Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Protein and DNA Modifications: Evolutionary Imprints of Bacterial Biochemical Diversification and Geochemistry on the Provenance of Eukaryotic Epigenetics

L. Aravind, A. Maxwell Burroughs, Dapeng Zhang, and Lakshminarayan M. Iyer

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

Correspondence: aravind@ncbi.nlm.nih.gov

Epigenetic information, which plays a major role in eukaryotic biology, is transmitted by covalent modifications of nuclear proteins (e.g., histones) and DNA, along with poorly understood processes involving cytoplasmic/secreted proteins and RNAs. The origin of eukaryotes was accompanied by emergence of a highly developed biochemical apparatus for encoding, resetting, and reading covalent epigenetic marks in proteins such as histones and tubulins. The provenance of this apparatus remained unclear until recently. Developments in comparative genomics show that key components of eukaryotic epigenetics emerged as part of the extensive biochemical innovation of secondary metabolism and intergenomic/interorganismal conflict systems in prokaryotes, particularly bacteria. These supplied not only enzymatic components for encoding and removing epigenetic modifications, but also readers of some of these marks. Diversification of these prokaryotic systems and subsequently eukaryotic epigenetics appear to have been considerably influenced by the great oxygenation event in the Earth's history.

It has long been recognized that eukaryotes possess several subcellular systems with no apparent equivalents in the two prokaryotic domains (superkingdoms) (Dacks and Doolittle 2001; Best et al. 2004; Mans et al. 2004; Walsh and Doolittle 2005; Aravind et al. 2006; Cavalier-Smith 2009). Hence, a major challenge in modern biology has been to explain the provenance of these uniquely eukaryotic features. Among these, the extensive use of epigenetic information in regulatory systems is a key paradigm that has fructified in the past two decades (Richards and Elgin 2002; Allis et al. 2007; Kouzarides 2007; Grewal 2010). Broadly defined, epigenetics might be viewed as transmission of biologically significant information over and beyond what is encoded by the standard bases in DNA (i.e., genetic information). It has become increasingly clear that the nucleus is the primary center for encoding of epigenetic information in eukaryotes (Denhardt et al. 2005; Allis et al. 2007; Kouzarides 2007). Here, it largely

Editors: Patrick J. Keeling and Eugene V. Koonin
Additional Perspectives on The Origin and Evolution of Eukaryotes available at www.cshperspectives.org

 doi: 10.1101/cshperspect.a016063
Cite this article as *Cold Spring Harb Perspect Biol* 2014;6:a016063

occurs via covalent modifications of DNA or DNA-associated proteins (chromatin proteins). Eukaryotes also show certain less-understood cytoplasmic forms of epigenetic transmission. These include modifications of cytoskeletal proteins, protein-based templating (i.e., prionic transmission) (Beauregard et al. 2009), and RNA-based information transmission in phenomena such as paramutation in plants (Brzeski and Brzeska 2011) and postconjugation macronucleus regeneration in ciliates (Mochizuki 2010).

Epigenetic information impinges on fundamental aspects of eukaryotic biology such as DNA replication, DNA-damage repair, transcription of specific genes, global control of gene expression, splicing and other types of RNA processing, and exhibition of metabolically or structurally distinct cellular states (Richards and Elgin 2002; Allis et al. 2007; Kouzarides 2007; Grewal 2010). In specialized eukaryotes, such as parasites, epigenetic information plays an important role in the display of variable cell-surface antigens to evade host immunity (Duraisingh et al. 2005; Jiang et al. 2013). In multicellular forms such information is central to maintenance of a structured body plan (Muller et al. 2002; Gehring et al. 2006; Allis et al. 2007), dedicated immune systems (Cedar and Bergman 2011), and phenomena such as neural memory in animals (Landry et al. 2013). Epigenetic modifications of secreted proteins paralleling those of chromatin proteins have also been shown to be the primary determinants for specification of structures of unique biomineralized matrices such as bones (Tagliabracci et al. 2013) and siliceous shells of diatoms (Kroger et al. 2002; Sumper et al. 2007). Thus, understanding the origins of epigenetics is a major element in reconstructing eukaryote origins, including the emergence of their quintessential feature, the nucleus.

Here, we offer a synthetic perspective on the origin of DNA and protein modification systems used to transmit epigenetic information in eukaryotes. Based on results from comparative genomics we emphasize their pervasive connections to bacterial secondary metabolism and interorganismal and genomic conflict systems.

THE LOGIC OF EPIGENETIC MARKS: ENCODERS, RESETTERS, AND READERS

Epigenetics in eukaryotes can be conceptualized as three distinct processes (Fig. 1): (1) encoding of epigenetic information into biopolymers; (2) resetting of these marks at key points in the life cycle of an organism; and (3) reading of these marks to convert them into biologically "relevant" outputs. The first process is almost entirely dependent on enzymes, which specifically modify bases of nucleic acids or protein side chains by a striking array of moieties (encoders) (Figs. 1–3). The former include methylation and subsequent oxidation of methylcytosine at the 5-position in DNA (Goll and Bestor 2005; Pastor et al. 2013). The best studied of the latter are modifications of nucleosomal histones by moieties ranging from small groups, such as methyl, acetyl, and phosphate, through medium-sized adducts, such as sugars, all the way to giant modifiers such as polyADP ribose (with more than 100 ADP-ribose units), polyglutamate/glycine, or whole polypeptides (i.e., ubiquitin [Ub] and ubiquitin-like proteins [Ubls]) (Allis et al. 2007; Kouzarides 2007; Yan et al. 2009; Zentner and Henikoff 2013). These modifications often occur at low-complexity or structurally disordered oligopeptides in proteins (Cumberworth et al. 2013), which serve as linear arrays wherein information is encoded in the form of various covalent modifications (e.g., positively charged histone tails). In some cases (e.g., histones), different combinations of modifications of particular side chains are often viewed as comprising a code ("histone code") (Dutnall 2003; Peterson and Laniel 2004; Kouzarides 2007). These patterns of modification are seen as "coding for" or specifying particular chromatin states (e.g., active transcription, repression, or poised for expression upon reception of additional signals).

Although epigenetic marks often persist through mitosis, and in certain cases through meiosis (Scott and Spielman 2006), they are reset during events such as zygote formation in multicellular eukaryotes (Hajkova et al. 2010). Like encoding, resetting of most marks involves

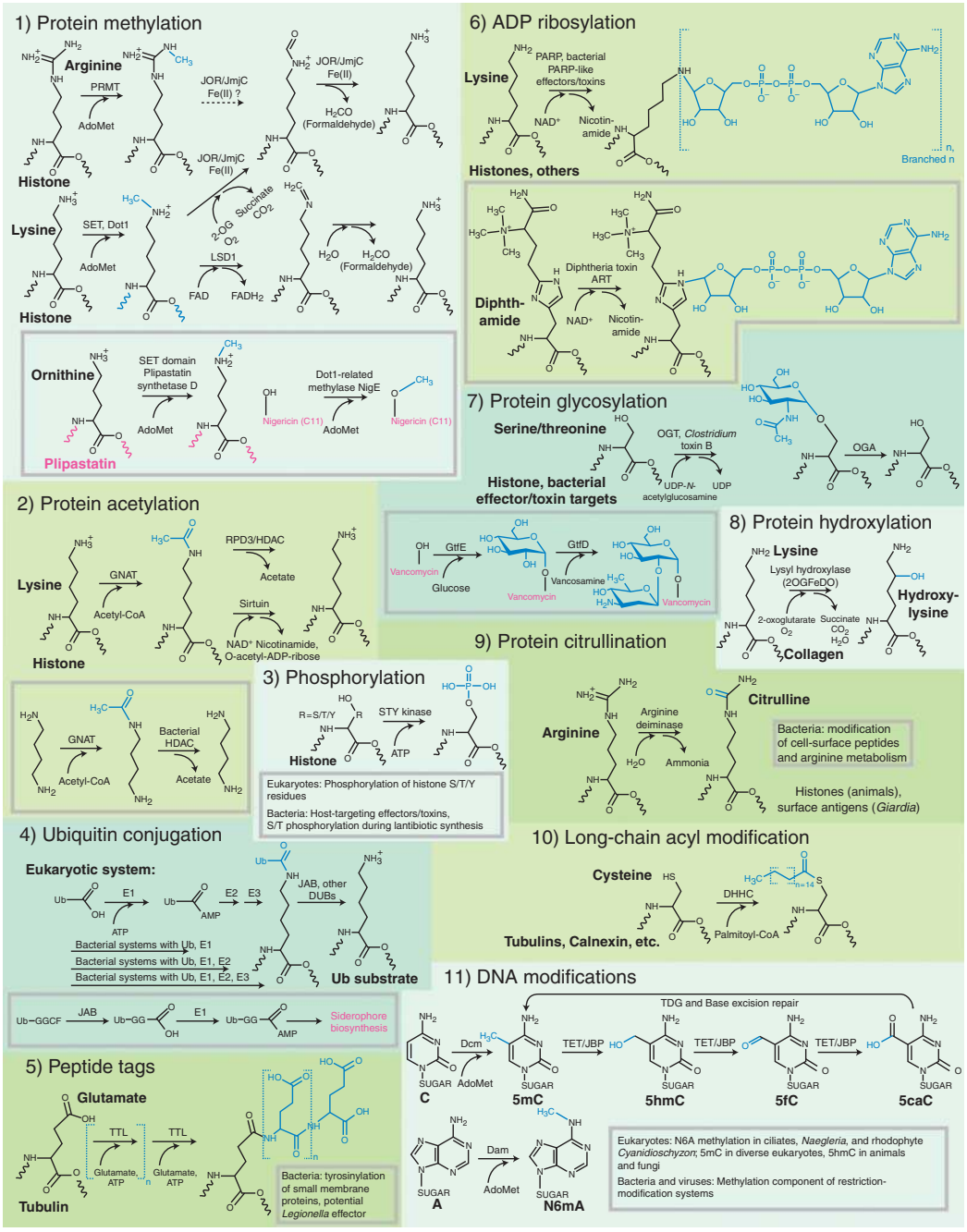

Figure 1. Reactions relating to epigenetic modifications. Reactions are numbered in the order in which they appear in the text. Targets of various modifications and, where applicable, reaction intermediates are labeled in bold. Modifying chemical groups attached during the reaction are colored in blue. Eukaryotic reactions are provided immediately below numbers and descriptions; comparable prokaryotic reactions or descriptions of prokaryotic substrates are provided to the left or below the eukaryotic reaction, boxed in gray. Modified prokaryotic molecules are labeled in pink to distinguish them from eukaryotic substrates. Ub, ubiquitin; TDG, thymine DNA glycosylase.

