Biochimica et Biophysica Acta 1799 (2010) 101–113

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

# HMGB proteins: Interactions with DNA and chromatin

Michal Štros *

*Laboratory of Analysis of Chromosomal Proteins, Academy of Sciences of the Czech Republic, Institute of Biophysics, Královopolská 135, 612 65 Brno, Czech Republic*

ARTICLE INFO

*Article history:*
Received 10 August 2009
Received in revised form 9 September 2009
Accepted 15 September 2009

*Keywords:*
HMGB
HMG-box
Architectural protein
DNA–protein interaction
Chromatin

ABSTRACT

HMGB proteins are members of the High Mobility Group (HMG) superfamily, possessing a unique DNA-binding domain, the HMG-box, which can bind non-B-type DNA structures (bent, kinked and unwound) with high affinity, and also distort DNA by bending/looping and unwinding. HMGBs (there are four HMGBs in mammals, HMGB1–4) are highly abundant and ubiquitously expressed non-histone proteins, acting as DNA chaperones influencing multiple processes in chromatin such as transcription, replication, recombination, DNA repair and genomic stability. Although HMGB1 is a nuclear protein, it can be secreted into the extracellular milieu as a signaling molecule when cells are under stress, in particular, when necrosis occurs. Mammalian HMGBs contain two HMG-boxes arranged in tandem, share more than 80% identity and differ in the length (HMGB1-3) or absence (HMGB4) of the acidic C-tails. The acidic tails consist of consecutive runs of only Glu/Asp residues of various length, and modulate the DNA-binding properties and functioning of HMGBs. HMGBs are subject to post-translational modifications which can fine-tune interactions of the proteins with DNA/chromatin and determine their relocation from the nucleus to the cytoplasm and secretion. Association of HMGBs with chromatin is highly dynamic, and the proteins affect the chromatin fiber as architectural factors by transient interactions with nucleosomes, displacement of histone H1, and facilitation of nucleosome remodeling and accessibility of the nucleosomal DNA to transcription factors or other sequence-specific proteins.



## 1. Introduction

The eukaryotic genomic DNA must be highly condensed in a dynamic supramolecular nucleoprotein structure – chromatin – to fit into the cell nucleus and perform its function. The fundamental repeat unit of chromatin, the nucleosome, represents a structure formed by coiling of the DNA around histone octamers [consisting of the central $(H3/H4)_2$ tetramer and two peripheral H2A/H2B dimers] [1]. Chromatin is characterized by repeating units of nucleosomes arranged in a beads-on-a-string nucleosomal chain, stabilized by histone H1 into the 30-nm chromatin fiber forming the higher-order chromatin structure [2]. Modulation of chromatin folding and remodeling of the chromatin structure affects access of regulatory factors to their cognate DNA binding sites which is required for regulation of fidelity of gene expression and establishing a gene phenotype. This is achieved by loosening the chromatin structure or even disruption of the nucleosome structure (by chromatin remodeling complexes), by DNA bending and unwinding, as well as by affecting the strength of DNA–histone interactions by DNA methylation, post-translational modifications of histones, or incorporation of specific histone variants to chromatin to allow access of specific transcription factors or other proteins [3].

Histone H1 family (or linker histones) represents the major architectural proteins that can bind most of the nucleosomes in metazoan cells restricting access to transcription machinery and other DNA-dependent processes, such as suppressing chromatin remodeling [4]. Many of the structural changes in chromatin are also mediated by a large and diverse superfamily of HMG (High Mobility Group) proteins that can bind to nucleosomes in a non-sequence specific manner [5]. The HMG proteins have been subdivided into three distinct structural families: HMGA (the HMG-AT-hook family), HMGN (the HMG-nucleosome binding family), and HMGB (the HMG-box family) [6,7]. Members of each family are abundantly and ubiquitously expressed in most eukaryotic cells, exhibit different structures and unique DNA or chromatin-binding motifs but they all affect the chromatin fiber as architectural factors. Association of HMG proteins is not confined to specific sites but it is rather highly dynamic, and the proteins can scan the potential chromatin binding sites and move from one chromatin site to another in a " hit and run" fashion [7].

In this review I focus on proteins of the HMGB family, with the main emphasis on nuclear functions of HMGB proteins, binding to DNA and chromatin, as well as their functioning as architectural factors ("DNA chaperones"). Most of the presented data on DNA/chromatin interactions are related to vertebrate HMGB1 and/or HMGB2 as similar binding experiments have not yet been performed with other mammalian HMGB proteins (HMGB3 and HMGB4).

