

HPC2 and ubinuclein define a novel family of histone chaperones conserved throughout eukaryotes†

S. Balaji,* Lakshminarayan M. Iyer and L. Aravind*

Received 18th September 2008, Accepted 8th December 2008

First published as an Advance Article on the web 21st January 2009

DOI: 10.1039/b816424j

While histone chaperones have been intensely studied, the roles of components of the Hir–Asf1 histone chaperone complex such as Hir3p and Hpc2p are poorly understood. Using sensitive protein sequence profile analyses we investigated the evolution of these proteins and showed that Hir3p and Hpc2p have a much wider phyletic pattern than was previously known. We established the animal histone-deacetylase-complex-interacting proteins, CAIN/CABIN, to be orthologs of Hir3p. They contain a conserved core of around 30 TPR-like bi-helical repeats that are likely to form a super-helical scaffold. We identified a conserved domain, the HUN domain, in all Hpc2p homologs, including animal ubinuclein/yemanuclein and the recently discovered vertebrate cell-cycle regulator FLJ25778. The HUN domain has a characteristic pattern of conserved acidic residues based on which we predict that it is a previously unrecognized histone-tail-binding chaperone. By analyzing various high-throughput data sets, such as RNAi knock-downs, genetic and protein interaction maps and cell-cycle-specific gene expression data, we present evidence that Hpc2p homologs might be deployed in specific processes of chromatin dynamics relating to cell-cycle progression in vertebrates and schizogony in *Plasmodium*. Beyond the conserved HUN domain these proteins show extensive divergence patterns in different eukaryotic lineages. Hence, we propose that Hpc2p homologs are probably involved in recruitment of the ancient conserved histone-loading Hir–Asf1 complex to different lineage-specific chromatin reorganization processes.

Introduction

The unique presence of long positively charged tails in eukaryotic nucleosomal histones appears to have provided the selective pressure for evolution of a panoply of histone chaperones.¹ This diverse group of proteins escorts histones right from the point of their synthesis, through their entry into the nucleus, to their repeated deposition on DNA during chromatin assembly and remodeling.^{1–3} The functions performed by different histone chaperones include histone storage (nucleoplasmin), shuttling of histone to the nucleus (NAP-1), and deposition of nucleosomes in conjunction with specific chromatin remodeling complexes (various nuclear histone chaperones).^{1,2,4–15} This latter group includes complexes with specific temporal roles, such as the CAF-1 complex required for histone deposition during replication, the FACT complex required for histone disassembly and reassembly during transcription elongation and the Hir complex that appears to be active throughout the cell cycle.^{1,2,10,16–21} Also included in this group are complexes specific to certain histone variants such as Swr1–Chz1 system involved in the incorporation of H2AZ and Scm3 required for cenH3 assembly.^{1,3,22–26} Of the characterized histone chaperones, Asf1 is required for both replication-dependent and independent nucleosome assembly *via* direct

binding of the H3–H4 dimer.^{2,19,20,27} In these two distinct processes it, respectively, associates with either of two paralogous WD40 repeat proteins, Cac2p and HirA (Hir1p and Hir2p in *Saccharomyces cerevisiae*) to form two distinct chaperone complexes namely CAF-1 and Hir.^{20,28–31} Biochemical studies in *S. cerevisiae* have shown that the complete Hir complex comprises 4 proteins (Hir1p, Hir2p, Hir3p and Hpc2p), which also show functional interactions with a wide range of chromatin associated complexes, such as the FACT, CAF-1, PAF1 and Swi2–Snf2 complexes.^{19,21,32–36} Additionally, genetic studies also imply that the Hir complex might function independently of Asf1 in maintenance of kinetochore chromatin structure.^{32,33,36,37} Taken together the evidence suggests that the Hir complex and Asf1p are important generalized histone chaperones providing functional backup for a number of functionally specialized complexes involved in nucleosome deposition and remodeling.

Evolutionary comparisons have shown that Asf1p is highly conserved throughout eukaryotes—it is present even in organisms with highly reduced functional systems, such as microsporidians, the nucleomorph of *Guillardia theta* and *Giardia*.^{27,38} This suggests that Asf1 is indeed the key generalized histone chaperone that was already present in the last common ancestor of extant eukaryotes. The evolutionary history of its functional partners in generalized histone deposition, *i.e.* the components of the Hir complex, and the differentiation of the replication-linked Caf-1 complex are less understood. The core WD40 subunits of these complexes HirA and Cac2 are known to be widely conserved in various lineages

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. E-mail: aravind@mail.nih.gov. E-mail: sbalaji@mail.nih.gov

† Electronic supplementary information (ESI) available: Text alignments and supporting information. See DOI: 10.1039/b816424j

of the eukaryotic crown group (fungi, animals and plants), with comparable functions and interactions.¹ But the exact role and evolutionary pattern of the other subunits, Hir3p and Hpc2p, are unclear. We were specifically interested in understanding the origins of these less-studied Hir complex components and potentially uncovering previously uncharacterized features of their role in nucleosome assembly and deposition. Utilizing several distinct computational approaches we show that Hpc2p contains an ancient conserved domain that is likely to have a key pan-eukaryotic role in the nucleosome assembly mediated by the Hir–Asf1 complex. We also show that Hir3p is a rapidly diverging member of the complex that has possibly undergone extensive functional differentiation in the eukaryotic crown group.

