Review

Diversity and evolution of chromatin proteins encoded by DNA viruses

Robson F. de Souza, Lakshminarayan M. Iyer, L. Aravind *

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

ABSTRACT

Double-stranded DNA viruses display a great variety of proteins that interact with host chromatin. Using the wealth of available genomic and functional information, we have systematically surveyed chromatin-related proteins encoded by dsDNA viruses. The distribution of viral chromatin-related proteins is primarily influenced by viral genome size and the superkingdom to which the host of the virus belongs. Smaller viruses usually encode multifunctional proteins that mediate several distinct interactions with host chromatin proteins and viral or host DNA. Larger viruses additionally encode several enzymes, which catalyze manipulations of chromosome structure, chromatin remodeling and covalent modifications of proteins and DNA. Among these viruses, it is also common to encounter transcription factors and DNA-packaging proteins such as histones and IHF/HU derived from cellular genomes, which might play a role in constituting virus-specific chromatin states. Through all size ranges a subset of domains in viral chromatin proteins appears to have been derived from those found in host proteins. Examples include the Zn-finger domains of the E6 and E7 proteins of papillomaviruses, SET domain methyltransferases and Jumonji-related demethylases in certain nucleocytoplasmic large DNA viruses and BEN domains in poxviruses and polydnaviruses. In other cases, chromatin-interacting modules, such as the LXCXE motif, appear to have been widely disseminated across distinct viral lineages, resulting in similar retinoblastoma targeting strategies. Viruses, especially those with large linear genomes, have evolved a number of mechanisms to manipulate viral chromosomes in the process of replication-associated recombination. These include topoisomerases, Rad50/SbcC-like ABC ATPases and a novel recombinase system in bacteriophages utilizing RecA and Rad52 homologs. Larger DNA viruses also encode SWI2/SNF2 and A18-like ATPases which appear to play specialized roles in transcription and recombination. Finally, it also appears that certain domains of viral provenance have given rise to key functions in eukaryotic chromatin such as a HEH domain of chromosome tethering proteins and the TET/JBP-like cytosine and thymine hydroxylases.

Published by Elsevier B.V.

1. Introduction

In cellular life forms DNA-packaging proteins bind DNA with low sequence specificity, promote its bending and organize it into highly compacted structures. This nucleoprotein ensemble or chromatin has a central role in facilitating and regulating biochemical processes including DNA replication and repair, transcription and RNA processing. Evolutionary comparisons have shown that the primary DNA-packaging proteins involved in organization of chromatin are different across the three superkingdoms of life. In bacteria the primary DNA-packaging proteins are members of the HU/IHF (also called DNABII) superfamily [1]. In contrast, several archaea and most eukaryotes contain histones, which form the characteristic octameric DNA compaction unit termed the nucleosome [2]. However, in some eukaryotes, such as certain dinoflagellates, bacterial type HU/IHF homologs, rather than histones, play a fundamental role in DNA

packaging [3]. Likewise, in certain archaeal lineages such as Sulfolobales the histones appear to have been displaced by other chromosome-packaging proteins [4]. Importantly, eukaryotic histones differ from archaeal histones in having long, low-complexity tails that are enriched in positively charged residues and contact the negatively charged backbone of DNA [5]. These histone tails are substrates for a large number of chromatin-modifying enzymes, which catalyze a bewildering array of covalent modifications on lysine, arginine, serine, threonine and glutamate [6,7]. These modifications range from low molecular weight adducts such as methyl, acetyl and phosphate groups to ligation of entire protein chains such as ubiquitin and SUMO. Akin to protein modifications, DNA modifications such as methylation, momylation and more recently hydroxymethylation, among others, are seen to play important roles in chromatin organization [8–10].

Modifications of histones (and other chromosomal proteins) and DNA appear to act as a "code" atop that specified by the genome and are thus termed epigenetic marks [11]. Eukaryotes also display a unique proliferation of diverse "adaptor" domains, for example, the Bromo, Chromo, PHD, MYB/SANT and BMB (PWWP) domains [6].

* Corresponding author. Tel.: +1 301 594 2445.
E-mail address: aravind@mail.nih.gov (L. Aravind).

These domains recognize modified or unmodified peptides in histone tails and other chromatin proteins. Likewise, eukaryotes are also known to possess DNA-binding proteins that specifically recognize modified DNA [12]. Thus, domains which specifically recognize such covalent modifications help in "reading" the epigenetic code and linking it to various downstream processes [11]. Supercoiling, topology and higher order arrangement of DNA in chromatin is also highly dynamic and considerably influenced by the action of multiple distinct topoisomerases [13]. Eukaryotes in particular, and to a certain degree prokaryotes, also contain other chromatin remodeling enzymes that use the free-energy of ATP hydrolysis to actively remodel DNA–protein contacts, unwind DNA or reorganize it into higher order loop-structures. Such enzymes, including SWI2/SNF2 ATPases, SMC ATPases and MORC-type ATPases, have a major role in chromosomal organization and alterations of nucleosomal positions across eukaryotes [14–16]. Proteins involved in these structural and dynamic processes of chromatin interact with other DNA-binding proteins, namely, basal or general transcription factors (which recruit the RNA polymerase to a promoter) and specific transcription factors, which recognize distinctive regulatory DNA sequences associated with particular genes [17]. Transcription factors (TFs) often share DNA-binding domains with proteins involved in chromatin structure and dynamics and functionally overlap with them [6]. Thus, transcription-related protein complexes might also be considered integral components of chromatin in both eukaryotes and prokaryotes. While intimately interacting with the transcription regulatory apparatus, chromatin structure and dynamics provide a distinct level of regulation with major consequences for all the cellular processes that operate on DNA [18]. This regulatory level, especially in the form of epigenetic marks, is highly developed in eukaryotes [7,18,19] and to lesser degree in the two prokaryotic superkingdoms [15].