## 2. The HMG-box family

All members of the HMG-box family possess a DNA-binding domain (the HMG-box) related to a motif originally identified in

* Tel.: +420 541517183; fax: +420541211293.
*E-mail address:* stros@ibp.cz.

1874-9399/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagrm.2009.09.008

vertebrate HMGB proteins. Mammalian HMG-box-containing proteins are usually classified into two major groups distinguished by their abundance, function and DNA specificity. The first group consists of proteins containing two HMG-boxes [e.g., HMGB1-4 proteins, the mitochondrial factors mtTF1 and ABF2, and similar proteins to be found in other species: *Drosophila* Dorsal Switch Protein 1 (DSP1) and yeast HMO1/2 proteins], and RNA polymerase I transcription factor UBF with four to six HMG-boxes [8]. In general, two or more HMG-boxes are mostly found in abundant HMG-box proteins with little or no DNA sequence specificity, albeit some *non*-sequence specific proteins contain a single HMG-box such as plant "HMGB"-type proteins, dipteran insects *Chironomus* cHMGB1/2, *Drosophila* HMG-D/Z, and *S. cerevisae* Nhp6A/B (reviewed in refs. [8–10]). The second group of mammalian HMG-box proteins is highly diverse and consists of much less abundant proteins having mostly a single HMG-box (such as TCF/LEF-1, TOX, sex-determining factor SRY and SOX proteins, chromatin remodeling factors BAF57 and PB1, reviewed in ref. [8]). The single HMG-box proteins recognize specific DNA sequences but the specificity is restricted due to the limited number of base-specific hydrogen bonds that can be formed within the minor groove (see Section 3). Search of the human protein database (http://www.ebi.ac.uk/swissprot/) using BLASTP and amino acid sequence of the HMGB1-A+B di-domain (HMGB1-ΔC) revealed ~50 HMG-box-containing proteins of 15–193 kDa, with HMGB proteins representing only a small and specific subset of HMG-box proteins [8].

HMGB1 and HMGB2 were discovered ~35 years ago as abundant nonhistone DNA-binding proteins in calf thymus and their name originate from their (anomalous) high electrophoretic mobility in triton-urea gels (High Mobility Group, HMG) due to a high content of positively and negatively charged amino acid residues explaining their extractability in diluted solutions of acids [11]. HMGB1 protein exhibits not only *high* mobility in polyacrylamide gels, but also in the nucleus being the most mobile and dynamic nuclear protein [12]. There are three canonical HMGB proteins in human and mice: HMGB1, HMGB2 and HMGB3 ([13], reviewed in ref. [8] and refs. therein), and recently discovered HMGB4 [14]. HMGB1-3 proteins have a molecular mass of ~25 KDa, contain two DNA-binding domains (the HMG-boxes A and B) and a long acidic C-terminal tail (Fig. 1). HMGB1-3 share more than 80% identity and mainly differ in the length of their acidic C-tails. HMGB4 protein has a molecular mass of ~21 kDa and also contains two HMG-boxes but lacks the acidic tail [14].

HMGB1 (which is identical to the neurite outgrowth-promoting protein called p30 or amphoterin [15] or sulphoglucuronyl carbohydrate binding protein, SBP-1, in rat cerebral cortex and cerebellum [16]) is the most abundant non-histone protein in the nucleus (approximately 1 molecule per 10–15 nucleosomes). HMGB1 is an evolutionarily highly conserved protein in mammals and amino acid sequences of all mammalian HMGB1 proteins are virtually identical (>99%), implying similar biological functions in distinct organisms.