Results and discussion

Phyletic and divergence patterns of components of the Hir complex suggest possible functional shifts in eukaryotic evolution

Phyletic patterns and evolutionary divergence of proteins often provide a good estimate of their functional relevance and changes in their interacting partners. They might also throw new light on their biochemical roles. To determine if components of the Hir complex showed a conservation pattern similar to Asf1p we systematically identified their orthologs across a panel of eukaryotes that included the crown group (animals, fungi, amoebozoans and plants), the *Guillardia* nucleomorph, chromalveolates (apicomplexans, ciliates, oomycetes and diatoms), heteroloboseans, kinetoplastids, parabasalids and diplomonads (Table S1, ESI†). The physically interacting partner of Asf1p, HirA is characterized by N- and C-terminal WD40 repeat β -propellers with an intervening linker containing a motif implicated in interactions with Asf1.¹ While this configuration is conserved throughout the crown group, in saccharomycete yeasts the HirA ortholog has undergone a lineage-specific duplication spawning two genes Hir1 and Hir2, which have, respectively, diverged rapidly in their C- and N-terminal β -propellers. The primary conserved subunit of the Caf-1 complex, Cac2p is a paralog of HirA that appears to contain only an N-terminal β -propeller followed by a C-terminal Asf1p-interacting region.^{1,31,39} We observed that *Trichomonas* and *Giardia* contained proteins that appeared to be comparably related to both Cac2p and HirA (ESI†) suggesting that the differentiation into replication-dependent and independent chaperones had not yet occurred in the ancestral eukaryotes. All other eukaryotic lineages showed a clear differentiation of two paralogous groups (ESI†), respectively, representing HirA and Cac2 implying that the differentiation of general and temporally specific histone chaperones first emerged prior to the separation of the heterolobosean–kinetoplastid clade from other eukaryotes. However, certain lineages show the apparent loss of either the Cac2p ortholog (*e.g.* the *Guillardia* nucleomorph) or the HirA ortholog (*e.g.* kinetoplastids) suggesting a continued functional overlap between them. Nevertheless, the presence of at least one of the two paralogs or their common ortholog (*Trichomonas* and *Giardia*) in all eukaryotes, even the

most reduced genomes, suggests that a HirA-like β -propeller provides an essential scaffold for H3–H4 assembly in conjunction with the Asf1 chaperone.

Hir3p orthologs are thus far only known from ascomycete fungi. By means of sequence profile searches with the PSI-BLAST program we were able to identify homologs of Hir3p in animals, amoebozoans, plants, chlorophyte alga, chromists and heteroloboseans (ESI†). Relative to the orthologs of the other Hir and Caf-1 complex components, such as Asf1, HirA and Cac2, these newly detected Hir3 homologs were extremely divergent in sequence (ESI†, requiring over 3 more PSI-BLAST iterations for recovery). However, this set of Hir3 homologs consistently recover each other in reciprocal searches and are present in a single copy in most eukaryotic lineages in which they are detected. This strongly suggests that they are the divergent orthologs of Hir3 in other eukaryotes. The animal ortholog of Hir3p is the calcineurin-binding protein Cabin1/Cain ($e = 10^{-5}$ in iteration 3). While Cabin1/Cain has as yet not been shown to interact with other histone chaperones, it does function as a corepressor for specific transcription factors like Mef-2 by binding Sin3 and recruiting histone deacetylases.^{40,41} This is consistent with the repressive role of the fungal Hir complex and supports a conserved histone-associated role for the orthologs of Hir3 in other eukaryotes. Interestingly, we observed that the Hir3–Cabin1/Cain orthologs from ascomycete fungi are rapidly diverging relative to the orthologs from other crown-group eukaryotes (ESI†). This indicates a functional shift in them, perhaps related to the loss in ascomycetes of the calcium-dependent function observed in animal Cabin1/Cain proteins.^{40–43} The phyletic patterns of Hir3–Cabin1/Cain orthologs suggest that it probably emerged a little later in eukaryotic evolution relative to HirA–Cac2 and Asf1. It was sporadically lost in several eukaryotic lineages (*e.g.* alveolates, basidiomycetes, kinetoplastids and reduced genomes like microsporidians and nucleomorphs) suggesting that it is not an absolutely required component of the Hir complex. We identified a conserved core in all eukaryotic Hir3p orthologs containing about 30 tetratricopeptide repeats (TPRs) that are likely to constitute multi-rung super-helical α – α superhelix (ESI†). It has been previously observed that in certain cellular complexes TPR and β -propeller proteins cooperate to form structural scaffolds.⁴⁴ Consistent with this, Hir3 might cooperate with the β -propellers of HirA to possibly constitute the scaffold of this histone chaperone complex with several distinct interaction sites.

Identification of a novel conserved domain in Hpc2 and its functional implications

Like Hir3p, Hpc2p has hitherto only been identified in ascomycete fungi. Compositional entropy analysis of the protein with the SEG program⁴⁵ showed that it is a predominantly low complexity protein with a small predicted folded region at the C-terminus. To identify potential Hpc2p homologs outside of fungi we performed PSI-BLAST searches with a position specific score matrix for this region. These searches recovered as significant hits from several eukaryotes including nuclear proteins from *Drosophila* and mammals, respectively,

yemanuclein- α and ubinuclein-1 (Ubn1) and the potential mammalian cell-cycle regulator FLJ25778 (KIAA2030).⁴⁶ We then prepared a hidden Markov Model of the alignment of this region culled from the PSI-BLAST searches and scanned all eukaryotic proteomes with the HMMER package to identify homologous regions. Consequently, we recovered significant hits ($e < 10^{-4}$) to this region in all eukaryotic lineages with the exception of *Giardia*, kinetoplastids and the *Guillardia* nucleomorph. The relationship of the region in these newly detected eukaryotic proteins with the globular region of ascomycete Hpc2 was further confirmed by means of profile-profile comparisons with the HHpred program (p -value 10^{-8}). We accordingly named this small conserved region (~ 45 –55 residues in length) the HUN

