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 intensively 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 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
of the eukaryotic crown group (fungi, animals and plants), with comparable functions and interactions. 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, oomyctes 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. While this configuration is conserved throughout the crown group, in saecharomyxete yeasts the HirA ortholog has undergone a lineage-specific duplication spanning 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. 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 algae, 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 co-repressor for specific transcription factors like Mef-2 by binding Sin3 and recruiting histone deacetylases. 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. 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 tetraoctapeptide 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. 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 program 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-2 and ubinuclein-1 (Ubn1) and the potential mammalian cell-cycle regulator FLJ25778 (KIAA2030). 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 ($e < 10^{-5}$) 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 all-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

### Table 1

<table>
<thead>
<tr>
<th>Domain</th>
<th>Species Abbreviations</th>
</tr>
</thead>
</table>

*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: QNSTCKHRDE) residues shaded blue, hydrophobic (h: ACFMWYIVL) 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: *Kluyveromyces lactis*; Mdom: *Monodelphis domestica*; Mglo: *Malassezia globosa*; Mgri: *Magnaporthe grisea*; Mmus: *Mus musculus*; Nce: *Neurospora crassa*; Nvec: *Nematostella vectensis*; Nvit: *Nasonia vitripennis*; Otat: *Orzya sativa*; Otau: *Ostreococcus tauri*; Pfat: *Plasmodium falciparum*; Pnd: *Phaeosphaeria nodorum*; Ppat: *Physcomitrella patens*; Ptet: *Paramecium tetraurelia*; Pvin: *Plasmodium vivax*; Pthe: *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 ζ-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 DX2DX3DX2EX2–3[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 negatively charged exposed ridge on the protein. Crystal structures of the HUN domain proteins from other eukaryotes, while not extensive, is consistent with the above proposed function.

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, TFIIIS, ELF1 and prefoldin complex proteins; (5) Several DNA replication initiation and repair proteins including PCNA and primase; (6) RNA-processing machinery such as the Ksn1 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.
suggested 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). This suggests that the Hpc2 chaperone activity 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 Iswi and Rsc chromatin remodeling complexes, which are important for nucleosome sliding (Fig. 2A and Fig. S3). 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 Bub27 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 Hpc2 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 machinery and the RNA-processing machinery that associates with the pre-mRNA or rRNA. 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 z-helical domains (Fig. 1), it is quite possible that the interactions with chromatin components are distinct for different euchromatic lineages. Consistent with this, vertebrate ubiquitin appears to have acquired a distinctive lineage-specific ability to bind specific transcription factors such as c-Jun. A computational analysis combining mutual information from phylectic patterns and RNAi knockdowns in Hela cell lines implicates the HUN domain protein FLJ25778 (but not ubiquitin) in cell-cycle progression and suggests potential functional interaction with several cell-cycle regulating transcription factors (Fig. 2B). 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 ubiquitin and FLJ25778, we found that the \textit{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 Hpc2 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 phylectic 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 knockdowns, 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 \textit{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. HMMS were constructed using hmmbuild and searches were performed using the hmmssearch from the HMMER package. Profile–profile comparisons were performed using the HHPred program. Multiple alignments were constructed using the Kalign 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. 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. were downloaded from the BIOGRID database. 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.; also see ESI for further details. Custom written PERL scripts were used to calculate mutual information coefficients (MICs) of phylectic profiles as described in Marcotte et al. and Pellegrini et al. The list of
36 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. and PlasmoDB.

Acknowledgements

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

References

34 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 deposition proteins Asf1/Hir1 by multiple DNA damage checkpoint

View Article Online


51 T. Lassmann and E. L. Sonnhammer, Kalign—an accurate and fast multiple sequence alignment algorithm, BMC Bioinformatics, 2005, 6, 298.