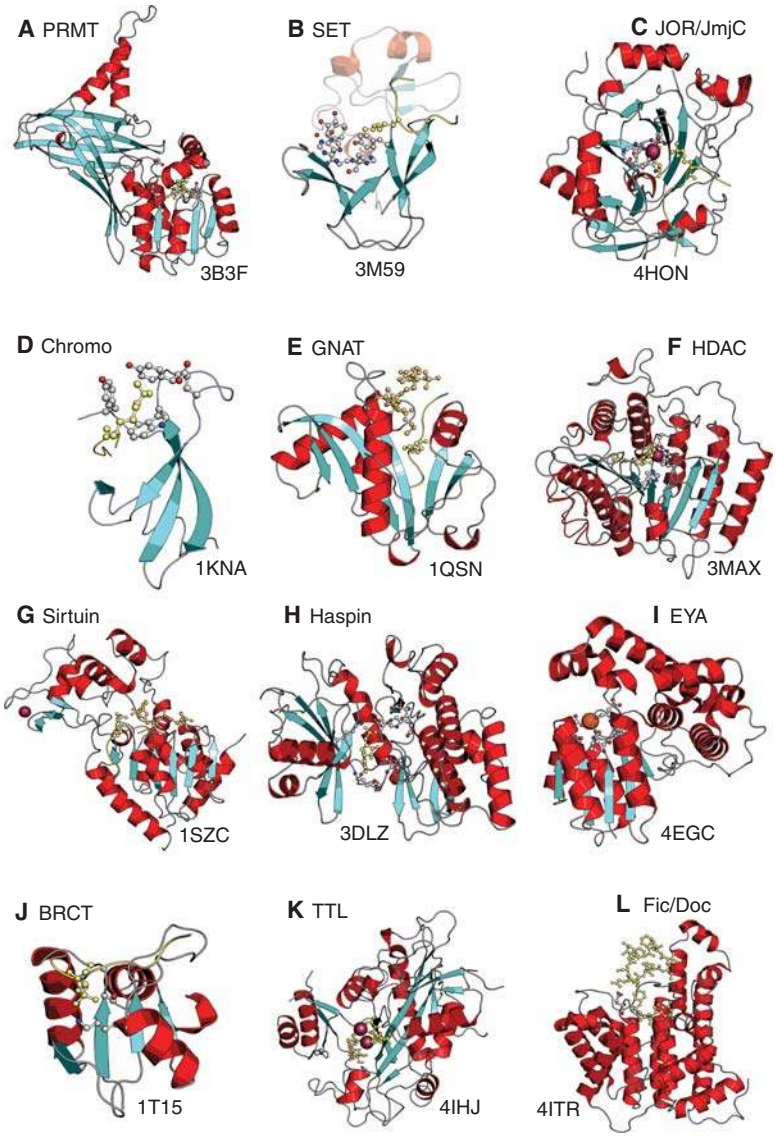

Figure 2. Structures of domains involved in various epigenetic modifications. Protein structures are depicted as cartoons. Domains are shown in the order they are discussed in the text. Domain names are provided above and Protein Databank id is provided to the right of each structure. Ligands are colored in yellow, core active site residues are rendered as ball-and-stick and colored according to atom type (carbon, white; nitrogen, blue; oxygen, red; cysteine, orange), and metal ions are rendered as spheres.

enzymes that catalyze removal of the covalently linked adducts (Figs. 1–3). However, resetting is also assisted by other processes such as dilution owing to semiconservative genome replication and consequent partitioning of chromatin proteins, proteasomal degradation, and repair, which erases epigenetic marks on DNA (Hajkova et al. 2010; Pastor et al. 2013). In contrast to encoders and resetters, readers of epigenetic marks are almost all noncatalytic, globular modules, which specifically discriminate between modified and nonmodified versions

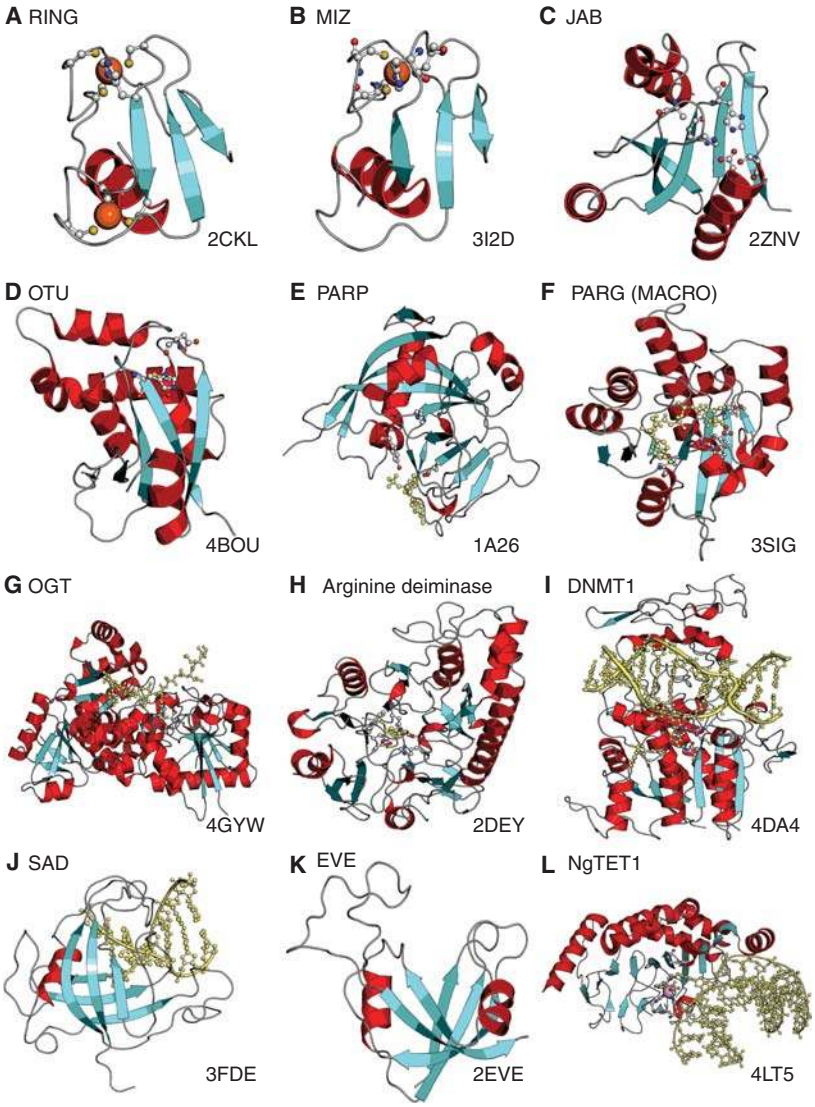

Figure 3. Structures of domains involved in various epigenetic modifications (see Fig. 2 legend for details).

of DNA or protein. Readers typically come in two forms: (1) small domains with high-specificity binding sites for peptides or DNA (Dhalluin et al. 1999; Maurer-Stroh et al. 2003; Chakravarty et al. 2009; Aravind et al. 2011); and (2) superstructure-forming repetitive units such as β-propellers and α-α repeats (Collins et al. 2008; Trievel and Shilatifard 2009; Aravind et al. 2011). Generation of biologically relevant outputs from epigenetic marks often depends on polypeptides combining reader domains with different types of enzymatic domains, which in addition to encoders and resetters, include enzymes that use free energy of ATP hydrolysis to remodel chromatin structure (e.g., SWI2/SNF2 [Hauk and Bowman 2011] and MORC ATPases [Iyer et al. 2008a]). Thus, epigenetic encoding involves multiple layers of interactions: marks generated by primary encoders recruit resetters and secondary encoders, with cross talk between epigenetic marks mediated by reader domains (Allis et al. 2007).

In the ensuing sections we provide a brief account of the major epigenetic marks in proteins and DNA along with the inferred evolutionary origin of encoders, resetters, and readers associated with them.

PROTEIN METHYLATION-DEPENDENT SYSTEMS

Methylation of proteins on lysines and arginines are universally present epigenetic marks in eukaryotes (Fig. 1, No. 1). Protein methylases, which encode these marks, belong to two structurally unrelated folds, namely, the Rossmann-fold methylases (Fig. 2A) and the SET domain methylases (Fig. 2B) containing the β-clip fold (Trievel et al. 2002; Manzur et al. 2003; Sawada et al. 2004; Lee and Stallcup 2009; Aravind et al. 2011). Other than histones, these marks also occur in several other chromatin and splicing/RNA-processing proteins with arginine-rich repeats (Miranda et al. 2005; Anne et al. 2007; Nicholson and Chen 2009). Rossmann-fold protein methylases belong to two distinct families—PRMT and Dot1; the former catalyze all known arginine methylations, whereas the later methylates histone H3 at the K79 position (Dlakic 2001; Sawada et al. 2004; Lee and Stallcup 2009; Aravind et al. 2011). Phylogenetic analysis suggests that the protein arginine methyltransferase (PRMT) family had already diversified in the last eukaryotic common ancestor (LECA), as indicated by the presence of multiple members of this family in the early-branching parabasalid *Trichomonas vaginalis* with at least one version catalyzing symmetric arginine dimethylation and two distinct versions catalyzing asymmetric dimethylation (Aravind et al. 2011). With the exception of *Trichomonas* and *Giardia*, Dot1 orthologs are present in most other major eukaryotic lineages suggesting that they could have been recruited after the divergence of basal eukaryotes. SET domains methylate histone H3 at K4, K9, K27 and K36 and histone H4 at K20; additionally they might also catalyze several other lesser understood lysine methylations in histones (Kouzarides 2007; Zentner and Henikoff 2013). Positions corresponding to H3K4, H3K9, H3K27, H3K36, and H4K20 are confidently inferred as being lysine even in the LECA. All eukaryotic genomes sequenced to date include genes for multiple SET domain proteins, with at least five distinct versions traceable to the LECA (Aravind et al. 2011). SET proteins also appear to have a role in cytoplasmic epigenetic marks on tubulin, ribosomal proteins, and RUBISCO (Trievel et al. 2002; Porras-Yakushi et al. 2007). An extraordinary role for methyl marks has come to light in the form of secreted SET domains in diatoms, which are predicted to be involved in establishing an "epigenetic" code in the secreted protein silaffin (Aravind et al. 2011). This code of modified residues in silaffin is a key determinant for the biomineralization patterns in silica shells of diatoms (Sumper et al. 2007).

Methyl marks are reset by two distinct families of demethylases (Fig. 1, No. 1): (1) The LSD1-like lysine demethylases are FAD-binding Rossmann-fold oxidoreductases and primarily demethylate H3K4me1 and H3K4me2 (Chen et al. 2006; Nicholson and Chen 2009). (2) The Jumonji-related (JOR or JmjC) enzymes are 2-oxoglutarate/iron-dependent dioxygenases of the double-stranded β-helix fold (Fig. 2C) and are by far the most prevalent demethylases in eukaryotic chromatin (Klose et al. 2006; Tsukada et al. 2006; Iyer et al. 2010; Aravind et al. 2011). Unlike LSD1-demethylases, these can demethylate mono-, di-, and trimethylated lysines, and perhaps the different forms of methylated arginines. Some members of this family also catalyze formation of other potential epigenetic marks such as hydroxylated asparagine in proteins and RNA modifications (e.g., modified base hydroxywybutosine in tRNA) (Elkins et al. 2003; Iyer et al. 2010). Strikingly, unlike the SET and Rossmann-fold methylases, both LSD1 and JOR/JmjC demethylases are absent in the parabasalids and diplomonads (Iyer et al. 2008b), raising the possibility that there was no active mechanism for resetting methyl marks in the LECA.

Readers of methyl marks include structurally diverse domains (Yap and Zhou 2010): (1) simple globular domains such as the chromo-like domains with the SH3 fold (Fig. 2D) and possibly catalytically inactive versions of the

JOR/JmjC domain (Jacobs and Khorasanizadeh 2002; Maurer-Stroh et al. 2003; Brehm et al. 2004; Shimojo et al. 2008). By far, chromo-like domains constitute the most versatile class of methylated histone-binding domains, recognizing methylation at H3K4, H3K9, H3K27, H3K36, and H4K20. The initial radiation of the chromo-like domains in eukaryotes appears to coincide with the expansion of protein methylases that happened before the LECA (Aravind et al. 2011). (2) Metal-chelation-supported domains include versions of the treble-clef fold typified by the PHD-finger domain and its structural derivatives, which specialize in recognition of H3K4me2/3 and H3K9me2/3 and might also bind H3K14ac, H4S, acetylated amino termini of histone, and nonacetylated peptides (Chakravarty et al. 2009). At least a single copy of the methylated H3K4-recognizing PHD finger is inferred as having been present in the LECA. (3) Superstructure-forming repeats that include versions of the WD40 and ankyrin repeats bind methylated histones (Collins et al. 2008; Trievel and Shilatifard 2009), of which at least the former might have been present in the LECA.