(for HPC2-Ubinuclein-1) domain. The above phyletic patterns also indicated that the HUN domain was already present in the last eukaryotic common ancestor (LECA) and that the majority of eukaryotic lineages contain a homolog of yeast Hpc2p (Fig. 1). While in *S. cerevisiae* and related ascomycetes the HUN domain is found at the C-terminus of a largely non-globular Hpc2p, homologs from basidiomycetes, animals and plants contain the HUN domain at the N-terminus (Fig. 1). These latter proteins contain a unique α -helical domain at their C-termini (Fig. 1, ESI†). Thus, ascomycete Hpc2 appears to have lost this domain after their divergence and undergone a drastic reorganization. Thus divergence patterns of the HUN domain proteins to certain extent parallel that of Hir3; *i.e.* rapid divergence in ascomycetes relative to other crown group

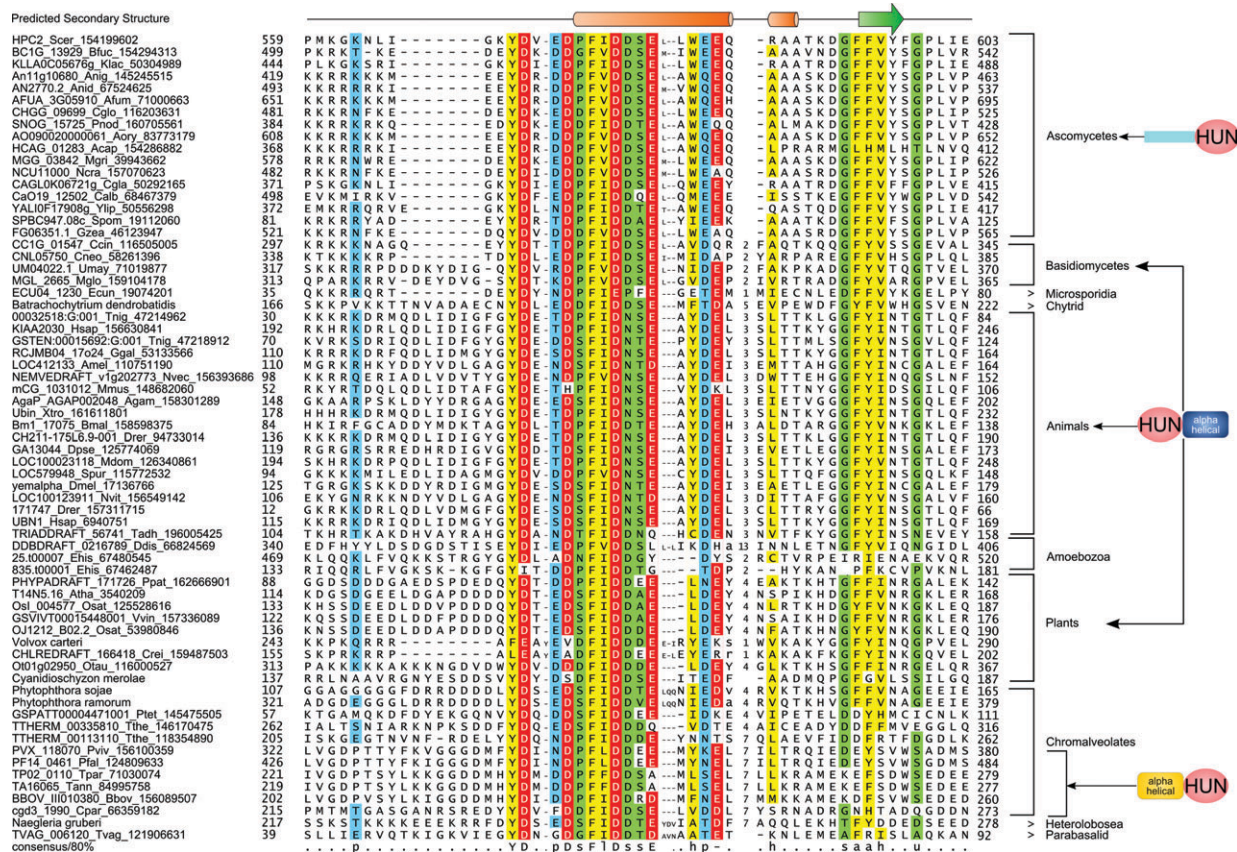


Fig. 1 Multiple sequence alignment and domain architectures of the HUN domain. Proteins are denoted by their gene name, species abbreviations and gi. Amino acids are colored based on their functional properties and conservation ($> 80\%$) in the alignment. The predicted secondary structure is shown at the top with α -helices represented as orange cylinders and β -strands as green block arrows. Domain architectures of HUN containing protein are shown to the right. Consensus abbreviations and coloring scheme are as follows: polar (p: QNSTCHKRDE) residues shaded blue, hydrophobic (h: ACFMWYIVL), aliphatic (l: IVL) and aromatic (a: FYWH) residues colored yellow, acidic ('-': DE) residues shaded red, small (s: GASVTDNPC) and tiny (u: GAS) residues colored green. Species abbreviations are as follows: Acap: *Ajellomyces capsulatus*; Afum: *Aspergillus fumigatus*; Agam: *Anopheles gambiae*; Amel: *Apis mellifera*; Anid: *Aspergillus nidulans*; Anig: *Aspergillus niger*; Aory: *Aspergillus oryzae*; Atha: *Arabidopsis thaliana*; Bbov: *Babesia bovis*; Bfuc: *Botryotinia fuckeliana*; Bmal: *Brugia malayi*; Calb: *Candida albicans*; Ccin: *Coprinopsis cinerea*; Cgla: *Candida glabrata*; Cglo: *Chaetomium globosum*; Cneo: *Cryptococcus neoformans*; Cpar: *Cryptosporidium parvum*; Crei: *Chlamydomonas reinhardtii*; Ddis: *Dictyostelium discoideum*; Dmel: *Drosophila melanogaster*; Dpse: *Drosophila pseudoobscura*; Drer: *Danio rerio*; Ecun: *Encephalitozoon cuniculi*; Ehis: *Entamoeba histolytica*; Ggal: *Gallus gallus*; Gzea: *Gibberella zeae*; Hsap: *Homo sapiens*; Klac: *Khuyveromyces lactis*; Mdom: *Monodelphis domestica*; Mglo: *Malassezia globosa*; Mgr1: *Magnaporthe grisea*; Mmus: *Mus musculus*; Nera: *Neurospora crassa*; Nvec: *Nematostella vectensis*; Nvit: *Nasonia vitripennis*; Osat: *Oryza sativa*; Otai: *Ostreococcus tauri*; Pfal: *Plasmodium falciparum*; Pnod: *Phaeosphaeria nodorum*; Ppat: *Physcomitrella patens*; Ptet: *Paramecium tetraurelia*; Pviv: *Plasmodium vivax*; Scer: *Saccharomyces cerevisiae*; Spom: *Schizosaccharomyces pombe*; Spur: *Strongylocentrotus purpuratus*; Tadh: *Trichoplax adhaerens*; Tann: *Theileria annulata*; Tnig: *Tetradodon nigroviridis*; Tpar: *Theileria parva*; Tthe: *Tetrahymena thermophila*; Tvag: *Trichomonas vaginalis*; Umay: *Ustilago maydis*; Vvin: *Vitis vinifera*; X(Si): *Xenopus* (Silurana); Ylip: *Yarrowia lipolytica*.

