Mini-Review

Chromodomains and LTR retrotransposons in plants

Olga Novikova

Institute of Cytology and Genetics SB RAS; Novosibirsk, Russia

**Abbreviations:** HP1, heterochromatin protein-1; MeK, methyllysine; H3 Me2K9, dimethyllysine at position 9 of histone H3; H3 Me3K9, trimethyllysine at position 9 of histone H3; CHD1, human chromo-helicase/ATPase DNA-binding protein 1; Pol, polyprotein; TEs, transposable elements; LTRs, long terminal repeats; PR, protease; RT, reverse transcriptase; RNaseH, ribonuclease H; Int, integrase; YFP, yellow fluorescent protein

**Key words:** chromodomain, LTR retrotransposons, targeted integration, plants, evolution

A chromodomain is a domain contained in various proteins involved in chromatin remodeling and the regulation of gene expression in eukaryotes during development. Chromodomains perform a wide range of diverse functions including chromatin targeting and interactions between different proteins, RNA and DNA. The chromodomains also have been found as an additional domain at the C-terminal region of Polyproteins (Pol) encoded by transposable elements, which belong to the Gypsy LTR retrotransposons superfamily. Chromoviruses or chromodomain-containing Gypsy LTR retrotransposons form the most widespread clade of Gypsy LTR retrotransposons and can be found in diverse eukaryotes including plants, fungi and vertebrates. The recent finding suggested that chromodomains can be responsible for the targeted integration of LTR retrotransposons and, thus, should be favorable for mobile elements by allowing them to avoid negative selection arising from insertion into coding regions.

**Chromatin Organization Modifiers and Chromatin Regulation**

The chromodomains (chromatin organization modifiers) are protein structural domains of about 40–50 amino acid residues commonly found in proteins associated with the remodeling of chromatin. Chromodomains have been identified in a variety of proteins including factors that activate and repress transcription. More than 40 examples of chromodomains are known so far. Chromodomains were originally identified as a conserved sequence motif between Polycomb, a protein required for homeotic gene repression, and heterochromatin protein-1 (HP1), a component of centromeric Polycomb, a protein required for homeotic gene repression, and were originally identified as a conserved sequence motif between Drosophila melanogaster and human chromosomes (Me2K9) or trimethyllysine (Me3K9). The MeK recognition domain of HP1 is essential for direct heterochromatin binding, whereas the ‘shadow’ domain is necessary for nuclear localization.

There are two major groups of chromodomains found in eukaryotic chromodomain-containing proteins. So-called ‘classical’ chromodomains carry the characteristic chromo-box motif [Yf][Ly][LIV][W][W]y[kr] (single letter code, capital letters standing for the most prominent amino acid). The second group of chromodomains is more variable and, includes chromo-related domains, which are well conserved in their central part (strand β2), but deviate significantly in other regions (strand β1 and α-helix) (Fig. 1). The best-known protein with both types of chromodomains is HP1.

HP1 is a hallmark of constitutive heterochromatin, a condensed and highly repressive type of chromatin that organizes the repetitive pericentromeric DNA. This protein contains an N-terminal ‘classical’ chromodomain and a C-terminal ‘shadow’ chromo-related domain. Mutational analysis showed that the N-terminal chromodomain of HP1 is essential for direct heterochromatin binding, whereas the ‘shadow’ domain is necessary for nuclear localization.

So far the function of Drosophila HP1 N-terminal chromodomain is the most investigated of all known chromodomains. This domain of HP1 is sufficient for specific interactions with the histone H3 tail, in a manner that depends on the methylation of Lys9, dimethyllysine (Me2K9) or trimethyllysine (Me3K9). The MeK recognition involves a conserved aromatic pocket, whereas interactions between H3 tail and the chromodomain consist of a series of backbone hydrogen bonds and complementary surfaces formed between their side chains. The aromatic pocket is formed by three conserved aromatic residues Y24, W45 and Y48.

http://www.landesbioscience.com/journals/cib/article/7702

Correspondence to: Olga Novikova; Department of Molecular Evolution; Institute of Cytology and Genetics SB RAS; Pr. Lavrentjeva 10; Novosibirsk 630090 Russia; Email: novikova@bionet.nsc.ru

Submitted: 12/19/08; Accepted: 12/23/08

Previously published online as a Communicative & Integrative Biology E-publication: http://www.landesbioscience.com/journals/cib/article/7702
of HP1 binds to the pentapeptide motif [PL][WRY][V][MIL][MLV] that has been seen in non-histone proteins which interact with HP1.16,17 Interestingly, HP1 may form homodimers and subsequently compact chromatin into higher order heterochromatin structures when bound to histone H3 MeK.18 Collectively this structure would contribute to an overall transcriptionally repressive chromatin environment.

