*; At, *Arabidopsis thaliana*; Bt, *Bos taurus* (cattle); Ca, *Candida albicans*; Ci, *Ciona intestinalis*; Cn, *Cryptococcus neoformans*; Cs, *Ciona savignyi*; Dh, *Debaryomyces hansenii*; Dr, *Danio rerio* (zebra fish); Gz, *Gibberella zeae*; Hs, *Homo sapiens*; Mac, *Macaca mulatta* (macaque); Mm, *Mus musculus*; Nc, *Neurospora crassa*; Os, *Oryza sativa* (rice); Pt, *Pan troglodytes*; Rn, *Rattus norvegicus*; Sc, *Saccharomyces cerevisiae*; Sp, *Strongylocentrotus purpuratus*; Tn, *Tetraodon nigroviridis* (puffer fish); Um, *Ustilago maydis*; Xl, *Xenopus laevis* (frog); and, Yl, *Yarrowia lipolytica*.



N-terminal 1–183 amino acids of sea urchin Meisetz (TblastN search: 38.5 bits;  $E = 0.47$ ).

Furthermore, 25 additional putative homologues ( $E < 0.1$ ) of this motif were collated in three iterations of database searches using a hidden Markov model constructed (Eddy, 1998) from a multiple alignment of 13 fungal YDL156W orthologues. A further 10 putative homologues were then identified by searching nr with *S.cerevisiae* YDL156W (1–90 amino acids) using PSI-BLAST in ungapped mode (options -G 32767 -E 32767 -h 0.005) in 13 iterations. The extent of the motif was defined on the basis that it encompassed the most N- and C-terminal amino acids of *Yarrowia lipolytica* YDL156W and *Cryptococcus neoformans* CNBA3050 (GenInfo identifiers [gi] 50556624 and 50261008), respectively.

These sequences were then found to be significantly similar to KRAB motifs using two methods that compare amino acid alignments. When an alignment of these 35 YDL156W-like motifs was compared with an alignment of 8 KRAB motifs from Meisetz- and SSX-like proteins using COMPASS (Sadreyev and Grishin, 2003), a highly significant  $E$ -value of  $2.9 \times 10^{-10}$  was obtained. Similarly, comparison of YDL156W-like and Meisetz-like KRAB motif alignments, using LAMA (Petrokovski, 1996) generated a highly significant  $Z$ -score of 10.1. In a random simulation involving 7 million block pair comparisons no alignments yielded  $Z$ -scores  $> 8.3$  (Petrokovski, 1996); hence, the expected number of random block alignments with a  $Z$ -score by chance equal to, or greater than, 10.1 is essentially zero.

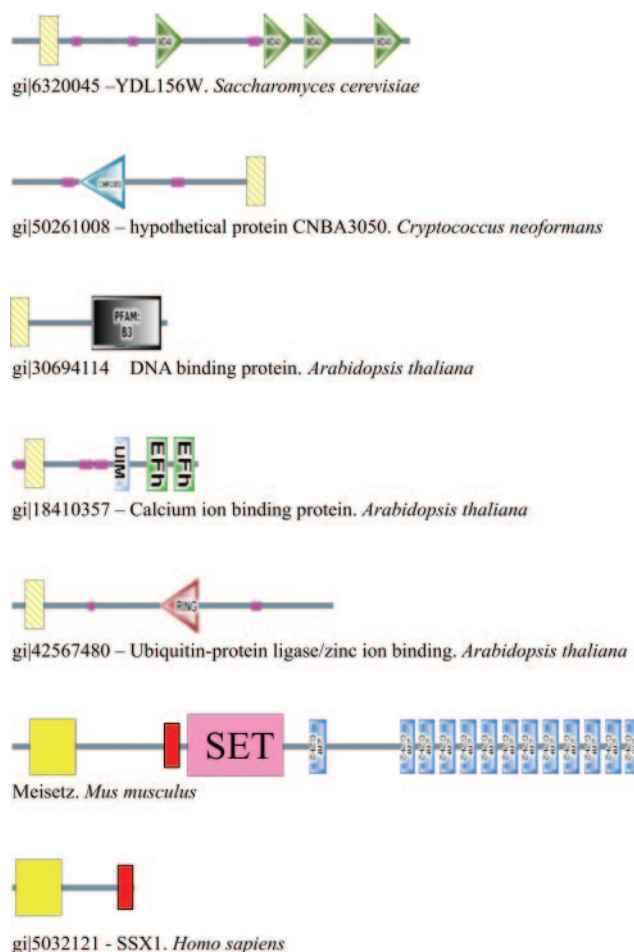
These significant sequence similarities between short regions of fungal YDL156W and animal Meisetz proteins (Fig. 1) indicate that they evolved convergently, or they are homologous, they share a common ancestor. Convergence appears less likely given the strong sequence similarities between these two motifs, their common phylogenetic distributions and the functional similarities between their co-occurring domains (see below). Assuming homology, we term the YDL156W, and similar, sequences KRI or KRAB Interior motifs since they align against the interior of Meisetz KRAB motifs. KRI motifs match the known  $\alpha$ -helical regions of KRAB motifs (Fig. 1).