eukaryotes. Likewise, in apicomplexan Hpc2 homologs the HUN domain is linked to a unique α -helical N-terminal domain that is not in any other lineage (Fig. 1).

The mutual information derived from the phyletic pattern of the HUN domain proteins is strongly correlated with Hir1/2 and Asf1 (>0.9) (Table S2 and Fig. S7, ESI†). These observations and the low copy number of the Hpc2 homologs point to a common conserved function for HUN domain throughout eukaryotes. The domain contains a prominent central helix followed by a region predicted to adopt an extended conformation (Fig. 1). The characteristic conserved acidic residues of the HUN domain display a periodic signature of the form $DX_2DX_3DX_2EX_{2-3}[ED]$ (where X is any amino acid) and are mainly associated with the central helix. Hence, these residues are most likely to form a contiguous negatively charged exposed ridge on the protein. Crystal structures as well as mutagenesis studies have implicated the two other associated histone chaperones Asf1p and HirA in interacting with the globular head domains of the histone H3–H4 complex, but there is no evidence for these chaperones directly interacting with the non-globular tails.^{27,38,47–49} Based on the presence of the conserved acidic ridge we predict that the HUN domain is likely to be the histone-tail-binding subunit of the Hir complex. This prediction implies that the conserved HUN domain was an important innovation in LECA that was needed for deposition of the highly charged eukaryotic histones. Available experimental evidence on functions of HUN domain proteins from other eukaryotes, while not extensive, is consistent with the above proposed function. Both yem- α and Ubn1 have been shown to be strongly

expressed or ubiquitous nuclear proteins (<http://flybase.org> and <http://genome-www.stanford.edu/Human-CellCycle/Hela/search.shtml>) and Ubn1 has also been shown to localize and to punctuate nuclear spots, which resembles the pattern of other histone remodeling and modifying complexes.^{50,51}

HPC2 is critical for histone deposition in multiple contexts

To understand the specific contexts in which the role of Hpc2p might be important, we constructed its interaction network using the available protein–protein interaction and genetic interaction data in yeast (Fig. 2A; Fig. S3, ESI†). This network showed that Hpc2p is a critical component of the Hir histone chaperone complex displaying functional interactions with genes encoding proteins belonging to various functional classes, namely: (1) chromatin remodeling complexes such as Swi–Snf, Iswi, Swr, Ino80, FACT, CAF-1, RSC and the boundary-element-associated factor Yta7; (2) histones H3, H4, H2A and H2AZ; (3) RNA polymerase subunits like RPA43 and CDC73, some basal transcription factors like TAFs and components of the CDK8, head and middle sub-complexes of the Mediator complex; (4) RNA-polymerase-elongation-associated machinery, including, TFIIS, ELF1 and prefoldin complex proteins; (5) Several DNA replication initiation and repair proteins including PCNA and primase; (6) RNA-processing machinery such as the Kem1 nuclease, PAT1 and LSM complexes associated with polyA tails and caps, and mRNA export components of the nuclear pore (Nup84); (7) kinetochore assembly and spindle checkpoint components.

These wide-ranging functional interactions of Hpc2p are comparable to those of Hir1p, Hir2p and Hir3p and Asf1p,

