Biochimica et Biophysica Acta xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Review

Structure-specific nucleic acid recognition by L-motifs and their diverse roles in expression and regulation of the genome

Roopa Thapar *

BioSciences at Rice-Biochemistry and Cell Biology, Rice University, Houston, TX 77251-1892, USA

ARTICLE INFO

Article history:
Received 18 December 2014
Received in revised form 5 February 2015
Accepted 24 February 2015
Available online xxxx

Keywords:
HMG-box domain
Stem-loop binding protein (SLBP)
RNA binding domain
Phosphorylation
DNA binding domain
Protein-nucleic acid interaction

ABSTRACT

The high-mobility group (HMG) domain containing proteins regulate transcription, DNA replication and recombination. They adopt L-shaped folds and are structure-specific DNA binding motifs. Here, I define the L-motif super-family that consists of DNA-binding HMG-box proteins and the L-motif of the histone mRNA binding domain of stem-loop binding protein (SLBP). The SLBP L-motif and HMG-box domains adopt similar L-shaped folds with three α-helices and two or three small hydrophobic cores that stabilize the overall fold, but have very different and distinct modes of nucleic acid recognition. A comparison of the structure, dynamics, protein-protein and nucleic acid interactions, and regulation by PTMs of the SLBP and the HMG-box L-motifs reveals the versatile and diverse modes by which L-motifs utilize their surfaces for structure-specific recognition of nucleic acids to regulate gene expression.

1. Introduction

Nucleic acid binding motifs in proteins generally recognize specific sequences in the cognate RNA or DNA using modular domains. While the structural basis for sequence-specific recognition of DNA and RNA is easier to understand, it is quite challenging to understand the mechanism of structure-specific recognition or "shape readout" of both DNA and RNA [1]. Tertiary interactions between secondary structure elements in RNAs such as the tetraloop-receptor interaction [2], A-minor interactions [3], kissing hairpin loops [4], pseudoknots [5], ribose zippers [2], adenosine wedges [6] etc can result in a infinite number of unique RNA folds [7–10]. Noncanonical DNA structures are also present as intermediates in DNA recombination, repair, and replication and are frequently stabilized by proteins [11–15]. Nucleic acid binding proteins that are structure-specific either recognize a pre-formed structure in RNA and DNA, or can induce DNA bending, DNA looping, helical distortions and unfolding of RNA [16]. One functional consequence of structure-specific nucleic acid recognition is to facilitate the formation of higher-order multi-protein DNA/RNA complexes whereby binding of one protein triggers conformational changes in the nucleic acid that are recognized by another protein in the macromolecular complex.

Here I highlight the structure and mechanism of RNA recognition of a new RNA binding fold, the L-motif, that is present in the histone mRNA specific processing factor, Stem-Loop Binding Protein (SLBP) [17–19], also known as Hairpin Binding Protein (HBP) [20] and compare it to the canonical L-motif observed in HMG-box domains [21] that bind DNA. The high mobility group (HMG) proteins are abundant non-histone nuclear proteins that associate with chromatin, as has been summarized in some excellent recent reviews [22–24]. In 1990, R. Tjian and colleagues proposed that the HMG-box motif from Upstream Binding Factor or UBF is a novel DNA binding motif [25]. Since then, the HMG-box domain has been identified in plants, yeast and vertebrates and is involved in transcription, DNA replication, recombination and repair [22,26]. The L-shaped scaffold is quite versatile and can bind and bend DNA in the minor groove. It also plays a role in DNA looping [27]. These boxes have high affinity (K_ds ~10^{-9} M) towards four-way junction DNA [28], cruciform DNA [29–31], cisplatin-modified DNA [32–34] as well as tRNA [30], double-stranded RNA [35,36] and single-stranded RNA [37]. There are two broad subfamilies of HMG-box containing proteins: those that bind bent or distorted DNA in a non-sequence-specific manner and have two or more tandemly arranged HMG-box domains followed by an acidic tail that has a regulatory function; and a second class of sequence-specific single HMG-box containing transcription factors usually with no acidic tail [38,39]. HMG-box proteins from the former class such as HMG1 [40], HMG2 [41], HMGD [42], and NHP6a [43] bind pre-bent DNA such as Holliday junctions or cisplatin-modified DNA [34,44]. HMG-box proteins from the second class, such as lymphoid enhancer-binding factors TCF-1 and LEF-1 [45–47], sex-determining region Y (SRY) [48], and the SRY-related HMG box (Sox) family [49], are observed frequently in transcription factors that bend specific DNA sequences. The HMG domains are highly