Strikingly, close homologs of all methylases and demethylases involved in eukaryotic epigenetic modification are found in bacteria as components of biosynthetic systems for secondary metabolites, such as antibiotics and siderophores, from modified amino acids or peptides (Aravind et al. 2011). In some cases, these domains might be embedded within gigantic multidomain synthetases for nonribosomally synthesized peptides (e.g., the SET domain in plipastatin synthetase subunit D) or in operons for biosynthesis of metabolites (e.g., the Dot1-related methylase NigE in the polyether antibiotic nigericin gene cluster) (Walsh 2003; Walsh et al. 2005; Tsuge et al. 2007; Aravind et al. 2011). Moreover, both Dot1-like and SET methylases are also found as effectors secreted into eukaryotic hosts by various endosymbiotic and parasitic bacteria (Fig. 4). Similarly, bacterial homologs of classical SH3 and chromo-like domains are found in secreted or periplasmic proteins associated with peptidoglycan (Fig. 4).

PROTEIN ACETYLATION-DEPENDENT SYSTEMS

The ancient superfamily of *N*-acetyltransferases, typified by the GCN5 (GNAT domain) (Fig. 2E), catalyze lysine and amino-terminal acetylation (Fig. 1, No. 2), which act as major epigenetic marks in chromatin (Neuwald and Landsman 1997; Dutnall et al. 1998; Liu et al. 2008). Some of these have also been recently shown to catalyze addition of other acyl groups (e.g., crotonyl) to histones (Montellier et al. 2012). At least 14 distinct families of the GNAT superfamily have specialized roles in eukaryotic chromatin, of which conservatively four can be traced back to the LECA (Iyer et al. 2008b). Of these, one targeted H3 (Gcn5-like) and a second H4 (Esa1-like) to respectively specify transcriptionally active and silent states (Durant and Pugh 2006). The third (Elp3-like) targeted both these histones simultaneously in the context of transcription elongation (Wittschieben et al. 1999; Winkler et al. 2002). In contrast, the fourth activity (Kre33) is apparently directed toward chromatin-associated ribonucleoprotein complexes and is needed for ribosomal assembly (Oeffinger et al. 2007; Ossareh-Nazari et al. 2010). Elp3p and Kre33p have clear archaeal cognates suggesting an inheritance from the archaeal progenitor of eukaryotes. On several occasions lineage-specific GNATs introducing acetyl marks in eukaryotic histones were acquired repeatedly from bacteria, in which several GNATs acetylate side chains of peptide-derived antibiotics as part of resistance mechanisms or polyamines as part of their assimilation (Fig. 1, No. 2) (Leipe and Landsman 1997; Forouhar et al. 2005; Ramirez and Tolmasky 2010).

Acyl modifications are reversed (Fig. 1, No. 2) by two groups of histone deacetylases belonging to structurally distinct superfamilies, namely, the RPD3/HDAC (Fig. 2F) superfamily and the sirtuin (Sir2) superfamily (Fig. 2G) (Leipe and Landsman 1997; Avalos et al. 2004; Blander and Guarente 2004), both of which are inferred as being present in the LECA (Iyer et al. 2008b). Prokaryotic members of both superfamilies from which the eukaryotic representatives were

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

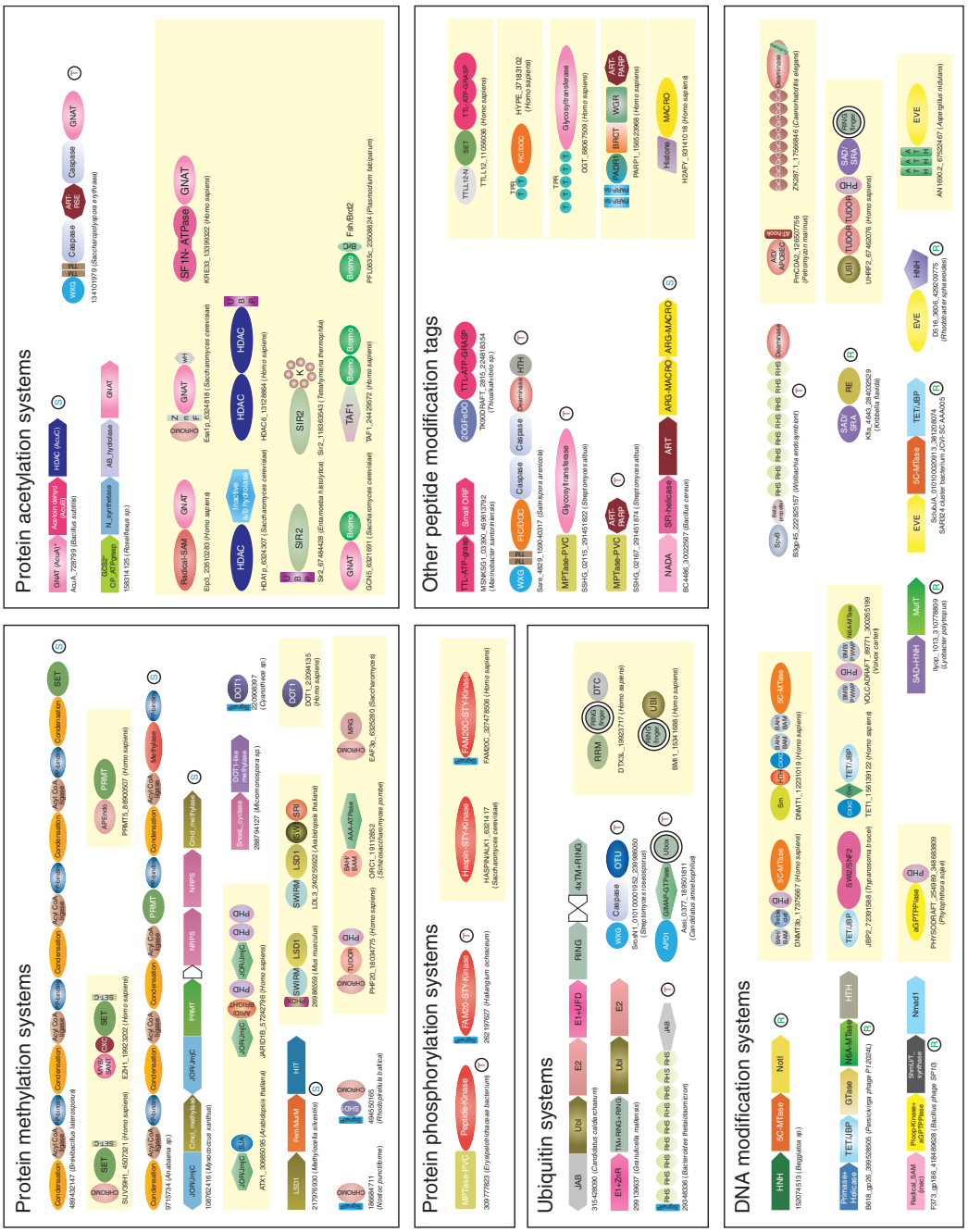

Figure 4. *(See next page for legend.)*

derived appear to have played predominantly metabolic roles. Representatives of the RPD3/HDAC superfamily appear to have a role in acetoin/polyamine metabolism (Ramirez and Tolmasky 2010), whereas those of the sirtuin superfamily regulate acyl CoA biosynthesis and nicotinamide dinucleotide (NAD) metabolism (Avalos et al. 2004; de Souza and Aravind 2012). In addition to a representative from the archaeal precursor, eukaryotes appear to have acquired several sirtuins from bacteria in the course of their evolution. Acetylated peptides on histone H3 and H4 are primarily recognized by the tetrahelical bromo domain (Zeng and Zhou 2002), of which at least four representatives can be traced to the LECA. The presence of a bromo domain in the basal transcription factor TAF1 and the Fsh/BDF1 protein that interacts with acetylated H4 in association with TFIID indicates an ancestral role for reading of acyl marks in regulation of transcription initiation (Martinez-Campa et al. 2004; Durant and Pugh 2006).

PHOSPHORYLATION-DEPENDENT EPIGENETIC MARKS

From the earliest studies on signaling it became clear that eukaryotes use a wide array of serine/threonine/tyrosine (STY) kinases (Fig. 1, No. 3) as opposed to histidine kinase-dependent phosphotransfer relays that are dominant in bacteria (West and Stock 2001; Manning et al. 2002). Phosphorylation of residues as epigenetic marks is typified by histone phosphorylation at a wide range of positions catalyzed by more than 10 distinct lineages of kinases in eukaryotes: histone H2AS1; variant histone H2AXS139 and Y142; histone H2B S14; histone H3T3, S10, S11, S28, Y41, and T45; and histone H4S1 (Rossetto et al. 2012; Zentner and Henikoff 2013). These marks are functionally distinct from other phosphorylation-dependent events in the sense that they specify distinct chromatin states as opposed to being a switch for initiation of a catalytic cascade (e.g., kinase cascades) (Johnson and Hunter 2005). Reconstruction of the phosphorylation landscape in the LECA is not easy owing to the general promiscuity of kinases in terms of targets and their lack of diagnostic fusions to chromatin-protein-specific domains. However, certain ancient STY-kinase clades such as the casein-kinase/NHK-1-like, the Ste20/Mst-like and ATM-like (belonging to the lipid kinase clade) clades appear to have been present in the LECA suggesting that histone phosphorylation marks catalyzed by them were probably present. Additionally, the LECA can also be inferred as possessing the unique H2AX Y142-specific WSTF kinase (Xiao et al. 2009). This enzyme is structurally and mechanistically unrelated to the STY kinases (Fig. 2H) and contains a catalytic cysteine that mediates phosphotransfer.

In the course of eukaryotic evolution several additional histone kinases were acquired: the CDK8 kinase with the cyclin-C partner as part of the Mediator Cdk8 complex (Conaway and Conaway 2011), CDK7 with the cyclin-H partner as part of the TFIIH complex (Egly and Coin

Figure 4. (*Continued*) A comparison of protein contexts of domains involved in eukaryotic epigenetic systems with domain and genome contexts of their prokaryotic counterparts. Proteins and operons are labeled by their names, Genbank index (GI), and species name. For operons, the name is derived from the principal domain being compared between the prokaryotic and eukaryotic versions. Eukaryotic versions are shown in a yellow background. Operons are shown as arrows with the arrowhead pointing to the gene in the 3′ orientation of the coding strand. Prokaryotic representatives were divided as those that are either involved in bacterial secondary metabolism systems (S), toxin/effector systems (T), or restriction-modification (R) systems. All domains are labeled as in the text or as per the Pfam database. HIT, Histidine triad; wH, winged helix-turn-helix; K-Kelch, TUDOR, and BAH/BAM are chromo-like SH3 fold domains; WXG, the ESX/Type 7 secretory system domain; NADA, the NADAR domain involved in ADP-ribose metabolism; ARG, ADP-glycohydrolase, GTase, glycosyltranferases; Cys, cysteine-rich domain; aGPTPPlase, α-glutamyl/putrescinyl thymine pyrophosphorylase; MutT, NUDIX, HNH, metal-chelating endonuclease domain; SpvB, a domain of the SpvB-type secretory system; PADR1, domain is found in PARPs (poly-ADP-ribose polymerases); brC, bromo-carboxy-terminal domain; GCS2, glutamyl cysteine synthetase-2; DTC, DTX-specific domain; APendo, AP endonuclease.

2011), haspin-like kinases with a distinctive amino-terminal domain of the kinase module (Higgins 2010), and the JAK tyrosine kinases with a duplication of the kinase module (Griffiths et al. 2011). In terms of resetters of phosphate marks, the situation in the LECA is rather unclear. It appears that the EYA-like HAD-fold phosphatase (Fig. 2I) that resets H2AX Y142P (Cook et al. 2009; Krishnan et al. 2009) is unlikely to have been present owing to its absence in most basal eukaryotic lineages. However, calcineurin-like ST phosphatases were definitely present in the LECA suggesting that these could have potentially reset any phosphate marks on serines or threonines. Among readers of phosphate marks the conserved 14-3-3 module comprised of α-α repeats, the all-β FHA, SJA/FYR, and BRCT domain (both with α/β folds) (Fig. 2J) are traceable to the LECA (Lloyd et al. 2009; Zippo et al. 2009; Garcia-Alai et al. 2010; Singh et al. 2012). However, the histone H3T3-binding BIR domain appears to be a later innovation in eukaryotic evolution, probably emerging concomitantly with haspin-like kinases that modify the position recognized by them (Jeyaprakash et al. 2011).