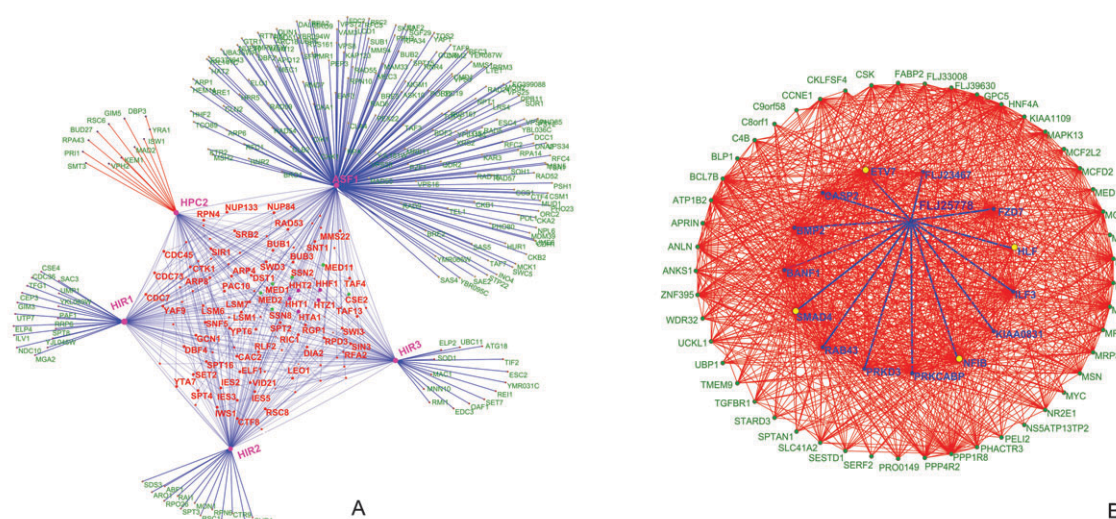


Fig. 2 (A) A network representation of genetic and protein interactions of Hir complex components (Hir1, Hir2, Hir3 and Hpc2) and Asf1. Nodes and edges in the network, respectively, represent genes/proteins and interactions between them. The shared interactions between two or more of the components of the Hir complex are shown within the central pentagon module, while the unique interactions of the components (that are not shared with other components of the Hir complex) are shown outside of the pentagon module. Asf1 having functions independent of the Hir complex (e.g. the Caf-1 complex) has a large set of unique interactions. Histones H2A, H3, H4, H2AZ are shown as purple nodes within the central pentagon module and components of the mediator/CDK8 complex are shown as green nodes (Med1, Med2, Ssn2, Cse2, Ssn8 and Med11) within the module. (B) Largest dense sub-network of proteins containing HUN domain protein FLJ25778 constructed based on mutual information in their phyletic patterns. The network has 64 human proteins implicated in cell-cycle function *via* RNAi knockdowns, which also contains FLJ25778, are shown as nodes. These nodes are connected by an edge if a given protein pair has a high MIC (>0.6) in their phyletic patterns (see section 4.1 in ESI† for details) and there are 908 edges in the network. The edges connecting protein FLJ25778 to other proteins in the network are shown in blue; among them transcription factors are shown as yellow nodes and the others are shown as blue nodes.

suggesting that it is an integral component of this nucleosome chaperone system (Fig. 2A). In particular genetic interactions (and in some cases physical interactions) with a range of specific chromatin remodeling complexes, replication, transcription and nuclear RNA processing complexes suggest that it can function as a generalized chaperone providing a backup for several specific histone deposition systems (Fig. 2A). While the Hir–Asf1 complex has been shown to physically interact with the H3–H4 dimer, their components, including Hpc2, also showed strong epistatic interactions with H2A specifically in the context of transcription (section 3.1 and Fig. S4, ESI†). Hence, it might be of interest to investigate if the above-proposed histone chaperone function of Hpc2p might also affect assembly of H2A in certain contexts. More strikingly, Hpc2 also showed specific genetic phenotypic enhancement to the exclusion of Hir1, Hir2, Hir3 and Asf1 with Isw1 and Rsc6. These genes, respectively, encode components of the Isw1 and Rsc chromatin remodeling complexes, which are important for nucleosome sliding (Fig. 2A and Fig. S3, ESI†). This raises the possibility that a histone-tail-specific chaperone function mediated by Hpc2 might be critical for this process. Independently of other histone chaperone subunits, Hpc2 also shows phenotypic suppression of synthetic lethality with specific interactions with kinetochore and spindle checkpoint genes like Mad2 and Bud27 suggesting that Hpc2's chaperone activity might also have a specific role in organization of centromeric chromatin (Fig. 2A). Likewise, other genetic interactions unique to Hpc2 are those with the prefoldin subunit Gim5p and RNA-processing components like Kem1p, Yra1p and DBP3p. These interactions suggest that Hpc2p is potentially critical in linking histone re-deposition upon passage of the RNA polymerases to both the prefoldin complex that associates with the transcription elongation machinery⁵² and the RNA-processing machinery that associates with the pre-mRNA or rRNA.⁵³

Given that the HUN domain itself is rather small, it is likely that it only mediates interactions with histones, whereas the remaining regions of these proteins are responsible for interactions with other chromatin components. In light of the wide divergence of the Hpc2 homologs beyond the HUN domain, including linkages to lineage-specific α -helical domains (Fig. 1, ESI†), it is quite possible that the interactions with chromatin components are distinct for different eukaryotic lineages. Consistent with this, vertebrate ubinuclein appears to have acquired a distinctive lineage-specific ability to bind specific transcription factors such as c-Jun.⁵⁰ A computational analysis combining mutual information from phyletic patterns and RNAi knockdowns in HeLa cell lines implicates the HUN domain protein FLJ25778 (but not ubinuclein) in cell-cycle progression and suggests potential functional interaction with several cell-cycle regulating transcription factors (Fig. 2B, ESI†). However, FLJ25778 mRNA levels show no significant changes throughout the cell-cycle implying that its chaperone activity or interactions could be post-translationally regulated during the course of cell-cycle progression (<http://genome-www.stanford.edu/Human-CellCycle/HeLa/search.shtml>). Unlike yeast Hpc2 and vertebrate ubinuclein and FLJ25778, we found that the *Plasmodium falciparum* Hpc2 homolog is strongly elevated in conjunction with a SET domain protein

(MAL6P1.131) and the RecQ/Wrn-like helicase during early schizogony and might have a specific role in chromatin reorganization or repair during this process. Thus, it is plausible that by virtue of rapid divergence and variability outside of the HUN domain Hpc2p homologs might link the ancient conserved Hir–Asf1 histone chaperone complex to distinct specific processes in different lineages.

General conclusions

Using sensitive protein sequence profile analyses we investigated the evolution of components of the Hir–Asf1 complex and showed that Hir3p and Hpc2p have a much wider phyletic pattern than was previously known. We identified the HUN domain in all Hpc2p homologs including ubinuclein and the recently discovered vertebrate cell-cycle regulator FLJ25778. The HUN domain has a characteristic pattern of acidic residues based on which we predict that it might function as a histone-tail-binding chaperone. By means of an analysis of various high-throughput data sets, such as RNAi knock-downs, genetic and protein interaction maps and cell-cycle wide gene expression, we were able to obtain evidence that Hpc2p homologs might be deployed in specific processes of chromatin dynamics relating to cell-cycle progression in vertebrates and schizogony in *Plasmodium*. Further, given the extensive divergence of regions of these proteins beyond the conserved HUN domain, we propose that Hpc2p homologs are important for recruitment of the ancient conserved Hir–Asf1 complex to different lineage-specific processes. We present evidence that both Hpc2 and Hir3 appear to have undergone a functional shift in ascomycete fungi implying that studying them in other model organisms might offer better clues regarding their general functions. The results and predictions presented here might help in exploring previously unexpected aspects of chromatin structure and organization.