KRI motifs are present in proteins from all major eukaryote lineages, including plants *Arabidopsis thaliana* and *Oryza sativa* (rice), and ascomycete fungi *Neurospora crassa* and *Aspergillus nidulans* (in addition to the *Saccharomyces* species described above). Despite Meisetz KRAB motifs being apparently absent in fish and frogs, we find KRI homologues in fish (e.g. *T.nigroviridis* sequence gi:472218092) and frog species (e.g. *Xenopus laevis* sequence gi:68533974) and otherwise throughout the vertebrates.

The wider phyletic distribution of KRI, compared to KRAB, motifs suggests that the 26 amino acid KRI sequence may be the progenitor of KRAB. In this model, accretion of additional sequence to KRI formed the KRAB A sequence in an early deuterostome Meisetz protein. This KRAB A motif, we predict, represents the founder member of all extant mammalian KRAB A motifs.

## 2.4 KRI motif function

In the main, the functions of KRI motifs remain obscure. However, in the Myc-interacting protein JPO2 (also known as R1, RAM2 or CDCA7L) the KRI motif is known, from experiments involving short 6 amino acid deletions within the motif, to be important for binding to the DNA transcription factor, c-Myc (Huang *et al.*, 2005). Other studies suggest that JPO2 localizes to chromatin (Maertens



**Fig. 2.** The five most common domain architectures found in KRI motif-containing proteins (not to scale) drawn using SMART (Schultz *et al.*, 2000). The position of the KRI motif is illustrated by a hatched yellow box. Domain architectures of mouse Meisetz and human SSX1 are also shown in comparison (KRAB motifs indicated by solid yellow boxes). The red, blue and pale pink boxes indicate SSXRD motif (Lim *et al.*, 1998), Cys<sub>2</sub>His<sub>2</sub> zinc fingers and SET domains, respectively. Bright pink horizontal lines represent regions of low compositional complexity as predicted using SEG (Wootton and Federhen, 1993).

*et al.*, 2006), and possesses transcriptional repressor activity (Chen *et al.*, 2005). A chromatin-related function for KRI motifs also suggested from consideration of the architectures of the proteins in which they occur (Fig. 2). KRI motif proteins contain, in addition, at least six other domain types. Of these, members of the CHROMO and WD40 domain families are known to bind lysine-methylated histones (Bannister *et al.*, 2001; Wysocka *et al.*, 2005); UIM domains bind ubiquitin, and methylation of lysine 4 histone H3 requires ubiquitylation of histone H2B (Sun and Allis, 2002) and RING finger proteins are E3 ubiquitin ligases, some of which promote ubiquitylation of histones H2A and H2B *in vitro* (Minsky and Oren, 2004). Just as a KRAB motif typically co-opts other proteins into DNA-associated protein complexes, so might a KRI motif, perhaps via its single predicted  $\alpha$ -helix that it shares in common with KRAB motifs.

### 3 CONCLUSIONS

Most KRAB domain proteins co-localize with the lysine 9 histone H3 methyltransferase SETDB1 via KAP1 and thus assist in transcriptional repression (Schultz *et al.*, 2002). By way of contrast, Meisetz is a lysine 4 histone H3 methyltransferase that appears to be a transcriptional activator of genes that are critical for the progression of meiotic prophase (Hayashi *et al.*, 2005). It is thus plausible that the progenitor Meisetz KRAB was involved in transcriptional activation but that this function was subverted by later evolving proteins which exploited the motif for transcription repression purposes.

In summary, we have identified a new family of KRI motifs from which, prior to the split of echinoderms and chordates, the Meisetz KRAB motif arose. This KRAB motif may represent the founder member for the hundreds of KRAB motifs encoded in multiple mammalian genomes. The functions of KRI and Meisetz KRAB motifs will need to be experimentally determined if we are to better appreciate the function and evolution of these important transcriptional regulators.

### ACKNOWLEDGEMENTS

The authors thank the Ponting group, in particular Leo Goodstadt, for helpful discussions and advice. This work was funded by the UK Medical Research Council (MRC). Funding to pay the Open Access publication charges for this article was provided by the MRC.