Although almost all archaea contain a few S/T/Y kinases, they lack the rich diversity of kinase domains that can be reconstructed as being present in the LECA (Leonard et al. 2004; Kannan et al. 2007; Aravind et al. 2010). However, certain bacteria such as myxobacteria, cyanobacteria, and actinobacteria have a rich array of STY kinases related to and comparable to those seen in eukaryotes (Kannan et al. 2007; Aravind et al. 2010). Such STY kinases are also seen as part of biosynthetic operons for lantibiotics where they phosphorylate S/T to generate an intermediate for dehydration (Goto et al. 2010). Related kinases are also the active principle of secreted host-targeting effectors of several pathogens and polymorphic toxins used in interbacterial conflicts (Zhang et al. 2012). This raises the possibility that, like eukaryotic protein methylases, S/T/Y kinases (except the unique WSTF) were also acquired by the stem eukaryote from bacteria followed by explosive proliferation before the LECA. Like kinases, even the FHA and BRCT domains appear to have been acquired from bacteria, in which the latter domain functions as a DNA end-binding domain in the context of DNA repair (Mueller et al. 2008). Other than histones, lineage-specific secreted STY kinases (e.g., Golgi casein-kinase/FAM20 and silaffin kinase) phosphorylate low-complexity extracellular matrix proteins with specific patterns of serine/threonine residues or sugars establishing an epigenetic code for directing biomineralization of tissue matrix (Sheppard et al. 2010; Tagliabracci et al. 2013). These kinases have been derived from related secreted kinases from bacteria (e.g., *Haliangium* gi: 262197627).

THE UBIQUITIN SYSTEM

Until recently, the triligase system comprised of E1, E2, and E3 enzymes and deubiquitinases (DUBs), which respectively link Ub or related Ubls to lysine through an isopeptide bond (more infrequently to cysteines, terminal NH_2, and lipids) and remove them through hydrolysis, were considered a unique feature of eukaryotes (Hochstrasser 2009). However, it has become clear that prokaryotes possess a diverse array of antecedents of eukaryotic Ub systems (Fig 1, No. 4), which include simple versions with just E1s, those with E1 and E2, and complete triligase systems with RING finger E3s (Fig. 3A) and JAB domain DUBs (Nunoura et al. 2011; Burroughs et al. 2012). Thus, the Ub system appears to have emerged in its complete form first in prokaryotes and was acquired in toto by the ancestral eukaryote from a prokaryotic source (Fig. 5). By the time of the LECA this system had undergone a spectacular expansion with about 7 E1, 20 E2s, and at least 18 RING domain E3s (Burroughs et al. 2012). Although poly-Ub tags were initially described for their role in proteasomal protein degradation, it is now clear that several of these modifications have signaling and epigenetic roles. In terms of the latter, the best known are the mono-Ub marks on histones: H2AK119, H2BK120, and H4K91, which are respectively introduced by the RING1/Bmi (part of polycomb repressive complex), RNF20/RNF40, and DTX3L clades (Lanzuolo and Orlando 2012; Wright et al.

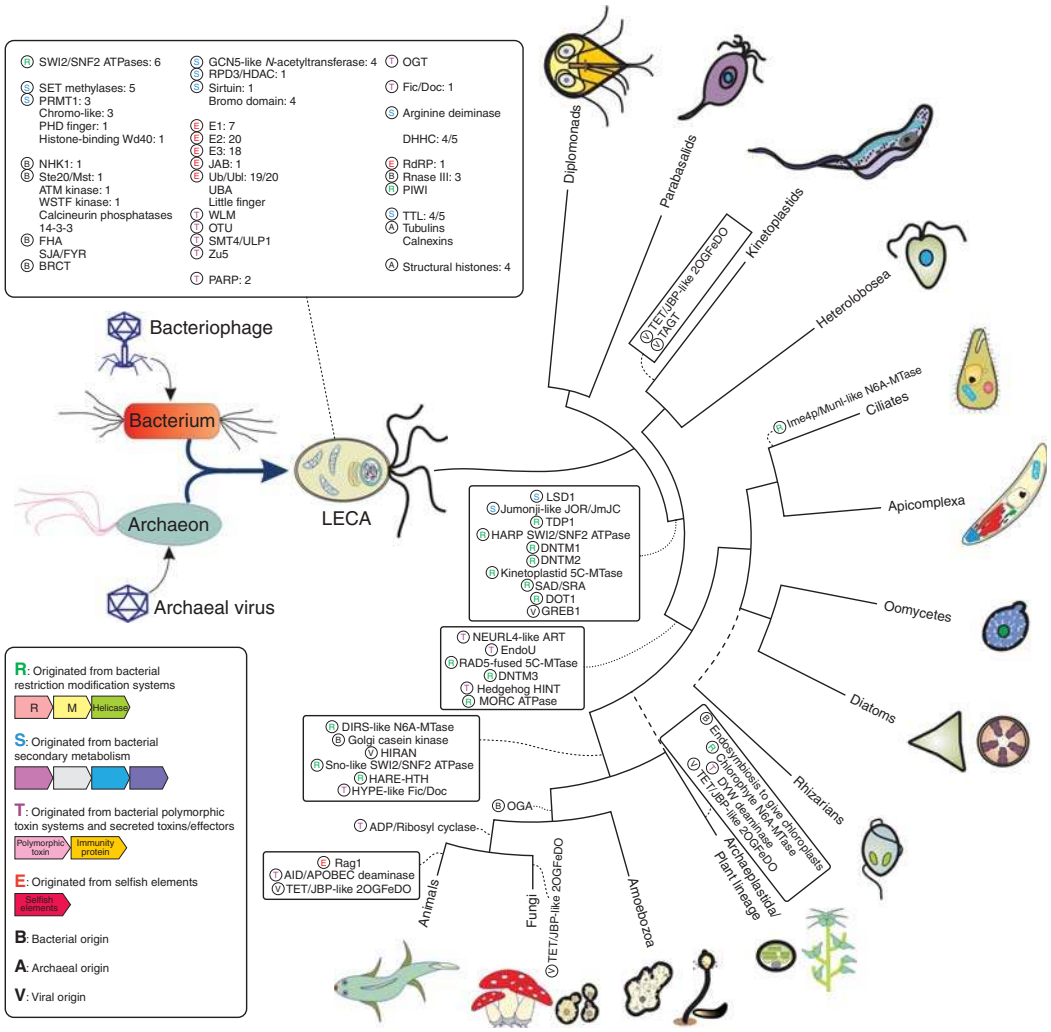

Figure 5. Evolutionary origins of various domains involved in eukaryotic epigenetic systems. Using a eukaryotic tree as reference, the source and the reconstructed points of acquisition of various domains involved in epigenetic systems are illustrated. Prokaryotic systems from which these domains were recruited include bacterial restriction-modification systems (R), secondary metabolism (S), various bacterial toxin systems (T) including polymorphic toxin systems, toxin-antitoxin systems, and secreted effector toxins secreted by various pathways, selfish elements (E), genes of bacterial origin (B), genes of archaeal origin (A), and those of viral origin (V). The inset showing the domains in the LECA additionally provide the reconstructed numbers of domain in this ancestor. The dashed lines indicate an uncertainty in the phylogenetic positions of the lineages indicated.

2012; Zentner and Henikoff 2013). Of these, only RNF20/RNF40, and hence, H2B ubiquitination, can be traced to the LECA (Fig. 5).

Other than histones, nuclear proteins, especially specific transcription factors and chromatin proteins, are significantly overrepresented among targets of the Ubl, SUMO, which is conjugated by a distinctive variant of the RING E3 ligase (Fig. 3B), the MIZ finger (Heun 2007; Venancio et al. 2009). At least a subset of these nuclear SUMO modifications potentially functions as epigenetic marks. Divergence of the classical RING and MIZ E3s predated the LECA: given the preponderance of SUMO tar-

gets among nuclear proteins, it is conceivable that Ub-SUMO divergence and the parallel divergence of their respective E3s corresponded to the emergence of the eukaryotic nucleus. Moreover, the origin of SUMO also probably marked the emergence of specific nuclear substructures such as the nucleolus and the so-called promyelocytic leukemia (PML) bodies (Heun 2007; Lallemand-Breitenbach and de The 2010). In the course of eukaryotic evolution, additional Ub E3 ligases emerged to introduce epigenetic marks in histones, such as the RING finger domain of the metazoan Rag1 recombinase, which modifies histone H3 as part of marking sites for immune receptor diversification (Grazini et al. 2010). Ub-binding UBA and the "little finger"-type Zn-ribbon domains are found in several chromatin proteins and are inferred as being present from the LECA itself (Fig. 5).

Although eukaryotes acquired JAB-type DUBs (Fig. 3C) as part of the multicomponent Ub systems inherited from prokaryotes (Burroughs et al. 2011), they also show two other peptidase superfamilies as DUBs, namely, WLM metallopeptidases and diverse papain-like peptidases (Fig. 3D). To date the latter superfamilies of peptidases have never been found in association with any prokaryotic Ub-like systems. Versions of both of these, including the OTU-like and SMT4/ULP1-like deSUMOylases are abundantly represented among effectors of endosymbiotic bacteria (especially amoebozoan endosymbionts such as *Amoebophilus* and *Odyssella*) (Fig. 4), and several bacterial toxins, suggesting that they were probably recruited independently from such systems (Schmitz-Esser et al. 2010; Zhang et al. 2012).

OTHER PEPTIDE TAGS ADDED TO PROTEINS

Eukaryotes also show a range of nonribosomally synthesized peptides that modify carboxyl termini or side chains of proteins (Fig. 1, No. 5) with tyrosyl, polyglutamyl, or polyglycyl moieties ranging in size from a single amino acid to more than 20 amino acids (Janke et al. 2008; Fukushima et al. 2009). These modifications are best known in the carboxy-terminal tails of α- and β-tubulin and are catalyzed by ATP-grasp enzymes of the TTL family (Fig. 2K). A subset of polyglutamylases and polyglycinases also modify the histone chaperones NAP1 and NAP2 (Janke et al. 2008). Thus, these modifications could serve as epigenetic marks both in the cytoplasm and nucleus. The TTL family is inferred as having already diversified in the stem eukaryote with at least four to five versions capable of different modifications in the LECA (Fig. 5). This is consistent with the reconstruction of the LECA as a multiflagellated organism utilizing epigenetic tags for organization of its microtubular cytoskeleton (Simpson et al. 2006; Zhang and Aravind 2012). Eukaryotes with expanded ciliary cytoskeletons, such as ciliates and parabasalids, show massive expansions of TTLs consistent with emergence of intricate, epigenetic mechanisms for ciliary positioning. Members of the TTL family arose as part of the vast radiation of ATP-grasp enzymes in bacteria secondary metabolism systems (Iyer et al. 2009). A subset of bacterial TTLs appears to function as modifiers of certain small membrane proteins with glutamate- and lysine-rich tails, whereas others appear to be effectors of endoparasitic legionellae (Fig. 4). It is probable that acquisition of such an enzyme from an endosymbiont was a key event in the emergence of the cytoskeleton of the stem eukaryote (Iyer et al. 2009).