Materials and methods

Profile searches were conducted using the PSI-BLAST program with a default profile inclusion expectation (*E*) value threshold of 0.01.⁵⁴ HMMs were constructed using hmmbuild and searches were performed using the hmmsearch from the HMMER package.⁵⁵ Profile–profile comparisons were performed using the HHpred program.⁵⁶ Multiple alignments were constructed using the Kalign⁵⁷ program followed by manual adjustments based on PSI-BLAST results. Protein secondary structure was predicted using a multiple alignment as the input for the JPRED program.⁵⁸ Then the BLASTCLUST program (<ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html>) was used to cluster proteins based on their sequence similarity score densities. The genetic and protein interactions data generated by Krogan *et al.* and Collins *et al.*^{59–61} were downloaded from the BIOGRID database.⁵⁶ We adopted a similar strategy to identify orthologs of proteins suggested to be involved in human cell cycle as described in Madan Babu *et al.*;⁶² also see ESI† for further details. Custom written PERL scripts were used to calculate mutual information coefficients (MICs) of phyletic profiles as described in Marcotte *et al.*⁶³ and Pellegrini *et al.*⁶⁴ The list of

63 genomes used in the analysis is listed in ESI† Table S3. Gene expression data for intraerythrocytic developmental cycle of *Plasmodium falciparum* was based on Bozdech *et al.*⁶⁵ and PlasmoDB.⁶⁶

Acknowledgements

SB, LMI and LA acknowledge the Intramural research program of National Institutes of Health, USA for funding their research.