*Conflict of Interest:* none declared.

### REFERENCES

- Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Bannister, A.J. *et al.* (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, **410**, 120–124.
- Beckstead, R. *et al.* (2001) Bonus, a *Drosophila* homolog of TIF1 proteins, interacts with nuclear receptors and can inhibit betaFTZ-F1-dependent transcription. *Mol. Cell*, **7**, 753–765.
- Bellefroid, E.J. *et al.* (1991) The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Natl Acad. Sci. USA*, **88**, 3608–3612.
- Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Blair, J.E. and Hedges, S.B. (2005) Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.*, **22**, 2275–2284.
- Chen, K. *et al.* (2005) R1, a novel repressor of the human monoamine oxidase A. *J. Biol. Chem.*, **280**, 11552–11559.
- Copley, R.R. *et al.* (2003) Eukaryotic domain evolution inferred from genome comparisons. *Curr. Opin. Genet. Dev.*, **13**, 623–628.
- Cuff, J.A. and Barton, G.J. (2000) Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*, **40**, 502–511.
- Cuff, J.A. *et al.* (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics*, **14**, 892–893.
- Delsuc, F. *et al.* (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*, **439**, 965–968.
- Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
- Faisst, S. and Meyer, S. (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.*, **20**, 3–26.
- Friedman, J.R. *et al.* (1996) KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev*, **10**, 2067–2078.
- Goodstadt, L. and Ponting, C.P. (2001) CHROMA: consensus-based colouring of multiple alignments for publication. *Bioinformatics*, **17**, 845–846.
- Hayashi, K. *et al.* (2005) A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*, **438**, 374–378.
- Huang, A. *et al.* (2005) Identification of a novel c-Myc protein interactor, JPO2, with transforming activity in medulloblastoma cells. *Cancer Res*, **65**, 5607–5619.
- Huh, W.K. *et al.* (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.
- Huntley, S. *et al.* (2006) A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res*, **16**, 669–677.
- Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res*, **12**, 656–664.
- Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Lim, F.L. *et al.* (1998) A KRAB-related domain and a novel transcription repression domain in proteins encoded by SSX genes that are disrupted in human sarcomas. *Oncogene*, **17**, 2013–2018.
- Looman, C. *et al.* (2002) KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Mol. Biol. Evol.*, **19**, 2118–2130.
- Maertens, G.N. *et al.* (2006) Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. *J. Cell. Sci.*, **119**, 2563–2571.
- Marchler-Bauer, A. and Bryant, S.H. (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.*, **32**, W327–W331.
- Margolin, J.F. *et al.* (1994) Kruppel-associated boxes are potent transcriptional repression domains. *Proc. Natl Acad. Sci. USA*, **91**, 4509–4513.
- Miller, J. *et al.* (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.*, **4**, 1609–1614.
- Minsky, N. and Oren, M. (2004) The RING domain of Mdm2 mediates histone ubiquitylation and transcriptional repression. *Mol. Cell*, **16**, 631–639.
- Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809–817.
- Petrokovski, S. (1996) Searching databases of conserved sequence regions by aligning protein multiple-alignments. *Nucleic Acids Res.*, **24**, 3836–3845.
- Ponting, C.P. and Russell, R.R. (2002) The natural history of protein domains. *Annu. Rev. Biophys. Biomol. Struct.*, **31**, 45–71.
- Ponting, C.P. *et al.* (1999) SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res.*, **27**, 229–232.
- Rousseau-Merck, M.F. *et al.* (2002) The KOX zinc finger genes: genome wide mapping of 368 ZNF PAC clones with zinc finger gene clusters predominantly in 23 chromosomal loci are confirmed by human sequences annotated in Ensembl. *Cytogenet. Genome Res.*, **98**, 147–153.
- Sadreyev, R. and Grishin, N. (2003) COMPASS: a tool for comparison of multiple protein alignments with assessment of statistical significance. *J. Mol. Biol.*, **326**, 317–336.
- Schultz, D.C. *et al.* (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev*, **16**, 919–932.
- Schultz, J. *et al.* (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.*, **28**, 231–234.
- Shannon, M. *et al.* (1998) Tandem zinc-finger gene families in mammals: insights and unanswered questions. *DNA Seq.*, **8**, 303–315.
- Sun, Z.W. and Allis, C.D. (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature*, **418**, 104–108.
- Urrutia, R. (2003) KRAB-containing zinc-finger repressor proteins. *Genome Biol.*, **4**, 231.
- Waterston, R.H. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
- Wootton, J.C. and Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem.*, **17**, 149–163.
- Wysocka, J. *et al.* (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, **121**, 859–872.