ADP RIBOSYLATION

Numerous eukaryotic proteins, including histones, are both mono- and poly-ADP-ribosylated (Fig. 1, No. 6). PolyADP ribosylation of lysine with more than 200 ADP-ribose units in nucleosomal histones helps specify open chromatin states associated with enhanced transcription and DNA repair (Ame et al. 2004; Zentner and Henikoff 2013). Although members of the sirtuin superfamily can catalyze mono-ADP ribosylation, the best-understood modifications of histones are catalyzed by the poly-ADP-ribose polymerase (PARP) family (Fig. 4E) that belongs to the ADP-ribosyltransferase (ART) superfamily (Laing et al. 2011; de Souza and Aravind 2012). At least two PARPs, including the histone-modifying PARP1 are

reconstructed as being present in the LECA (Citarelli et al. 2010). However, in the course of eukaryotic evolution they expanded extensively giving rise to other clades such as the telomere-protein-modifying tankyrase and the vPARP, which is a subunit of the vault, a small noncoding RNA-associated organelle (Citarelli et al. 2010). Thus, use of ADP ribosylation as an epigenetic mark might extend beyond histones. Clear bacterial antecedents of the PARPs were recently identified among the toxin domains of bacterial toxins secreted via the novel PVC secretory system (Zhang et al. 2012). Thus, PARPs emerged as part of the dramatic diversification of ARTs among bacterial toxins, which have evolved a wide range of specificities to target diverse proteins including specific highly modified amino acids such as diphthamide (*Corynebacterium diptheriae* ART) (Fig 1, No. 6) (Laing et al. 2011). Indeed, ARTs related to toxins and effectors of endosymbionts, such as *Waddlia*, and entomotoxic bacteria have been independently acquired by certain eukaryotic lineages (e.g., Neuralized, a eukaryotic protein ART and ARTs-modifying guanine in DNA acquired by lepidopterans as regulators of apoptosis) (de Souza and Aravind 2012; Zhang et al. 2012). Like ARTs, a potential reader of this modification, the MACRO domain, and two types of ADP-ribohydrolases (Fig. 3F), which remove this mark, have been acquired from bacterial type-II toxin-antitoxin systems (Fig. 4) (de Souza and Aravind 2012).

PROTEIN TAGGING BY GLYCOSYL, NUCLEOTIDYL, AND RELATED MOIETIES

The majority of eukaryotic proteins are believed to be glycosylated, with distinct pathways for *N*-linked and *O*-like glycosylation, and separate nuclear and endoplasmic systems for the latter (Zoldos et al. 2013). However, epigenetic roles of glycosylation are poorly understood, except for recent studies on β-*N*-acetylglucosamine modification of serine/threonine in histones H2A, H2B, and H4 by the glycosyltransferase OGT, and its removal by the glycohydrolase OGA (Fig. 1, No. 7) (Hanover et al. 2012; Zentner and Henikoff 2013). This modification can compete with phosphorylation and has been implicated in specifying both repressive and active chromatin states probably modulated by nutrient availability (Hanover et al. 2012). OGT (Fig. 3G) with carboxy-terminal tetratricopeptide repeats can be traced to the LECA (Fig. 5), but appears to have been lost in several lineages suggesting that it is not an essential modification in several eukaryotes. OGA in contrast is a later acquisition from bacteria probably entering only in the common ancestor of animal, fungi, and amoebozoa (Fig. 5). Recently, another widespread eukaryotic OGT clade prototyped by GREB1 was identified (Iyer et al. 2013). It combines an *O*-glycosyltransferase to an amino-terminal circularly permuted superfamily-II helicase domain and binds DNA, suggesting that it might generate potential epigenetic marks by modifying chromatin proteins or hydroxylated DNA bases.

Like the OGTs, the Fic/Doc superfamily (Fig. 2L) targets hydroxyl groups of protein side chains. At least a single member of this superfamily can be inferred as present in the LECA (Fig. 5). Members of this family possess diverse protein modification activities such as AMPylation, UMPylation, and phosphorylcholination of hydroxyl groups of serine, threonine, or tyrosine in target proteins (Engel et al. 2012; Feng et al. 2012; Campanacci et al. 2013). Although the specificity of versions traceable to the LECA is not yet known, eukaryotes appear to have subsequently acquired a second version from bacteria (seen in animals, fungi, and plants) typified by the human HYPE protein, which AMPylates tyrosine side chains of the cytoskeleton-regulating GTPase Rho (Worby et al. 2009). This modification thus has the potential for functioning as a cytoplasmic epigenetic mark. The Fic/Doc domains are major toxin domains in bacteria among type-II toxin-antitoxin systems, effectors targeting eukaryotic hosts, and polymorphic toxins (Zhang et al. 2012). Like their eukaryotic cellular counterparts, most studied bacterial enzymes target host-signaling enzymes and interfere with their action (Engel et al. 2012; Campanacci et al. 2013).

In addition to reversible modifications, certain epigenetic marks are also generated by

irreversible modification of amino acid side chains such as hydroxylation of asparagine, lysine and proline (Fig. 1, No. 8), and deimination of arginine (Fig. 1, No. 9) to form citrulline (Dann et al. 2002; Vossenaar et al. 2003; Iyer et al. 2010). Although 2-oxoglutarate-Fe-dependent dioxygenases (2OGFeDOs) and JOR/JmjC enzymes catalyzing the former modifications are apparently not found in parabasalids and diplomonads, they are found in all other eukaryotes; hence, it is not clear if these modifications existed in the LECA (Iyer et al. 2010). The citrulline-generating arginine deiminase (Fig. 3H) appears to have been acquired by certain eukaryotic lineages such as metazoans, fungi, and *Giardia* on multiple occasions independently from bacterial precursors involved in modification of cell-surface peptides or arginine metabolism (Aravind et al. 2011). In metazoans, citrullination of arginines on histone H2A, H3, and H4 appear to be nuclear epigenetic marks (Vossenaar et al. 2003), whereas in *Giardia* it is used to mark cytoplasmic tails of expressed variant surface antigens (Touz et al. 2008).

LONG-CHAIN ACYL MODIFICATIONS: POSSIBLE MEMBRANE-ASSOCIATED EPIGENETICS?

In principle, modification of proteins by hydrophobic moieties could target them to specific regions of inner or plasma membranes, thereby facilitating transmission of epigenetic information, especially in the context of asymmetric cell divisions (Resh 2006). Although this is poorly understood, an attractive candidate for such a modification is *S*-palmitoylation of cysteines (Fig. 1, No. 10) by the eukaryote-specific DHHC domains (Blaskovic et al. 2013; Linder and Jennings 2013). Indeed, proteins ancestral to all eukaryotes (i.e., tubulins and the endoplasmic reticular chaperone calnexin [implicated in transmission of epigenetic states]) (Beauregard et al. 2009) are targets for multiple palmitoylations by these enzymes (Blaskovic et al. 2013; Linder and Jennings 2013). Conservatively, the ancestral eukaryote can be predicted as possessing at least four to five DHHC palmitoyltransferases (Fig. 5), with a major expansion of these in the course of their later evolution. In contrast, depalmitoylating enzymes of the α/β-hydrolase fold probably arose much later via lateral transfer from bacteria (Blaskovic et al. 2013).

EPIGENETIC MODIFICATIONS OF DNA

The most common DNA modification in eukaryotes is cytosine methylation (5mC) (Fig. 1, No. 11) (Gommers-Ampt and Borst 1995). Some eukaryotes also show methylation of adenine on the NH_2 group at the sixth position of the purine ring (N6mA) (Fig. 1, No. 11) (Gommers-Ampt and Borst 1995). Pyrimidines with exocyclic methyl groups are also the focus of oxidative modifications: hydroxylation of thymine followed by *O*-glycosylation to β-D-glucosyl-hydroxymethyluracil (base J) occurs in euglenozoans (Borst and Sabatini 2008) and comparable serial hydroxylation of 5mC (Fig. 1, No. 11) to generate 5-hydroxymethylcytosine (5hmC), formylcytosine, and carboxycytosine in several eukaryotes (Pastor et al. 2013). Recent studies also predict the potential formation of hypermodified thymines (conjugated to glutamate or putrescine) in basidiomycete fungi and certain chlorophytes from hydroxylated thymine (Iyer et al. 2013). Another modification, catalytic deamination of cytosine in DNA, has thus far only been confirmed in vertebrates (Rogozin et al. 2007), although detection of divergent deaminase domains related to the DNA deaminases point to a possibly more widespread distribution of this modification (Iyer et al. 2011b). Strikingly, unlike protein modifications, currently available genomic evidence does not point to any conserved DNA-modification machinery traceable to the LECA (Iyer et al. 2011a). Barring the possibility of drastic gene loss, DNA methylation arose probably first after the most basal "excavate" lineages had diverged from the rest of Eukarya.

All other DNA-modification systems appear to have been elaborated later and are often restricted to a few eukaryotic lineages (Fig. 5). For example, N6A methylation has been detected only in ciliates and chlorophytes (Gommers-Ampt and Borst 1995), and is predicted in the

heterolobosean *Naegleria* and the rhodophyte alga *Cyanidioschyzon* (Iyer et al. 2011a). The biggest resource from which eukaryotic DNA-modification enzymes were recruited comprises systems deployed in intergenomic conflicts between bacteria and invasive DNA (Fig. 4), such as viruses (Bickle and Kruger 1993; Roberts et al. 2010; Iyer et al. 2011a). Eukaryotic DNA 5mC methylases have been recruited from restriction-modification systems on at least five distinct occasions, including the widely conserved versions prototyped by DNMT1, DNMT2, and DNMT3 (Figs. 3I and 5), as well as several other lineage-specific versions (Iyer et al. 2011a). Likewise, reader domains with the PUA-like fold, which discriminate between methylated and unmethylated DNA (SAD/SRA) (Fig. 3J) or recognize oxidized mC derivatives (EVE) (Fig. 3K) have been derived from hemimethylation and hmC-specific restriction systems, which are deployed by bacteria to counter bacteriophage hmC-containing DNA (Bickle and Kruger 1993; Iyer et al. 2013; Spruijt et al. 2013). 2OGFeDOs of the TET/JBP family (Fig. 3L), which oxidize methylpyrimidines in DNA, have been acquired from bacterial or phage sources on at least three independent occasions. In the case of the euglenozoan base J, phylogenetic analysis strongly suggests that both the thymine hydroxylase (JBP) and the subsequently acting glycosyltransferase have been acquired by transfer of a complete operon encoding both these enzymes from a phage similar to the *Persicivirga* phage P12024L (Iyer et al. 2013). Interestingly, TET/JBP domains are deployed as apparent host-directed effectors by the eukaryote endoparasite *Legionella* (Iyer et al. 2013). In a similar vein, the origin of the eukaryotic cytosine deaminating AID-APOBEC enzymes can be traced to the radiation of such enzymes among bacterial effectors deployed against eukaryotes (Iyer et al. 2011b).

EARLIEST EUKARYOTES POSSESSED A RICH ARRAY OF EPIGENETIC MODIFICATIONS

Epigenetic protein modifications are ubiquitous across major eukaryotic lineages and a subset of enzymes catalyzing modifications, such as methylation, acetylation, phosphorylation, and Ub/Ubl conjugation, is even present in minimal eukaryotic genomes such as microsporidians and *Entamoeba* (Iyer et al. 2008b). Thus, transmission of certain types of epigenetic information appears to be essential for eukaryotic life. Irrespective of uncertainties in the higher order phylogeny of eukaryotes (Simpson et al. 2006; Iyer et al. 2008b; Cavalier-Smith 2009), the LECA can be reconstructed as possessing a richly diversified apparatus for epigenetic tagging of proteins (Fig. 5). This includes multiple encoder enzymes for methylation, acetylation, phosphorylation, glycosylation, Ub/Ubl conjugation, peptide tagging, and ADP ribosylation (Manning et al. 2002; Iyer et al. 2008b; Citarelli et al. 2010). Likewise, at least one type of reader domain for most of these modifications can also be inferred in the ancestral eukaryote (Iyer et al. 2008b). However, in most cases there is less certainty about the presence of resetters in the LECA (e.g., strong evidence for the absence of demethylases) (Fig. 5). Thus, both chromatin-based and cytoskeletal epigenetic transmissions were already in place in the ancestral eukaryote and probably accompanied the initial expansion of low-complexity regions typical of eukaryotic proteins (Babu 2012; Cumberworth et al. 2013). By targeting such regions for modification, these encoder enzymes and reader domains considerably extended the "information content" and functional significance of conserved targets originally inherited from the archaeal progenitor (histones, tubulins, and ribosomal proteins) or innovated early in eukaryotic evolution (e.g., calnexins) (Fig. 5) (Dacks and Doolittle 2001; Sandman and Reeve 2005; Yutin and Koonin 2012). Thus, these ancient mediators of epigenetics tend to show a strong vertical pattern of inheritance alongside their conserved targets. However, subsequent evolution of the encoders is marked by numerous instances of duplications to form new conserved groups of paralogs, gene loss, and above all lineage-specific expansions (e.g., secreted SET domains in both diatoms and the rhizarian *Bigelowiella* and JOR/JmjC domains in the latter associated with spec-

ifying extracellular self-organizing structures). Continuing lateral transfers from bacteria, such as of the secreted FAM20 kinases, have been central to the emergence of specific eukaryotic adaptations, such as phosphate-rich biomineralized matrices (Fig. 5).