References

- M. Eitoku, L. Sato, T. Senda and M. Horikoshi, Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly, *Cell. Mol. Life Sci.*, 2008, **65**, 414–444.
- A. Loyola and G. Almouzni, Histone chaperones: a supporting role in the limelight, *Biochim. Biophys. Acta*, 2004, **1677**, 3–11.
- S. Henikoff, Nucleosome destabilization in the epigenetic regulation of gene expression, *Nat. Rev. Genet.*, 2008, **9**, 15–26.
- C. W. Akey and K. Luger, Histone chaperones and nucleosome assembly, *Curr. Opin. Struct. Biol.*, 2003, **13**, 6–14.
- Y. J. Park and K. Luger, The structure of nucleosome assembly protein 1, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 1248–1253.
- G. J. Gemmen, R. Sim, K. A. Haushalter, P. C. Ke, J. T. Kadonaga and D. E. Smith, Forced unraveling of nucleosomes assembled on heterogeneous DNA using core histones, NAP-1, and ACF, *J. Mol. Biol.*, 2005, **351**, 89–99.
- K. Ura and K. Ohsumi, Chromatin assembly *in vitro* with a linker histone chaperone NAP-1, *Seikagaku*, 2006, **78**, 421–425.
- E. Lejeune, M. Bortfeld, S. A. White, A. L. Pidoux, K. Ekwall, R. C. Allshire and A. G. Ladurner, The chromatin-remodeling factor FACT contributes to centromeric heterochromatin independently of RNAi, *Curr. Biol.*, 2007, **17**, 1219–1224.
- L. Selth and J. Q. Svejstrup, Vps75, a new yeast member of the NAP histone chaperone family, *J. Biol. Chem.*, 2007, **282**, 12358–12362.
- D. Angelov, V. A. Bondarenko, S. Almagro, H. Menoni, F. Mongelard, F. Hans, F. Muetton, V. M. Studitsky, A. Hamiche, S. Dimitrov and P. Bouvet, Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes, *EMBO J.*, 2006, **25**, 1669–1679.
- Y. Lorch, B. Maier-Davis and R. D. Kornberg, Chromatin remodeling by nucleosome disassembly *in vitro*, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 3090–3093.
- P. Korber, S. Barbaric, T. Luckenbach, A. Schmid, U. J. Schermer, D. Blaschke and W. Horz, The histone chaperone Asf1 increases the rate of histone eviction at the yeast PHO5 and PHO8 promoters, *J. Biol. Chem.*, 2006, **281**, 5539–5545.
- P. Korber and W. Horz, SWRred not shaken; mixing the histones, *Cell*, 2004, **117**, 5–7.
- X. Shen, R. Ranallo, E. Choi and C. Wu, Involvement of actin-related proteins in ATP-dependent chromatin remodeling, *Mol. Cell*, 2003, **12**, 147–155.
- Y. M. Moshkin, J. A. Armstrong, R. K. Maeda, J. W. Tamkun, P. Verrijzer, J. A. Kennison and F. Karch, Histone chaperone ASF1 cooperates with the Brahma chromatin-remodelling machinery, *Genes Dev.*, 2002, **16**, 2621–2626.
- M. A. Martinez-Balbas, T. Tsukiyama, D. Gdula and C. Wu, Drosophila NURF-55, a WD repeat protein involved in histone metabolism, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 132–137.
- A. Nourani, F. Robert and F. Winston, Evidence that Spt2/Sin1, an HMG-like factor, plays roles in transcription elongation, chromatin structure, and genome stability in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.*, 2006, **26**, 1496–1509.
- R. Pavri, B. Zhu, G. Li, P. Trojer, S. Mandal, A. Shilatifard and D. Reinberg, Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II, *Cell*, 2006, **125**, 703–717.
- J. A. Sharp, E. T. Fouts, D. C. Krawitz and P. D. Kaufman, Yeast histone deposition protein Asf1p requires Hir proteins and PCNA for heterochromatic silencing, *Curr. Biol.*, 2001, **11**, 463–473.
- E. M. Green, A. J. Antczak, A. O. Bailey, A. A. Franco, K. J. Wu, J. R. Yates, 3rd and P. D. Kaufman, Replication-independent histone deposition by the HIR complex and Asf1, *Curr. Biol.*, 2005, **15**, 2044–2049.
- J. A. Sharp, G. Rizki and P. D. Kaufman, Regulation of histone deposition proteins Asf1/Hir1 by multiple DNA damage checkpoint kinases in *Saccharomyces cerevisiae*, *Genetics*, 2005, **171**, 885–899.
- G. Mizuguchi, X. Shen, J. Landry, W. H. Wu, S. Sen and C. Wu, ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex, *Science*, 2004, **303**, 343–348.
- E. Luk, N. D. Vu, K. Patteson, G. Mizuguchi, W. H. Wu, A. Ranjan, J. Backus, S. Sen, M. Lewis, Y. Bai and C. Wu, Chz1, a nuclear chaperone for histone H2AZ, *Mol. Cell*, 2007, **25**, 357–368.
- G. Mizuguchi, H. Xiao, J. Wisniewski, M. M. Smith and C. Wu, Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes, *Cell*, 2007, **129**, 1153–1164.
- S. Stoler, K. Rogers, S. Weitze, L. Morey, M. Fitzgerald-Hayes and R. E. Baker, Scm3, an essential *Saccharomyces cerevisiae* centromere protein required for G2/M progression and Cse4 localization, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10571–10576.
- L. Aravind, L. M. Iyer and C. Wu, Domain architectures of the Scm3p protein provide insights into centromere function and evolution, *Cell Cycle*, 2007, **6**, 2511–2515.
- R. Natsume, M. Eitoku, Y. Akai, N. Sano, M. Horikoshi and T. Senda, Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4, *Nature*, 2007, **446**, 338–341.
- A. Verreault, P. D. Kaufman, R. Kobayashi and B. Stillman, Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4, *Cell*, 1996, **87**, 95–104.
- P. D. Kaufman, R. Kobayashi and B. Stillman, Ultraviolet radiation sensitivity and reduction of telomeric silencing in *Saccharomyces cerevisiae* cells lacking chromatin assembly factor-I, *Genes Dev.*, 1997, **11**, 345–357.
- V. Lamour, Y. Lecluse, C. Desmaze, M. Spector, M. Bodescot, A. Aurias, M. A. Osley and M. Lipinski, A human homolog of the *S. cerevisiae* HIR1 and HIR2 transcriptional repressors cloned from the DiGeorge syndrome critical region, *Hum. Mol. Genet.*, 1995, **4**, 791–799.
- H. DeSilva, K. Lee and M. A. Osley, Functional dissection of yeast Hir1p, a WD repeat-containing transcriptional corepressor, *Genetics*, 1998, **148**, 657–667.
- T. Formosa, S. Ruone, M. D. Adams, A. E. Olsen, P. Eriksson, Y. Yu, A. R. Rhoades, P. D. Kaufman and D. J. Stillman, Defects in SPT16 or POB3 (yFACT) in *Saccharomyces cerevisiae* cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure, *Genetics*, 2002, **162**, 1557–1571.
- J. A. Sharp, A. A. Franco, M. A. Osley and P. D. Kaufman, Chromatin assembly factor I and Hir proteins contribute to building functional kinetochores in *S. cerevisiae*, *Genes Dev.*, 2002, **16**, 85–100.
- D. C. Krawitz, T. Kama and P. D. Kaufman, Chromatin assembly factor I mutants defective for PCNA binding require Asf1/Hir proteins for silencing, *Mol. Cell. Biol.*, 2002, **22**, 614–625.
- D. Dimova, Z. Nackerdien, S. Furgeson, S. Eguchi and M. A. Osley, A role for transcriptional repressors in targeting the yeast Swi/Snf complex, *Mol. Cell*, 1999, **4**, 75–83.
- P. Prochasson, L. Florens, S. K. Swanson, M. P. Washburn and J. L. Workman, The HIR corepressor complex binds to nucleosomes generating a distinct protein/DNA complex resistant to remodeling by SWI/SNF, *Genes Dev.*, 2005, **19**, 2534–2539.
- T. Krude, Chromatin assembly: the kinetochore connection, *Curr. Biol.*, 2002, **12**, R256–R258.
- S. M. Daganzo, J. P. Erzberger, W. M. Lam, E. Skordalakes, R. Zhang, A. A. Franco, S. J. Brill, P. D. Adams, J. M. Berger and P. D. Kaufman, Structure and function of the conserved core of histone deposition protein Asf1, *Curr. Biol.*, 2003, **13**, 2148–2158.
- S. Lorain, S. Demczuk, V. Lamour, S. Toth, A. Aurias, B. A. Roe and M. Lipinski, Structural Organization of the WD repeat