The fact that the aromatic cage residues as well as the residues that responsible for H3 tail sequence specificity can be found in the number of chromodomains from other chromodomain-containing proteins suggests a common function. The studies did not reveal a universal principle by which all chromodomains act, but uncovered a potential for diverse molecular interactions.19 For example, the Polycomb chromodomains interacts with H3 MeK27, 20,21 whereas the tandem chromodomains of human chromo-helicase/ATPase DNA-binding protein 1 (CHD1) act cooperatively to specifically bind H3 MeK4,22 a mark characteristic of euchromatin. MOF protein is histone acetylase which targeted to the male X chromosome in Drosophila where it acetylates histone H4 at Lys16.23 This acetylation is critical to the activation of X-linked genes as a mechanism to compensate for the reduced gene dosage in male versus female somatic cells.24 MOF uses its chromodomains to bind roX RNA.9

Chromoviruses as LTR Retrotransposons with a Chromodomain

Surprisingly, the chromodomains also have been found as an additional domain at the C-terminal region of Polyproteins (Pol) encoded by transposable elements from Gypsy LTR retrotransposon superfamily.25,27 Transposable elements (TEs) have the ability to replicate and spread in genome. They were found virtually in all investigated eukaryotes and represent the ubiquitous components of eukaryotic genomes. For example, at least 46% of the human genome is represented by TE-derived sequences28 and more than 70% of the plants genomes can be composed by TEs.29 The relative abundance and diversity of TEs have contributed to the differences in the structure and size of eukaryotic genomes, especially in the plant kingdom.30 Transposable elements have had a profound influence on the evolution of eukaryote genomes.31 Recent evidence suggests that TEs may provide the genome with potent agents to generate genetic and genomic plasticity. Insertions of TEs near genes can lead to alterations in gene expression patterns—since the elements usually contain transcriptional regulatory sequences—while insertions within genes can directly alter gene structure. Recombination between elements at different sites can lead to large-scale chromosomal rearrangements. TEs may have reshaped the human genome by ectopic rearrangements, by creating new genes, and by modifying and shuffling existing genes.28 In some cases, TEs perform critical biological functions in their host. For example, the preferential insertion of some retrotransposons in Drosophila at telomeric locations has removed the need for telomerase function.32

Two classes of TEs are known currently. Class II TEs, or DNA transposons, utilize DNA-based modes of transposition including ‘cut-and-paste’ mechanism, rolling-circle replication, and a mechanism which involves DNA polymerase and is not yet well understood.33,34 Class I, or retrotransposons, are mobile genetic
Chromodomains and LTR retrotransposons

Although the genomic data for basal eukaryotes is very limited, it is believed that chromoviruses are absent in the genomes of the most basal eukaryotic lineages (Diplomonadida, Euglenozoa, Alveolata and Bacillariophyta). It was proposed that chromoviruses appeared no earlier than in Cercozoa.27

There are two distinct groups of retrotransposon chromodomains. Group I is characterized by the presence of the three conserved aromatic residues, which are the same as the conserved residues form aromatic cage of ‘classical’ N-terminal chromodomain from HP1.4,15,43 This group of chromodomains was found in diverse eukaryotic LTR retrotransposons including fungal and vertebrate Gypsy elements. Representatives of the group II of chromodomains lack the first conserved aromatic residue and usually the third. Group II was identified only in plant retrotransposons. The structure-based alignment allows attributing the group II retrotransposons chro-

Chromoviruses in Plants

The chromosomes of plants are littered with retrotransposons that, in many cases, constitute as much as 80% of plant genomes. LTR retrotransposons have been especially successful colonizers of plant chromosomes. Examination of their function, evolution and dispersal is essential to the understanding the evolution of eukaryotic genomes. A number of studies were performed in attempt to describe the diversity of Gypsy and Copia retrotransposons in the plant kingdom. Comprehensive studies of LTR retrotransposons in plant genomes are regularly published.44-47 However, these works
concerned mainly agricultural species or model organisms and the majority of the known families of LTR retrotransposons have been found and fully characterized in angiosperms while other large groups are less covered. The recently published survey of chromodomain-containing Gypsy LTR retrotransposons from mosses performed in our laboratory suggested that the diversity of plants LTR retrotransposons is highly underestimated and the evolutionary history of Gypsy LTR retrotransposons from plants is still not fully resolved.