ANTECEDENTS OF EUKARYOTIC EPIGENETIC SYSTEMS IN BACTERIAL CONFLICT SYSTEMS

Interestingly, barring few apparent exceptions like palmitoylation, the vast majority of epigenetic systems from eukaryotes appear to share key components with systems involved in prokaryotic genomic and interorganismal conflicts (Fig. 4) (Aravind et al. 2011, 2012). The latter cover a disparate set of systems including (1) those deployed in conflicts between bacteria and their phages (e.g., restriction and restriction-modification systems and phage DNA hypermodification systems) (Makarova et al. 2012). A comparable set of effector domains, typified by those from toxin-antitoxin systems, are deployed in both intra- and intergenomic conflict systems (Leplae et al. 2011). (2) Secreted toxins or effectors deployed in intraspecific and interspecific interactions. These include polymorphic toxins used in interbacterial conflicts and effectors secreted by pathogens and endosymbionts to control/alter host cellular processes (Zhang et al. 2012). (3) Systems for synthesis and modification of secondary metabolites (Walsh 2003). Some of these (antibiotics) are directly used in conflicts with competitors, whereas others like siderophores and signaling molecules are subjects of conflicts owing to siderophore "stealing" and predatory interactions (Barry and Challis 2009). Further, several defense mechanisms against toxic secondary metabolites involve their enzymatic modification by acetylases, methylases, kinases, and other enzymatic domains also seen in epigenetic encoders (Ramirez and Tolmasky 2010).

Links between DNA-based defenses and countermeasures in bacteria-phage conflicts and DNA-modifying epigenetic systems are rather direct because modifiers and readers are used in a similar capacity (Fig. 4). In particular, phages have evolved a large array of unusual, modified DNA bases, which apparently serve both as epigenetic marks for headful packaging of their genomes into capsids and evasion of host restriction (Gommers-Ampt and Borst 1995; Lobocka et al. 2004; Iyer et al. 2013). Likewise, secreted toxins/effectors function just like their eukaryotic epigenetic counterparts in terms of the modifications they catalyze. Notably, bacterial endosymbionts/parasites of amoebozoa and metazoans, such as *Legionella*, *Amoebophilus*, *Protochlamydia*, and *Odyssella* secrete a large panoply of effectors that possess catalytic domains spanning most of the protein and DNA-modifying epigenetic systems deployed by eukaryotes (Fig. 4). Importantly, some of these endosymbionts like *Odyssella* are phylogenetically close to the alphaproteobacterial progenitor of the mitochondrion (Georgiades et al. 2011), suggesting that the latter might have been an important player in the acquisition of such domains by the stem eukaryote (Aravind et al. 2012). Prokaryotic metabolites are often synthesized via serial enzymatic modifications from amino acid or short peptide precursors (e.g., β-lactam antibiotics from tripeptides) (Walsh 2003). The short peptide substrates of these enzymes resemble peptide segments, especially those in free conformations from low-complexity tails of eukaryotic proteins (Iyer et al. 2009; Cumberworth et al. 2013). Thus, actions of eukaryotic epigenetic modifiers closely mimic modifications of precursors in secondary metabolite biosynthesis (Fig. 1).

These conflict systems are important foci for intense innovation of new activities because of constant selective pressures arising from resistance and the need for evasion in the case of siderophore stealing and predatory targeting of secreted signals. In contrast to the innovation-fostering positive selection that confronts the bacterial systems, the eukaryotic systems are often characterized by stronger conservation, indicative of an innovation-curtailing tendency for purifying selection (Iyer et al. 2008b). Thus, the bacterial systems can be conceptualized as crucibles for generation of a startling array of new activities that were then "import-

ed" by lateral transfer to eukaryotes. This provides a coherent explanation for the apparent spurt of innovation of new systems in early eukaryotes, which might be contrasted to the later evolutionary trends characterized by conservation or at most lineage-specific expansion of preexisting domains. A notable finding furnished by comparative genomics is that the above three apparently distinct categories of bacterial conflict systems in turn share several catalytic components between themselves (Fig. 4). For example, kinases, different types of methylases, acetylases, glycosyltransferases, and base deaminases, are used both in secondary metabolism systems and secreted toxins/effectors (Reinert et al. 2005; Iyer et al. 2011b; Zhang et al. 2012). Similarly, diverse endonucleases, TET/JBP proteins, and glycosyltransferases are shared between DNA restriction/modification systems and secreted toxins/effectors. This indicates that innovations occurring in one system can be channeled to another, thereby increasing the potential for newer catalytic inventions. Importantly, channeling of adaptations from secondary metabolism and DNA restriction/modification systems to effector systems delivered into eukaryotic hosts by endosymbionts could have provided the direct conduit for such adaptations right from the earliest endosymbiotic events in eukaryogenesis.

BIOGEOCHEMICAL CONSIDERATIONS

Diversification of secondary metabolism in bacteria is conspicuous for the high oxygen content and oxidation states of the emergent metabolites (Walsh 2003). This is a direct consequence of introduction of molecular oxygen into metabolites by enzymes; especially double-stranded β-helix dioxygenases and NAD/FAD-dependent Rossmann-fold dehydrogenases (Iyer et al. 2010). Two major superfamilies (JOR/JmjC and 2OGFeDO) of the former dioxygenases also use 2-oxoglutarate, a tricarboxylic acid (TCA) cycle metabolite as a cosubstrate. These observations indicate availability of free molecular oxygen and a functional TCA cycle (Iyer et al. 2010). Although hotly debated, prokaryotic life was perhaps present on Earth for at least ~3.5 billion years (Altermann and Kazmierczak 2003; Brasier et al. 2006); however, molecular oxygen became abundant only after the great oxygenation event (GOE) around 2.4 billion years ago (Frei et al. 2009), perhaps released by cyanobacterial photosynthesis (Flannery and Walter 2012). This suggests that the explosive expansion of secondary metabolism was a direct consequence of the GOE, probably occurring shortly after fixation of the TCA cycle. The TCA cycle, in addition to supplying metabolites, probably provided greater energetic capabilities to support lifestyles with an expanded secondary metabolism.

Importantly, diversification of these oxygen-utilizing enzymes gave rise to key components of epigenetic systems, such as the LSD1 and JOR/JmjC histone demethylases, TET/JBP methylpyrimidine 2OGFeDOs, prolyl, lysyl, and asparaginyl dioxygenases. This suggests that the emergence of oxidative epigenetic modifications of proteins and nucleic acids was coeval with or postdated the GOE (Iyer et al. 2010). Importantly, these modifications are central to resetting methyl marks in both proteins and nucleic acids (Klose et al. 2006; Iyer et al. 2010; Pastor et al. 2013); hence, their emergence might have allowed methyl modifications to be used dynamically, making them a major component of the epigenetic code in nuclear, cytoplasmic, and extracellular contexts. The GOE was also responsible for the genesis of more than half the known minerals on Earth (Hazen 2010), including phosphates that probably formed in shallow seas owing to oxidation by molecular oxygen (Papineau et al. 2012). Given the importance of phosphates in biomineralized matrices, it is likely that the GOE also facilitated the use of secreted kinases as a mechanism for generation of such.

CONCLUDING REMARKS

The origin of eukaryotes has long been regarded as a "major evolutionary transition" requiring explanations beyond the quotidian evolutionary processes shaping the forms and genomes of organisms under usual conditions (Maynard Smith and Szathmáry 1999). Teleologies, such

as relaxation of selective constraints with concomitant paralog formation and accumulation of low-complexity segments in proteins, have been proposed as general explanations of this event (Lynch 2007). However, until recently there has been little proximal understanding of the actual steps leading to it, in particular, the numerous new and parallel inventions that occurred independently of paralog formation and rapid divergence in the stem eukaryotes. One key set of these innovations (i.e., epigenetic modifications), which are closely tied to the origin of the nucleus can now be seen in large part as emerging from a rich pool of biochemistries, and which first evolved in the context of bacterial conflict systems—in a sense "peacetime" use of "war-time" inventions (Aravind et al. 2012). The explosive spurt of innovation seen in these systems also displays imprints of a major geochemical event that shaped the Earth, the GOE. Thus, in the early endosymbiotic associations resulting in eukaryogenesis bacterial endosymbionts were not passive partners in a metabolic mutualism (Lopez-Garcia and Moreira 1999) but active manipulators of the archaeal host's biology, much like extant endosymbionts. This very manipulation, perhaps in more than one way, favored the emergence of a protective barrier for the host genome, the nucleus (Koonin 2006; Jekely 2008; Aravind et al. 2012), while enabling "domestication" of many of these manipulation strategies as purveyors of epigenetic information.

ACKNOWLEDGMENTS

Work by the authors is supported by the intramural funds of the National Library of Medicine, National Institutes of Health.

REFERENCES

Allis CD, Jenuwein T, Reinberg D. 2007. *Epigenetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Altermann W, Kazmierczak J. 2003. Archean microfossils: A reappraisal of early life on Earth. *Res Microbiol* **154:** 611–617.

Ame JC, Spenlehauer C, de Murcia G. 2004. The PARP superfamily. *Bioessays* **26:** 882–893.

Anne J, Ollo R, Ephrussi A, Mechler BM. 2007. Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in *Drosophila*. *Development* **134:** 137–146.

Aravind L, Iyer LM, Koonin EV. 2006. Comparative genomics and structural biology of the molecular innovations of eukaryotes. *Curr Opin Struct Biol* **16:** 409–419.

Aravind L, Anantharaman V, Iyer LM. 2010. *Sensory mechanisms in bacteria: Molecular aspects of signal recognition*. Caister Academic, Norfolk, UK.

Aravind L, Abhiman S, Iyer LM. 2011. Natural history of the eukaryotic chromatin protein methylation system. *Prog Mol Biol Transl Sci* **101:** 105–176.

Aravind L, Anantharaman V, Zhang D, de Souza RF, Iyer LM. 2012. Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Front Cell Infect Microbiol* **2:** 89.

Avalos JL, Boeke JD, Wolberger C. 2004. Structural basis for the mechanism and regulation of Sir2 enzymes. *Mol Cell* **13:** 639–648.

Babu MM. 2012. Intrinsically disordered proteins. *Mol Biosyst* **8:** 21.

Barry SM, Challis GL. 2009. Recent advances in siderophore biosynthesis. *Curr Opin Chem Biol* **13:** 205–215.

Beauregard PB, Guerin R, Turcotte C, Lindquist S, Rokeach LA. 2009. A nucleolar protein allows viability in the absence of the essential ER-residing molecular chaperone calnexin. *J Cell Sci* **122:** 1342–1351.

Best AA, Morrison HG, McArthur AG, Sogin ML, Olsen GJ. 2004. Evolution of eukaryotic transcription: Insights from the genome of *Giardia lamblia*. *Genome Res* **14:** 1537–1547.

Bickle TA, Kruger DH. 1993. Biology of DNA restriction. *Microbiol Rev* **57:** 434–450.

Blander G, Guarente L. 2004. The Sir2 family of protein deacetylases. *Annu Rev Biochem* **73:** 417–435.