In contrast to cellular life forms, DNA viruses package their genome into externally situated protein coats (capsids) or lipid membranes situated inside such protein coats. Studies of different bacteriophages, such as lambda, P22 and T4, suggest that DNA is packaged in viral capsids as naked DNA close to the maximum possible density observed in a pure DNA crystal [20–22]. In contrast, cores of large eukaryotic poxviruses have much greater available space than in the bacteriophage capsids and DNA is packaged at lower density [23,24]. However, even in this case the bulk of DNA in the core appears to be primarily in the form of naked strands although there might be limited linkages to proteins [23]. A similar partial linkage to a protein (conserved protein VII) has been reported in adenoviral capsids [25]. Studies on T4 DNA packaging have shown that, though positively charged proteins of the capsid play some role in the process, majority of the charge-neutralization during viral DNA packaging comes from polyamines and monovalent metal ions included in the capsid [22]. Hence, viral DNA in capsids is packaged very differently from that of their cellular hosts. However, viral DNA, while replicating either as an episome or integrated into host DNA, is often subject to packaging similar to host chromatin.

In recent years, major advances in viral genomics have made available complete genome sequences of numerous large DNA viruses. Comparative viral genomics has gone a long way in revealing the nature of the viral proteome and previously unclear vertical and horizontal relationships between diverse dsDNA viruses [26,27]. These studies point to a complex web of relationships in which a variety of proteins are shared between otherwise phylogenetically distinct groups of viruses as a result of extensive lateral gene exchanges between viruses and their hosts. In the past, sequences of viral proteins have been difficult to analyze due to rapid divergence relative to one and other and their cellular counterparts. Availability of numerous genome sequences and structure solution efforts has mitigated this to a certain extent and allowed recovery of distant relationships [28–31]. These studies have shown that both eukaryotic and prokaryotic viruses encode a diverse set of chromatin proteins, each of which might have functional consequences for the host or the virus. To date studies on both eukaryotic and bacterial dsDNA viruses have revealed that they encode proteins that are involved in chromatin structure and dynamics [26,32–34]. These included various P-loop ATPases that could function as chromatin remodelers, topoisomerases, histone-modifying enzymes and DNA-binding proteins with packaging and structure-modifying potential. Experimental studies on some such virally encoded chromatin proteins have demonstrated critical roles for them in expression of host or viral genes [32–36].

In this article we attempt to systematically review virally encoded chromatin proteins from a comparative genomics perspective. In doing so we hope to bring attention to previously underappreciated viral chromatin proteins and place what is already known in a broader context. As can be seen from the above discussion, the category "chromatin proteins (CPs)" can be a bit diffuse, overlapping with other processes such as replication, recombination and transcription. In this article we stick mainly to those involved in chromatin structure and dynamics, largely refraining from detailed discussion on enzymes catalyzing DNA and RNA synthesis or mediating DNA repair. However, we do briefly consider several transcription factors and their DNA-binding domains due to their functional overlap with chromatin proteins. We begin by providing an overview of large dsDNA viral relationships and phyletic patterns of chromatin proteins encoded by them. We follow this with a summary of the various functional classes of chromatin proteins encoded by viruses and their potential significance for viral biology. Finally, we attempt to integrate this information into our current understanding of viral evolution.

2. A general survey of DNA viruses and phyletic patterns of viral chromatin proteins

2.1. Commonalities and differences of DNA viruses

Double-stranded DNA viruses are enormously diverse in terms of virion morphology, genome size/coding capacity, genome structure and replication strategies (Fig. 1). Yet several disparate groups of viruses might display one or more shared features that include [26,28,30,37] (1) β-jellyroll domain capsid proteins, (2) DNA-packaging ATPases either of the terminase large subunit or FtsK-HerA superfamily, (3) portal proteins, (4) DNA polymerases, (5) replication-related DNA helicases belonging to the AAA+ superfamily, and (6) primases either of the eukaryote-type primase superfamily or TOPRIM domain superfamily (DnaG-like). Most large DNA viruses additionally encode one or more DNA metabolism enzymes that might help in more efficiently providing precursors for DNA synthesis [26]. These features appear to have spread in viruses as a result of a combination of common origin and extensive gene exchange between disparate groups [26,27,38]. Beyond these core proteins, major viral groups might considerably differ from each other in their protein complements. The main morphological and genomic differences appear to mirror the three superkingdoms of cellular life; thus, viruses infecting bacteria, archaea and eukaryotes do show considerable differences between each other (Fig. 1) [38,39]. For example, the caudate morphology (tailed-bacteriophages) is primarily restricted to bacterial viruses, whereas several unusual morphologies such as bottle-shaped (ampullavirus), lemon-shaped (fusellovirus), two tailed (bicaudavirus) and hooked-filamentous forms (lipothrixvirus) are unique to archaeal viruses [39]. Some viral groups such as certain caudate phages in bacteria, poxviruses, iridoviruses, phycodnaviruses and herpesviruses are observed across phylogenetically diverse sets of host species. However, other groups such as baculoviruses and polydnaviruses appear to be restricted to certain arthropod lineages.