Abbreviations: 3′ UTR, 3′ untranslated region; PTMs, posttranslational modifications; r.m.s.d., root-mean-square deviation

* Tel.: +1 713 201 7875; fax: +1 713 348 5154.
E-mail address: rthapar@rice.edu.

http://dx.doi.org/10.1016/j.bbagrm.2015.02.006
1874-9399/

selective but in a subtle fashion such that it is difficult to predict DNA binding preferences. For example, the HMG1 box A binds four-way junction DNA with a K_d of 200 nM, but the UBF box has low affinity (K_d 1.5 μM) towards the same DNA [50]. The sequences and structures of HMG-box domains from both subgroups are remarkably similar.

Interestingly, recent crystal structures of the histone mRNA specific RNA processing factor, Stem-Loop Binding Protein (SLBP) [18] bound to histone mRNA stem-loop and the exonuclease 3′hExo/ERI1 reveal that its RNA binding domain (RBD) is structurally related to the HMG-box domain. SLBP recognizes the structure of the A-form RNA hairpin and distorts and unfolds the RNA tetraloop. There is no sequence similarity between the SLBP L-motif and HMG-box domains, indicating they are evolutionarily distant. However the overall folding topologies and their architectural functional roles are similar. The similarities in structure, dynamics, and regulation by posttranslational modifications of the SLBP RNA binding L-motif and the DNA-binding HMG-box domains lead to new hypotheses. Do HMG-box proteins bind RNA, and do they play a role in RNA processing? There is some experimental evidence in the literature that this may be a plausible idea. Does SLBP play a direct role in DNA replication? No functional roles for SLBP besides its role in histone mRNA metabolism have been described. Herein I compare the SLBP L-motif with HMG-box domains and discuss their distinct modes of structure-specific recognition of RNA and DNA, respectively.

2. The HMG-box L-motif fold

High-resolution structures of several HMG-box domains have been solved by NMR spectroscopy [21,51–58] and X-ray crystallography [59,60]. The HMG-box domain is an L-shaped DNA-binding motif (Fig. 1A, B) consisting of ~75–80 amino acids and three α-helices [21]. The angle between the two arms (helix-2 and helix-3 or the major wing) of the "L" shape is ~80° and can show ~20° variation between different HMG-box structures [61]. The long arm of the "L" (also known as the minor wing) consists of helix-3 and the extended N-terminus, whereas the short arm is made up of helix-1 and helix-2. An extended N-terminal strand packs against helix-3. The orientation of helix-1 differs slightly in many HMG-box structures. The loops that connect the helices can vary in length. The position of helix-2 and helix-3 to form the L-shape is maintained by interactions between conserved aromatic and aliphatic residues that form a compact hydrophobic core (Fig. 1C, D). The solution NMR structure of the B domain of HMG1 (PDB code 1HME) [21] shows stacking interactions between Phe14, Phe17, Trp45, Lys53, and Tyr56 side chains to form the major tightly packed hydrophobic cluster (HC1) (Fig. 1C). In addition, Pro7 and Pro10 form a second hydrophobic core (HC2) in the N-terminal extension that stack against Tyr67 of helix-3, thereby stabilizing the overall fold (Fig. 1D).