Blaskovic S, Blanc M, van der Goot FG. 2013. What does *S*-palmitoylation do to membrane proteins? *FEBS J* **280:** 2766–2774.

Borst P, Sabatini R. 2008. Base J: Discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* **62:** 235–251.

Brasier M, McLoughlin N, Green O, Wacey D. 2006. A fresh look at the fossil evidence for early Archaean cellular life. *Philos Trans R Soc Lond B Biol Sci* **361:** 887–902.

Brehm A, Tufteland KR, Aasland R, Becker PB. 2004. The many colours of chromodomains. *Bioessays* **26:** 133–140.

Brzeski J, Brzeska K. 2011. The maze of paramutation: A rough guide to the puzzling epigenetics of paramutation. *Wiley Interdiscip Rev RNA* **2:** 863–874.

Burroughs AM, Iyer LM, Aravind L. 2011. Functional diversification of the RING finger and other binuclear treble clef domains in prokaryotes and the early evolution of the ubiquitin system. *Mol Biosyst* **7:** 2261–2277.

Burroughs AM, Iyer LM, Aravind L. 2012. Structure and evolution of ubiquitin and ubiquitin-related domains. *Methods Mol Biol* **832:** 15–63.

Campanacci V, Mukherjee S, Roy CR, Cherfils J. 2013. Structure of the Legionella effector AnkX reveals the mechanism of phosphocholine transfer by the FIC domain. *EMBO J* **32:** 1469–1477.

Cavalier-Smith T. 2009. Megaphylogeny, cell body plans, adaptive zones: Causes and timing of eukaryote basal radiations. *J Eukaryot Microbiol* **56:** 26–33.

Cedar H, Bergman Y. 2011. Epigenetics of haematopoietic cell development. *Nat Rev Immunol* **11:** 478–488.

Chakravarty S, Zeng L, Zhou MM. 2009. Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure* **17:** 670–679.

Chen Y, Yang Y, Wang F, Wan K, Yamane K, Zhang Y, Lei M. 2006. Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc Natl Acad Sci* **103:** 13956–13961.

Citarelli M, Teotia S, Lamb RS. 2010. Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. *BMC Evol Biol* **10:** 308.

Collins RE, Northrop JP, Horton JR, Lee DY, Zhang X, Stallcup MR, Cheng X. 2008. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat Struct Mol Biol* **15:** 245–250.

Conaway RC, Conaway JW. 2011. Function and regulation of the Mediator complex. *Curr Opin Genet Dev* **21:** 225–230.

Cook PJ, Ju BG, Telese F, Wang X, Glass CK, Rosenfeld MG. 2009. Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. *Nature* **458:** 591–596.

Cumberworth A, Lamour G, Babu MM, Gsponer J. 2013. Promiscuity as a functional trait: Intrinsically disordered regions as central players of interactomes. *Biochem J* **454:** 361–369.

Dacks JB, Doolittle WF. 2001. Reconstructing/deconstructing the earliest eukaryotes: How comparative genomics can help. *Cell* **107:** 419–425.

Dann CE 3rd, Bruick RK, Deisenhofer J. 2002. Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci* **99:** 15351–15356.

Denhardt DT, Chaly N, Walden DB. 2005. The eukaryotic nucleus: A thematic issue. *BioEssays* **9:** 43.

de Souza RF, Aravind L. 2012. Identification of novel components of NAD-utilizing metabolic pathways and prediction of their biochemical functions. *Mol Biosyst* **8:** 1661–1677.

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

Dlakic M. 2001. Chromatin silencing protein and pachytene checkpoint regulator Dot1p has a methyltransferase fold. *Trends Biochem Sci* **26:** 405–407.

Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, Scherf A, Crabb BS, Cowman AF. 2005. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* **121:** 13–24.

Durant M, Pugh BF. 2006. Genome-wide relationships between TAF1 and histone acetyltransferases in *Saccharomyces cerevisiae*. *Mol Cell Biol* **26:** 2791–2802.

Dutnall RN. 2003. Cracking the histone code: One, two, three methyls, you're out! *Mol Cell* **12:** 3–4.

Dutnall RN, Tafrov ST, Sternglanz R, Ramakrishnan V. 1998. Structure of the yeast histone acetyltransferase Hat1: Insights into substrate specificity and implications for the Gcn5-related *N*-acetyltransferase superfamily. *Cold Spring Harb Symp Quant Biol* **63:** 501–507.

Egly JM, Coin F. 2011. A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor. *DNA Repair (Amst)* **10:** 714–721.

Elkins JM, Hewitson KS, McNeill LA, Seibel JF, Schlemminger I, Pugh CW, Ratcliffe PJ, Schofield CJ. 2003. Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1α. *J Biol Chem* **278:** 1802–1806.

Engel P, Goepfert A, Stanger FV, Harms A, Schmidt A, Schirmer T, Dehio C. 2012. Adenylylation control by intra- or intermolecular active-site obstruction in Fic proteins. *Nature* **482:** 107–110.

Feng F, Yang F, Rong W, Wu X, Zhang J, Chen S, He C, Zhou JM. 2012. A Xanthomonas uridine 5′-monophosphate transferase inhibits plant immune kinases. *Nature* **485:** 114–118.

Flannery DT, Walter RM. 2012. Archean tufted microbial mats and the Great Oxidation Event: New insights into an ancient problem. *Aust J Earth Sci* **59:** 1–11.

Forouhar F, Lee IS, Vujcic J, Vujcic S, Shen J, Vorobiev SM, Xiao R, Acton TB, Montelione GT, Porter CW, et al. 2005. Structural and functional evidence for *Bacillus subtilis* PaiA as a novel N1-spermidine/spermine acetyltransferase. *J Biol Chem* **280:** 40328–40336.

Frei R, Gaucher C, Poulton SW, Canfield DE. 2009. Fluctuations in Precambrian atmospheric oxygenation recorded by chromium isotopes. *Nature* **461:** 250–253.

Fukushima N, Furuta D, Hidaka Y, Moriyama R, Tsujiuchi T. 2009. Post-translational modifications of tubulin in the nervous system. *J Neurochem* **109:** 683–693.

Garcia-Alai MM, Allen MD, Joerger AC, Bycroft M. 2010. The structure of the FYR domain of transforming growth factor β regulator 1. *Protein Sci* **19:** 1432–1438.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL. 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124:** 495–506.

Georgiades K, Madoui MA, Le P, Robert C, Raoult D. 2011. Phylogenomic analysis of *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclimonas americana* mitochondrion. *PLoS ONE* **6:** e24857.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74:** 481–514.

Gommers-Ampt JH, Borst P. 1995. Hypermodified bases in DNA. *FASEB J* **9:** 1034–1042.

Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA. 2010. Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol* **8:** e1000339.

Grazini U, Zanardi F, Citterio E, Casola S, Goding CR, McBlane F. 2010. The RING domain of RAG1 ubiquitylates histone H3: A novel activity in chromatin-mediated regulation of V(D)J joining. *Mol Cell* **37:** 282–293.

Grewal SI. 2010. RNAi-dependent formation of heterochromatin and its diverse functions. *Curr Opin Genet Dev* **20:** 134–141.

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Griffiths DS, Li J, Dawson MA, Trotter MW, Cheng YH, Smith AM, Mansfield W, Liu P, Kouzarides T, Nichols J, et al. 2011. LIF-independent JAK signalling to chromatin in embryonic stem cells uncovered from an adult stem cell disease. *Nat Cell Biol* **13:** 13–21.

Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA. 2010. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329:** 78–82.

Hanover JA, Krause MW, Love DC. 2012. Bittersweet memories: Linking metabolism to epigenetics through *O*-GlcNAcylation. *Nat Rev Mol Cell Biol* **13:** 312–321.

Hauk G, Bowman GD. 2011. Structural insights into regulation and action of SWI2/SNF2 ATPases. *Curr Opin Struct Biol* **21:** 719–727.

Hazen RM. 2010. Evolution of minerals. *Sci Am* **302:** 58–65.

Heun P. 2007. SUMOrganization of the nucleus. *Curr Opin Cell Biol* **19:** 350–355.

Higgins JM. 2010. Haspin: A newly discovered regulator of mitotic chromosome behavior. *Chromosoma* **119:** 137–147.

Hochstrasser M. 2009. Origin and function of ubiquitin-like proteins. *Nature* **458:** 422–429.

Iyer LM, Abhiman S, Aravind L. 2008a. MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol Direct* **3:** 8.

Iyer LM, Anantharaman V, Wolf MY, Aravind L. 2008b. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* **38:** 1–31.

Iyer LM, Abhiman S, Maxwell Burroughs A, Aravind L. 2009. Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: Synthesis of novel metabolites and peptide modifications of proteins. *Mol Biosyst* **5:** 1636–1660.

Iyer LM, Abhiman S, de Souza RF, Aravind L. 2010. Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. *Nucleic Acids Res* **38:** 5261–5279.

Iyer LM, Abhiman S, Aravind L. 2011a. Natural history of eukaryotic DNA methylation systems. *Prog Mol Biol Transl Sci* **101:** 25–104.

Iyer LM, Zhang D, Rogozin IB, Aravind L. 2011b. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res* **39:** 9473–9497.

Iyer LM, Zhang D, Maxwell Burroughs A, Aravind L. 2013. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res* **41:** 7635–7655.

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

Janke C, Rogowski K, van Dijk J. 2008. Polyglutamylation: A fine-regulator of protein function? "Protein modifications: Beyond the usual suspects" review series. *EMBO Rep* **9:** 636–641.

Jekely G. 2008. Origin of the nucleus and Ran-dependent transport to safeguard ribosome biogenesis in a chimeric cell. *Biol Direct* **3:** 31.

Jeyaprakash AA, Basquin C, Jayachandran U, Conti E. 2011. Structural basis for the recognition of phosphorylated histone h3 by the survivin subunit of the chromosomal passenger complex. *Structure* **19:** 1625–1634.

Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, Yang W, Turner L, Lavstsen T, Theander TG, et al. 2013. PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature* **499:** 223–227.

Johnson SA, Hunter T. 2005. Kinomics: Methods for deciphering the kinome. *Nat Methods* **2:** 17–25.

Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G. 2007. Structural and functional diversity of the microbial kinome. *PLoS Biol* **5:** e17.

Klose RJ, Kallin EM, Zhang Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7:** 715–727.

Koonin EV. 2006. The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate? *Biol Direct* **1:** 22.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128:** 693–705.

Krishnan N, Jeong DG, Jung SK, Ryu SE, Xiao A, Allis CD, Kim SJ, Tonks NK. 2009. Dephosphorylation of the C-terminal tyrosyl residue of the DNA damage-related histone H2A.X is mediated by the protein phosphatase eyes absent. *J Biol Chem* **284:** 16066–16070.

Kroger N, Lorenz S, Brunner E, Sumper M. 2002. Self-assembly of highly phosphorylated silaffins and their function in biosilica morphogenesis. *Science* **298:** 584–586.

Laing S, Unger M, Koch-Nolte F, Haag F. 2011. ADP-ribosylation of arginine. *Amino Acids* **41:** 257–269.

Lallemand-Breitenbach V, de The H. 2010. PML nuclear bodies. *Cold Spring Harb Perspect Biol* **2:** a000661.

Landry CD, Kandel ER, Rajasethupathy P. 2013. New mechanisms in memory storage: piRNAs and epigenetics. *Trends Neurosci* **36:** 535–542.

Lanzuolo C, Orlando V. 2012. Memories from the polycomb group proteins. *Annu Rev Genet* **46:** 561–589.

Lee YH, Stallcup MR. 2009. Minireview: Protein arginine methylation of nonhistone proteins in transcriptional regulation. *Mol Endocrinol* **23:** 425–433.

Leipe DD, Landsman D. 1997. Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily. *Nucleic Acids Res* **25:** 3693–3697.

Leonard TA, Butler PJ, Lowe J. 2004. Structural analysis of the chromosome segregation protein Spo0J from *Thermus thermophilus*. *Mol Microbiol* **53:** 419–432.