- protein-encoding gene HIRA in the DiGeorge syndrome critical region of human chromosome 22, *Genome Res.*, 1996, **6**, 43–50.
- 40 A. Han, F. Pan, J. C. Stroud, H. D. Youn, J. O. Liu and L. Chen, Sequence-specific recruitment of transcriptional co-repressor Cabin1 by myocyte enhancer factor-2, *Nature*, 2003, **422**, 730–734.
 - 41 H. D. Youn and J. O. Liu, Cabin1 represses MEF2-dependent Nur77 expression and T cell apoptosis by controlling association of histone deacetylases and acetylases with MEF2, *Immunity*, 2000, **13**, 85–94.
 - 42 M. J. Kim, D. G. Jo, G. S. Hong, B. J. Kim, M. Lai, D. H. Cho, K. W. Kim, A. Bandyopadhyay, Y. M. Hong, D. H. Kim, C. Cho, J. O. Liu, S. H. Snyder and Y. K. Jung, Calpain-dependent cleavage of cain/cabin1 activates calcineurin to mediate calcium-triggered cell death, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9870–9875.
 - 43 H. D. Youn, C. M. Grozinger and J. O. Liu, Calcium regulates transcriptional repression of myocyte enhancer factor 2 by histone deacetylase 4, *J. Biol. Chem.*, 2000, **275**, 22563–22567.
 - 44 Y. Zhang and D. C. Chan, Structural basis for recruitment of mitochondrial fission complexes by Fis1, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18526–18530.
 - 45 J. C. Wootton, Non-globular domains in protein sequences: automated segmentation using complexity measures, *Comput. Chem.*, 1994, **18**, 269–285.
 - 46 R. Kittler, L. Pelletier, A. K. Heninger, M. Slabicki, M. Theis, L. Miroslaw, I. Poser, S. Lawo, H. Grabner, K. Kozak, J. Wagner, V. Surendranath, C. Richter, W. Bowen, A. L. Jackson, B. Habermann, A. A. Hyman and F. Buchholz, Genome-scale RNAi profiling of cell division in human tissue culture cells, *Nat. Cell Biol.*, 2007, **9**, 1401–1412.
 - 47 A. D. Malay, T. Umehara, K. Matsubara-Malay, B. Padmanabhan and S. Yokoyama, Crystal structures of fission yeast histone chaperone Asf1 complexed with the Hip1 B-domain or the Cac2 C terminus, *J. Biol. Chem.*, 2008, **283**, 14022–14031.
 - 48 H. J. Kim, J. H. Seol, J. W. Han, H. D. Youn and E. J. Cho, Histone chaperones regulate histone exchange during transcription, *EMBO J.*, 2007, **26**, 4467–4474.
 - 49 P. D. Kaufman, J. L. Cohen and M. A. Osley, Hir proteins are required for position-dependent gene silencing in *Saccharomyces cerevisiae* in the absence of chromatin assembly factor I, *Mol. Cell. Biol.*, 1998, **18**, 4793–4806.
 - 50 S. Aho, M. Buisson, T. Pajunen, Y. W. Ryoo, J. F. Giot, H. Gruffat, A. Sergeant and J. Uitto, Ubinuclein, a novel nuclear protein interacting with cellular and viral transcription factors, *J. Cell. Biol.*, 2000, **148**, 1165–1176.
 - 51 O. Ait-Ahmed, B. Bellon, M. Capri, C. Joblet and M. Thomas-Delaage, The yemanuclein-alpha: a new Drosophila DNA binding protein specific for the oocyte nucleus, *Mech. Dev.*, 1992, **37**, 69–80.
 - 52 C. T. Parusel, E. A. Kritikou, M. O. Hengartner, W. Krek and M. Gotta, URI-1 is required for DNA stability in *C. elegans*, *Development*, 2006, **133**, 621–629.
 - 53 W. Luo, A. W. Johnson and D. L. Bentley, The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model, *Genes Dev.*, 2006, **20**, 954–965.
 - 54 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
 - 55 S. R. Eddy, Profile hidden Markov models, *Bioinformatics*, 1998, **14**, 755–763.
 - 56 J. Soding, A. Biegert and A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res.*, 2005, **33**, W244–W248.
 - 57 T. Lassmann and E. L. Sonnhammer, Kalign—an accurate and fast multiple sequence alignment algorithm, *BMC Bioinformatics*, 2005, **6**, 298.
 - 58 J. A. Cuff and G. J. Barton, Application of multiple sequence alignment profiles to improve protein secondary structure prediction, *Proteins*, 2000, **40**, 502–511.
 - 59 N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. R. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rillstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili and J. F. Greenblatt, Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*, *Nature*, 2006, **440**, 637–643.
 - 60 S. R. Collins, K. M. Miller, N. L. Maas, A. Roguev, J. Fillingham, C. S. Chu, M. Schuldiner, M. Gebbia, J. Recht, M. Shales, H. Ding, H. Xu, J. Han, K. Ingvarsdottir, B. Cheng, B. Andrews, C. Boone, S. L. Berger, P. Hieter, Z. Zhang, G. W. Brown, C. J. Ingles, A. Emili, C. D. Allis, D. P. Toczyski, J. S. Weissman, J. F. Greenblatt and N. J. Krogan, Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map, *Nature*, 2007, **446**, 806–810.
 - 61 S. R. Collins, P. Kemmeren, X. C. Zhao, J. F. Greenblatt, F. Spencer, F. C. Holstege, J. S. Weissman and N. J. Krogan, Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*, *Mol. Cell. Proteomics*, 2007, **6**, 439–450.
 - 62 M. Madan Babu, S. A. Teichmann and L. Aravind, Evolutionary dynamics of prokaryotic transcriptional regulatory networks, *J. Mol. Biol.*, 2006, **358**, 614–633.
 - 63 E. M. Marcotte, M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates and D. Eisenberg, Detecting protein function and protein–protein interactions from genome sequences, *Science*, 1999, **285**, 751–753.
 - 64 M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg and T. O. Yeates, Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 4285–4288.
 - 65 Z. Bozdech, M. Llinas, B. L. Pulliam, E. D. Wong, J. Zhu and J. L. DeRisi, The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*, *PLoS Biol.*, 2003, **1**, E5.
 - 66 A. Bahl, B. Brunk, J. Crabtree, M. J. Fraunholz, B. Gajria, G. R. Grant, H. Ginsburg, D. Gupta, J. C. Kissinger, P. Labo, L. Li, M. D. Mailman, A. J. Pearson, D. S. Roos, J. Schug, C. J. Stoeckert, Jr and P. Whetzel, PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data, *Nucleic Acids Res.*, 2003, **31**, 212–215.