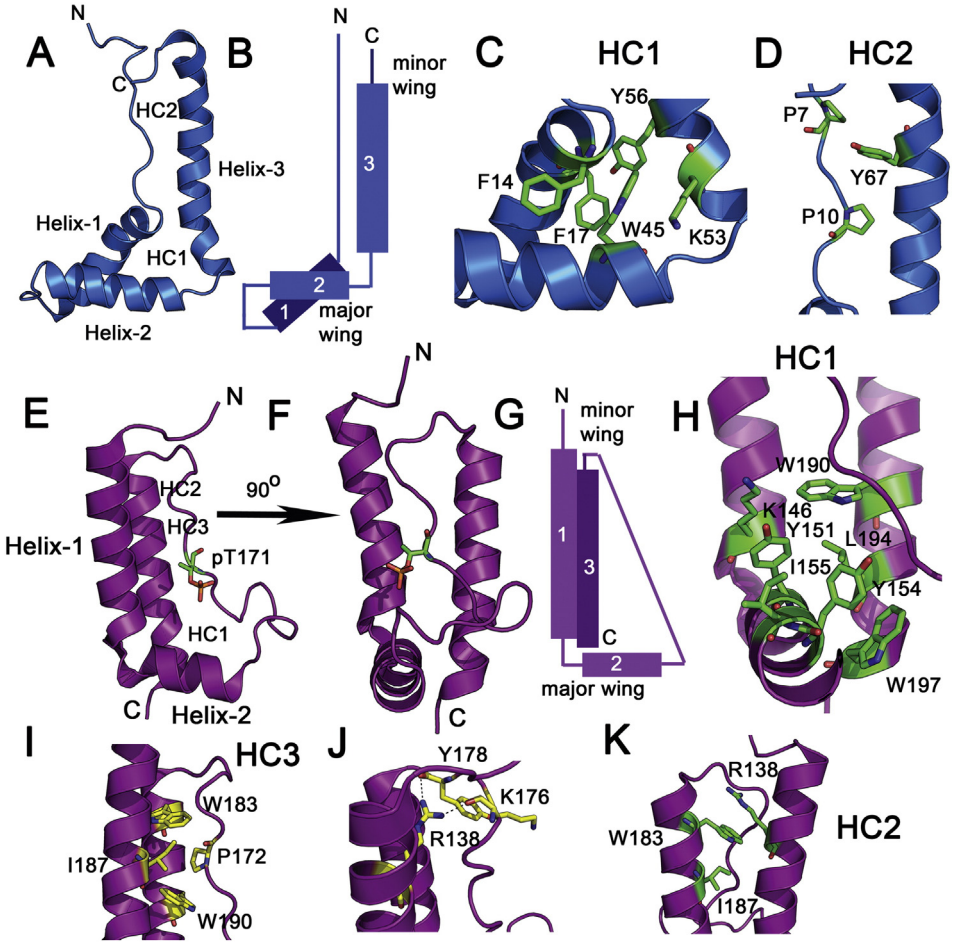

Fig. 1. (A) Structure of the HMG-box fold from HMG1A (PDB code 1HME) is shown. The location of the two hydrophobic cores, the major core 1 (HC1) and minor core 2 (HC2) are labeled. α-Helices 1 and 2 form the major wing and Helix-3 along with the N-terminal strand forms the minor wing. (B) Schematic of the L-shape fold of the HMG-box domain is depicted. (C) Interaction between hydrophobic side chains in HC1 of HMG1A located in the major wing. (D) Interaction between Y67, P7, and P10 in HC2 of HMG1A is depicted. (E) Structure of the L-motif of SLBP from the structure of the T171 phosphorylated SLBP-histone mRNA stem-loop-3′hExo ternary complex (PDB code 4QOZ) is shown. The phosphothreonine is shown in green stick. The location of the three hydrophobic cores, HC1, HC2, and HC3 is labeled. The front view is shown in (E) and the side view is shown in (F). (G) Schematic of the L-motif fold of SLBP is depicted. (H) Interaction between hydrophobic side chains in HC1 of SLBP located in the major wing. Interaction between hydrophobic side chains in HC2 and HC3 of SLBP located in the minor wing is shown in (I) and (K). (J) A stabilizing salt bridge interaction made by the side chain of R138 with the backbone carbonyl oxygens of Y178 and K176 locks the turn in place in the minor wing.

Please cite this article as: R. Thapar, Structure-specific nucleic acid recognition by L-motifs and their diverse roles in expression and regulation of the genome, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.02.006