Leplae R, Geeraerts D, Hallez R, Guglielmini J, Dreze P, Van Melderen L. 2011. Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res* **39:** 5513–5525.

Linder ME, Jennings BC. 2013. Mechanism and function of DHHC *S*-acyltransferases. *Biochem Soc Trans* **41:** 29–34.

Liu X, Wang L, Zhao K, Thompson PR, Hwang Y, Marmorstein R, Cole PA. 2008. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* **451:** 846–850.

Lloyd J, Chapman JR, Clapperton JA, Haire LF, Hartsuiker E, Li J, Carr AM, Jackson SP, Smerdon SJ. 2009. A supra-

modular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA damage. *Cell* **139:** 100–111.

Lobocka MB, Rose DJ, Plunkett G 3rd, Rusin M, Samojedny A, Lehnherr H, Yarmolinsky MB, Blattner FR. 2004. Genome of bacteriophage P1. *J Bacteriol* **186:** 7032–7068.

Lopez-Garcia P, Moreira D. 1999. Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem Sci* **24:** 88–93.

Lynch M. 2007. *The origins of genome architecture*. Sinauer Associates, Sunderland, MA.

Makarova KS, Anantharaman V, Aravind L, Koonin EV. 2012. Live virus-free or die: Coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biol Direct* **7:** 40.

Manning G, Plowman GD, Hunter T, Sudarsanam S. 2002. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* **27:** 514–520.

Mans BJ, Anantharaman V, Aravind L, Koonin EV. 2004. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* **3:** 1612–1637.

Manzur KL, Farooq A, Zeng L, Plotnikova O, Koch AW, Sachchidanand, Zhou MM. 2003. A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nat Struct Biol* **10:** 187–196.

Martinez-Campa C, Politis P, Moreau JL, Kent N, Goodall J, Mellor J, Goding CR. 2004. Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1. *Mol Cell* **15:** 69–81.

Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. 2003. The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28:** 69–74.

Maynard Smith J, Szathmáry E. 1999. *The origins of life: From the birth of life to the origin of language*. Oxford University Press, Oxford, New York.

Miranda TB, Webb KJ, Edberg DD, Reeves R, Clarke S. 2005. Protein arginine methyltransferase 6 specifically methylates the nonhistone chromatin protein HMGA1a. *Biochem Biophys Res Commun* **336:** 831–835.

Mochizuki K. 2010. DNA rearrangements directed by noncoding RNAs in ciliates. *Wiley Interdiscip Rev RNA* **1:** 376–387.

Montellier E, Rousseaux S, Zhao Y, Khochbin S. 2012. Histone crotonylation specifically marks the haploid male germ cell gene expression program: Post-meiotic male-specific gene expression. *Bioessays* **34:** 187–193.

Mueller GA, Moon AF, Derose EF, Havener JM, Ramsden DA, Pedersen LC, London RE. 2008. A comparison of BRCT domains involved in nonhomologous end-joining: Introducing the solution structure of the BRCT domain of polymerase λ. *DNA Repair (Amst)* **7:** 1340–1351.

Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. 2002. Histone methyltransferase activity of a *Drosophila* polycomb group repressor complex. *Cell* **111:** 197–208.

Neuwald AF, Landsman D. 1997. GCN5-related histone *N*-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem Sci* **22:** 154–155.

Nicholson TB, Chen T. 2009. LSD1 demethylates histone and non-histone proteins. *Epigenetics* **4:** 129–132.

Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee GJ, Hattori M, Kanai A, Atomi H, et al. 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39:** 3204–3223.

Oeffinger M, Wei KE, Rogers R, DeGrasse JA, Chait BT, Aitchison JD, Rout MP. 2007. Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat Methods* **4:** 951–956.

Ossareh-Nazari B, Bonizec M, Cohen M, Dokudovskaya S, Delalande F, Schaeffer C, Van Dorsselaer A, Dargemont C. 2010. Cdc48 and Ufd3, new partners of the ubiquitin protease Ubp3, are required for ribophagy. *EMBO Rep* **11:** 548–554.

Papineau D, Purohit R, Fogel ML, Shields-Zhou GA. 2012. High phosphate availability as a possible cause for massive cyanobacterial production of oxygen in the Paleoproterozoic atmosphere. *Earth Planet Sci Lett* **362:** 225–236.

Pastor WA, Aravind L, Rao A. 2013. TETonic shift: Biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* **14:** 341–356.

Peterson CL, Laniel MA. 2004. Histones and histone modifications. *Curr Biol* **14:** R546–R551.

Porras-Yakushi TR, Whitelegge JP, Clarke S. 2007. Yeast ribosomal/cytochrome *c* SET domain methyltransferase subfamily: Identification of Rpl23ab methylation sites and recognition motifs. *J Biol Chem* **282:** 12368–12376.

Ramirez MS, Tolmasky ME. 2010. Aminoglycoside modifying enzymes. *Drug Resist Updat* **13:** 151–171.

Reinert DJ, Jank T, Aktories K, Schulz GE. 2005. Structural basis for the function of *Clostridium difficile* toxin B. *J Mol Biol* **351:** 973–981.

Resh MD. 2006. Palmitoylation of ligands, receptors, and intracellular signaling molecules. *Sci STKE* **2006:** re14.

Richards EJ, Elgin SC. 2002. Epigenetic codes for heterochromatin formation and silencing: Rounding up the usual suspects. *Cell* **108:** 489–500.

Roberts RJ, Vincze T, Posfai J, Macelis D. 2010. REBASE—A database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res* **38:** D234–D236.

Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, Aravind L, Pancer Z. 2007. Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol* **8:** 647–656.

Rossetto D, Avvakumov N, Cote J. 2012. Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics* **7:** 1098–1108.

Sandman K, Reeve JN. 2005. Archaeal chromatin proteins: Different structures but common function? *Curr Opin Microbiol* **8:** 656–661.

Sawada K, Yang Z, Horton JR, Collins RE, Zhang X, Cheng X. 2004. Structure of the conserved core of the yeast Dot1p, a nucleosomal histone H3 lysine 79 methyltransferase. *J Biol Chem* **279:** 43296–43306.

Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M. 2010. The genome of the amoeba

symbiont "*Candidatus* Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* **192:** 1045–1057.

Scott RJ, Spielman M. 2006. Genomic imprinting in plants and mammals: How life history constrains convergence. *Cytogenet Genome Res* **113:** 53–67.

Sheppard V, Poulsen N, Kroger N. 2010. Characterization of an endoplasmic reticulum-associated silaffin kinase from the diatom *Thalassiosira pseudonana*. *J Biol Chem* **285:** 1166–1176.

Shimojo H, Sano N, Moriwaki Y, Okuda M, Horikoshi M, Nishimura Y. 2008. Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. *J Mol Biol* **378:** 987–1001.

Simpson AG, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol Biol Evol* **23:** 615–625.

Singh N, Basnet H, Wiltshire TD, Mohammad DH, Thompson JR, Heroux A, Botuyan MV, Yaffe MB, Couch FJ, Rosenfeld MG, et al. 2012. Dual recognition of phosphoserine and phosphotyrosine in histone variant H2A.X by DNA damage response protein MCPH1. *Proc Natl Acad Sci* **109:** 14381–14386.

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F, et al. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152:** 1146–1159.

Sumper M, Hett R, Lehmann G, Wenzl S. 2007. A code for lysine modifications of a silica biomineralizing silaffin protein. *Angew Chem Int Ed Engl* **46:** 8405–8408.

Tagliabracci VS, Pinna LA, Dixon JE. 2013. Secreted protein kinases. *Trends Biochem Sci* **38:** 121–130.

Touz MC, Ropolo AS, Rivero MR, Vranych CV, Conrad JT, Svard SG, Nash TE. 2008. Arginine deiminase has multiple regulatory roles in the biology of *Giardia lamblia*. *J Cell Sci* **121:** 2930–2938.

Trievel RC, Shilatifard A. 2009. WDR5, a complexed protein. *Nat Struct Mol Biol* **16:** 678–680.

Trievel RC, Beach BM, Dirk LM, Houtz RL, Hurley JH. 2002. Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111:** 91–103.

Tsuge K, Matsui K, Itaya M. 2007. Production of the nonribosomal peptide plipastatin in *Bacillus subtilis* regulated by three relevant gene blocks assembled in a single movable DNA segment. *J Biotechnol* **129:** 592–603.

Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439:** 811–816.

Venancio TM, Balaji S, Iyer LM, Aravind L. 2009. Reconstructing the ubiquitin network: Cross-talk with other systems and identification of novel functions. *Genome Biol* **10:** R33.

Vossenaar ER, Zendman AJ, van Venrooij WJ, Pruijn GJ. 2003. PAD, a growing family of citrullinating enzymes: Genes, features and involvement in disease. *Bioessays* **25:** 1106–1118.

Walsh C. 2003. *Antibiotics: Actions, origins, resistance*. American Society for Microbiology, Washington, DC.

Walsh DA, Doolittle WF. 2005. The real "domains" of life. *Curr Biol* **15:** R237–R240.

Walsh CT, Garneau-Tsodikova S, Gatto GJ. 2005. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angewandte Chemie* **2005:** 7342–7372.

West AH, Stock AM. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26:** 369–376.

Winkler GS, Kristjuhan A, Erdjument-Bromage H, Tempst P, Svejstrup JQ. 2002. Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. *Proc Natl Acad Sci* **99:** 3517–3522.

Wittschieben BO, Otero G, de Bizemont T, Fellows J, Erdjument-Bromage H, Ohba R, Li Y, Allis CD, Tempst P, Svejstrup JQ. 1999. A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol Cell* **4:** 123–128.

Worby CA, Mattoo S, Kruger RP, Corbeil LB, Koller A, Mendez JC, Zekarias B, Lazar C, Dixon JE. 2009. The fic domain: Regulation of cell signaling by adenylylation. *Mol Cell* **34:** 93–103.

Wright DE, Wang CY, Kao CF. 2012. Histone ubiquitylation and chromatin dynamics. *Front Biosci (Landmark Ed)* **17:** 1051–1078.

Xiao A, Li H, Shechter D, Ahn SH, Fabrizio LA, Erdjument-Bromage H, Ishibe-Murakami S, Wang B, Tempst P, Hofmann K, et al. 2009. WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. *Nature* **457:** 57–62.

Yan Q, Dutt S, Xu R, Graves K, Juszczynski P, Manis JP, Shipp MA. 2009. BBAP monoubiquitylates histone H4 at lysine 91 and selectively modulates the DNA damage response. *Mol Cell* **36:** 110–120.

Yap KL, Zhou MM. 2010. Keeping it in the family: Diverse histone recognition by conserved structural folds. *Crit Rev Biochem Mol Biol* **45:** 488–505.

Yutin N, Koonin EV. 2012. Archaeal origin of tubulin. *Biol Direct* **7:** 10.

Zeng L, Zhou MM. 2002. Bromodomain: An acetyl-lysine binding domain. *FEBS Lett* **513:** 124–128.

Zentner GE, Henikoff S. 2013. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20:** 259–266.

Zhang D, Aravind L. 2012. Novel transglutaminase-like peptidase and C2 domains elucidate the structure, biogenesis and evolution of the ciliary compartment. *Cell Cycle* **11:** 3861–3875.

Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. 2012. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct* **7:** 18.

Zippo A, Serafini R, Rocchigiani M, Pennacchini S, Krepelova A, Oliviero S. 2009. Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* **138:** 1122–1136.

Zoldos V, Novokmet M, Beceheli I, Lauc G. 2013. Genomics and epigenomics of the human glycome. *Glycoconj J* **30:** 41–50.

Protein and DNA Modifications: Evolutionary Imprints of Bacterial Biochemical Diversification and Geochemistry on the Provenance of Eukaryotic Epigenetics

L. Aravind, A. Maxwell Burroughs, Dapeng Zhang and Lakshminarayan M. Iyer

Cold Spring Harb Perspect Biol 2014; doi: 10.1101/cshperspect.a016063

For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/