The phylogenetic analysis of number of chromodomain-containing Gypsy LTR retrotransposons showed that they represented by four separate clades in the genomes of higher (seed) plants: Tekay, Galadriel, CRM and Reina (Fig. 2). At least two additional clades, indicated as A and B on the Figure 2, can be found in genomes of non-seed plants such as mosses, ferns and lycophytes (unpublished data). The analyses of whole genomic sequence of moss Physcomitrella patens (Bryophyta, Funariales) along with experimental investigation of diverse moss species allowed us to describe several new chromodomain-containing Gypsy LTR retrotransposons belonging to clades A and B. The most interesting finding in our investigation of Gypsy LTR retrotransposons from non-seed plants is that two types of retrotransposon chromodomains (group I in clade A and group II in clade B) indeed are represented in non-seed plant genomes. At the same time, only group II chromodomains are known in angiosperm and gymnosperm Gypsy LTR retrotransposons. It has to be noted that representatives of CRM clade have no detectable chromodomains. Instead, they showed the presence of the conservative protein domain, which was named ‘CR motif’. The role of this motif is unknown.

The information on the role of chromodomains in retrotransposition of LTR retrotransposons is limited. The chromo-integrase of Tf1 LTR retrotransposon from Schizosaccharomyces pombe lacking the chromodomains demonstrated a significantly higher activity and substantially reduced substrate specificity. This suggests that both the activity and specificity of enzymes can be modulated by their chromodomains. Conversely, the transposition activity of MAGGY retrotransposon of the rice blast fungus Magnaporthe oryzae dramatically decreased with the loss or alteration of chromodomain. As was demonstrated recently, MAGGY chromodomains, which belonged to the group I chromodomains, interacts with H3 Me2K9 and H3 Me3K9 similarly to HP1 ‘classical’ chromodomains. Moreover, it was proposed that chromodomains can target integration of chromoviruses into heterochromatin regions. Almost nothing is known about activity and role of group II chromodomains from plant retrotransposons. The mostly heterochromatic distribution of plant chromoviruses along with data suggesting the localization of chromodomain-YFP fused protein in heterochromatin can be used as an indirect evidence for recognizing heterochromatin and directing the integration role of group II chromodomains. Nevertheless, the mechanisms by which chromodomains group II act are still unknown.

As a result of potentially deleterious effects of transposable element proliferation, host organisms have often evolved mechanisms that limit TE activity, such as specific methylation. Similarly, since TEs are, in general, only rarely transmitted across species boundaries, their continued existence is usually dependent upon the continued survival of their hosts. As a result, transposable elements themselves often appear to have evolved mechanisms, such as directing their integration to specific parts of the genome, which keep the damage they cause to a minimum.

The further studies of retrotransposon chromodomains role should provide insights not only into the mechanism of targeted integration of LTR retrotransposons but also to general evolutionary issues. Several evolutionary questions can be addressed by such investigations. For example, how the shift from LTR retrotransposons with group I chromodomains to those with group II containing has effect on the organization of plant genomes? Is the difference in chromodomains activity played key role in this shift? Why do mosses contain predominantly group I retrotransposon chromodomains while seed plants contain only group II retrotransposon chromodomains? Is there a difference in beneficial features provided for LTR retrotransposons by group I and group II chromodomains?

We have a unique opportunity for in vivo investigation of the activity of the retrotransposon group I and group II chromodomains in a homogenous system. Moss Physcomitrella patens can be used as a model organism for the analysis of retrotransposon chromodomains activity and targeted integration. The haploidy of the dominant gametophyte stage in moss development makes mosses attractive material for genetic studies because isolation of mutants and genetic analysis are more straightforward than in species with a dominant diploid phase. The potential of mosses as model systems to study plant biological processes is reinforced by their suitability for cell lineage analysis and similar responses to plant growth factors and environmental stimuli to those observed in other plants. Moreover, the P. patens genome sequence has been published recently, which provides the basis for a comprehensive survey of transposable elements, and investigation of retrotransposon distribution in genomic sequences and their target site preferences.

Acknowledgements
Author thanks Dr. Marie-Angèle Grandbastien (Institut Jean-Pierre Bourguin, INRA, France) for the helpful comments and Dr. Victor Fet (Marshall University, West Virginia, USA) for his stylistic suggestions.

References
1. Koonin EV, Zhou S, Lucchesi JC. The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. Nucleic Acids Res 1995; 23:4229-33.
2. Aasland R, Stewart AF. The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. Nucleic Acids Res 1995; 23:3168-74.
3. Pato R, Hogen D. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc Natl Acad Sci USA 1991; 88:263-7.
4. Jacobs SA, Khorasanizadeh S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science 2002; 295:2080-3.
5. Chen CC, Hwang JK, Yang JM. (PS)2: protein structure prediction server. Nucleic Acids Res 2006; 34:152-7.
6. Bull IJ, Mutzina NV, Broadhurst RW, Raine AR, Archer SJ, Scott FJ, et al. Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. EMBO J 1997; 16:2473-81.
7. Brehm A, Tufveland KR, Aasland R, Becker PB. The many colours of chromodomains. Bioessays 2004; 26:133-40.
