

Published in final edited form as:

Cell. 2009 June 26; 137(7): 1173–1174. doi:10.1016/j.cell.2009.06.010.

Common Ancestry of the CENP-A Chaperones Scm3 and HJURP

Luis Sanchez-Pulido¹, Alison L. Pidoux², Chris P. Ponting^{1,*}, and Robin C. Allshire^{2,*}

¹MRC Functional Genomics Unit, University of Oxford, Oxford OX1 3QX, UK ²Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, Scotland

The centromere is a unique chromosomal locus that ensures accurate segregation of chromosomes during cell division. The centromere supports assembly of a multiprotein complex called the kinetochore, which attaches to spindle microtubules. The kinetochore has specialized nucleosomes in which histone H3 is replaced by the centromere-specific H3 variant CENP-A/cenH3 (reviewed in Allshire and Karpen, 2008). Two recent papers in *Cell* (Dunleavy et al., 2009; Foltz et al., 2009) have identified a new protein partner for soluble human CENP-A called HJURP/hFLEG/FAKTS that promotes the incorporation of CENP-A at centromeres.

Comparing the mechanism of deposition of CENP-A at centromeres across different organisms is difficult due to species-specific differences in the cell-cycle timing of CENP-A incorporation. Moreover, across eukaryotes, apparently diverse proteins associate with CENP-A and mediate its assembly into nucleosomes. There is, however, a lack of evidence for common ancestry (i.e., homology) among these proteins. In the budding yeast *Saccharomyces cerevisiae* and fission yeast *Schizosaccharomyces pombe*, the centromere-associated Scm3 protein binds to CENP-A and is required for incorporation of CENP-A into centromeric chromatin (Camahort et al., 2007; Mizuguchi et al., 2007; Stoler et al., 2007; Pidoux et al., 2009; Williams et al., 2009).

Similarities in the behavior and roles of Scm3 and HJURP suggest that they occupy the same functional niche (Dunleavy et al., 2009). Might these two functionally analogous proteins also be homologs whose shared common ancestry has been difficult to discern owing to substantial sequence divergence? Their non-overlapping phylogenetic ranges suggest that this is a distinct possibility.

To address this question we sought to determine the evolutionary provenances of fungal Scm3 and of mammalian HJURP. In an initial database search (Supplemental Experimental Procedures available online), the first Scm3 homolog to be detected in organisms other than fungi was found in a marine choanoflagellate *Monosiga brevicollis*. Tantalizingly, this search also showed marginal (but nonsignificant) sequence similarities between the Scm3

*Correspondence: chris.ponting@dpag.ox.ac.uk (C.P.P.) robin.allshire@ed.ac.uk (R.C.A.).

Supplemental Data include Supplemental Experimental Procedures, one figure, and Supplemental References and can also be found at [http://www.cell.com/supplemental/S0092-8674\(09\)00708-9](http://www.cell.com/supplemental/S0092-8674(09)00708-9).

family and bovine HJURP. Conversely, BLASTp and tBLASTn searches identified homologs of human HJURP in frogs and birds ($E < 2 \times 10^{-3}$), the first homologs reported in organisms other than mammals.

A sensitive profile sequence search identified additional, more divergent Scm3-like sequences beyond fungi and *Monosiga*. This established marginal similarity between fungal Scm3 and bovine HJURP with an E value of 0.7 (Figure S1) and implies that less than one sequence is expected to be found in this database search with an equivalent or better alignment score simply by chance. This finding, together with the known functional similarities between Scm3 and HJURP proteins, would be consistent with their common ancestry.

Finally, we provide confirmation of this prediction by comparing the profile of the fungal Scm3 protein alignment with that for the metazoan HJURP alignment, and vice versa, using HHpred (Soding et al., 2005). In each comparison, sequence similarity between these two families was statistically highly significant ($E < 10^{-5}$) (Figure S1). This level of significance implies that these proteins are homologous members of a wider Scm3/HJURP protein family.

The inclusion within sequence alignments of both families of more divergent, yet homologous, sequences from, for example, the frog *Xenopus tropicalis* and the choanoflagellate *Monosiga brevicollis*, appears to have allowed more sensitive sequence searches. This explains why this remote homologous relationship had hitherto escaped detection (Aravind et al., 2007). Our searches failed to identify members of this family in the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*. Although more rapid evolution occurs in these lineages, it cannot be discounted that these organisms possess alternative, functionally analogous molecules for loading CENP-A at centromeres. Besides Scm3 or HJURP orthologs, no additional homologs were discernible for any species.

Our analyses were unable to confirm the prediction by others of tryptophan-aspartic acid (WD40) repeats in HJURP, similar to those found in chromatin assembly factors (Foltz et al., 2009). Indeed, the first such proposed repeat encompasses the Scm3-HJURP homologous domain, which instead of containing β sheets as in WD40 repeats is predicted to contain an α helix. Moreover, the proposed conserved tryptophan residue is substituted in frog HJURP (Figure S1). We provide statistical support for a mixed $\alpha + \beta$ domain, instead of a WD40 β sheet repeat, in the N-terminal region of Scm3 and HJURP proteins. The only domain that all Scm3 and HJURP proteins have in common is the “Scm3 domain” (Figure S1), which may harbor a CENP-A-binding site. Indeed, the region in Scm3^{Sc} that binds to CENP-A^{Cse4} (residues 90–193) encompasses the Scm3 domain (residues 90–142) (Mizuguchi et al., 2007; Aravind et al., 2007). Moreover, two Scm3^{Sp} amino acid substitutions within the Scm3 domain that disrupt CENP-A^{Cnp1} localization in *S. pombe* are of leucine residues (Pidoux et al., 2009), which are conserved in other species as small hydrophobic side chains (leucine, isoleucine, or methionine; Figure S1). Other mutations within the Scm3 domain of Scm3^{Sp} (N50S) or just C-terminal to it (N100S) also disrupt the localization of CENP-A^{Cnp1} (Williams et al., 2009).

Our finding that the Scm3 domain is shared between, and is unique to, human HJURP and yeast Scm3 unites previously disparate lines of research. Descriptions of functionally similar, yet seemingly distinct, proteins from *S. cerevisiae*, *S. pombe*, and human cells might imply that centromere-specific histone incorporation differs greatly between fungi and animals. Instead, control of CENP-A incorporation at centromeres via Scm3/HJURP appears to be common to these eukaryotes. There are, however, likely to be some derived, lineage-specific features of CENP-A incorporation because of the various domains added to and deleted from Scm3 in different fungi (Aravind et al., 2007) and in vertebrates (Figure S1).

Our analyses reconcile previous observations by demonstrating that fungal Scm3 proteins are indeed distant counterparts of human HJURP. Thus, investigation of Scm3 and associated proteins is likely to be directly relevant to understanding the mechanism of HJURP-mediated CENP-A chromatin assembly at human centromeres.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

L.S.-P. is supported by an EMBO Long Term Fellowship, and the C.P.P. and R.C.A. laboratories by the Medical Research Council and the Wellcome Trust, respectively. R.C.A. is a Wellcome Trust Principal Research Fellow.

References

- 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 Research*. 1997; 25:3389–3402. [PubMed: 9254694]
- Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research*. 2006; 34:D187–191. [PubMed: 16381842]
- Eddy SR. Hidden Markov models. *Curr. Opin. Struc. Biol*. 1996; 6:361–365. [PubMed: 8804822]
- Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*. 2005; 33:W244–248. [PubMed: 15980461]