## 3. Structure of the HMG-box

The solution structures of individual HMG-boxes, A and B, of HMGB1 have been determined by NMR spectroscopy [17,18,25]. An HMG-box contains ~75 amino acids and has a characteristic L-shaped fold consisting of three α-helices with an angle of ~80° between the arms. The long arm includes the extended N-terminal strand and helix III (the minor wing), while the short arm is composed of helices I and II (the major wing) (Fig. 1B) The overall structure of the HMG-box is far more conserved than the corresponding amino acid sequences of the different HMG-boxes [26]. The HMG-boxes A an B of mammalian HMGB1 differ slightly in the relative geometries of helices I and II with respect to the rest of the molecule [17,18,25]. The length of the loop between helices I and II is longer in the HMGB1-box A [17,18,25] but this difference is not critical for binding affinity of the two HMGB1-boxes for bent (cisplatin-modified) DNA [27]. In addition, helix I is short and straight in HMG-box A (but bent in HMG-box B), and the distribution of charged surface residues of the two HMG-boxes is different (HMG-box A is more positively charged in the helices I/II than the HMG-box B [28]). The latter differences likely account for a clear preference (high affinity) of the HMGB1-box A for distorted DNA substrates such as cisplatin-modified DNA or four-way junctions [27–29]. The published NMR structures of the HMG-boxes of *Drosophila* HMGB1-like protein HMG-D [30], yeast HMGB1-like Nhp6a [31], lymphoid transcriptional activator SOX-4 [32], and human UBF HMG-boxes [33] resemble closely the HMG-box B of HMGB1 [17,25]. Several structures of single HMG-boxes in complex with linear DNA (LEF-1 [34]; sex-determining factor SRY [35]; yeast NHP6A [31]; *Drosophila* HMG-D [36]), or with bent, cisplatin-modified, DNA (domain A of HMGB1 [29]) have been determined. The structures of free HMG-boxes or bound to DNA are similar (structures of other HMG-boxes are to be found at http://www.ncbi.nlm.nih.gov), and binding of the HMG-box to DNA is accompanied by intercalation of bulky hydrophobic amino acid residues of the HMG-boxes between successive base-pairs within the DNA minor groove (see Section 4.2).

Mammalian HMG-box containing proteins consist of two major groups, distinguished by their abundance, function and sequence-specificity in DNA binding. Domain swap experiments between the non-sequence specific HMGB1-box B and the sequence specific HMG-box of TCF1α [T cell receptor α chain enhancer] revealed that the sequence specificity of the HMG-box lies in the minor wing of the DNA-binding motif [37]. The differences between sequence-specific and non-sequence specific minor groove DNA binding of HMG-boxes depend on changes in only a few amino acid residues (e.g., sequence-specific HMG-boxes of SRY or LEF-1 contain asparagine whereas non-sequence-specific HMGB proteins have mostly serine; hydrophilic and hydrophobic residues are present within the loop between helices I and II in sequence-specific or non-sequence specific HMG-boxes, respectively, reviewed in ref. [[38]). Although intercalating residues of the HMG-box and also residues that buttress the intercalating residues (see Section 4.2) are implicated as specificity determinants, it is likely that the specificity rather depends on other factors such as the ability of the sequence-specific HMG-box to achieve better shape complementarity with the DNA binding surface ([39] and refs. therein).

## 4. Interaction of HMGB proteins with DNA

Relatively soon after isolation of HMGB1 and HMGB2 [11] it was found that the proteins could bind DNA with a clear preference to noncanonical DNA structures such as single-stranded DNA [40], DNA containing cruciforms or bent structures [41], supercoiled DNA and Z-DNA [42]. However, the ability of HMGB1 to bind preferentially Z-DNA [42] has been challenged in later studies ([43] and refs. therein). The highest affinity of HMGB1 to date was reported for hemicatenated DNA loops ($K_D < 0.2 \times 10^{-12}$ M ) [44], followed by DNA minicircles ($1 \times 10^{-10}$ M) [45], four-way junctions ($1 \times 10^{-9}$ M) [41] and DNA damaged by chromium(VI) adducts (~$10^{-9}$ M) [46] as compared to the affinity of the protein to linear (B-type) DNA ($5 \times 10^{-5}$ M). HMGB1/2 proteins interact with DNA via their HMG-boxes, A and B. HMG-boxes A and B exhibit differences in their DNA binding and bending activities, partially due to the modulatory effect of their flanking sequences [47–53]. Although individual HMGB-boxes can bind to DNA, the binding is enhanced when the two domains are covalently linked in the A+B di-domain [50,51,54,55].

A characteristic feature of HMGB1/2 proteins and all known HMG-boxes is binding to distorted (bent and unwound) DNA substrates ([8,9,56] and refs. therein). Despite identification of numerous high-affinity binding DNA structures for HMGB1 and HMGB2 in vitro, the natural DNA binding sites of the HMGBs have not been identified so far. However, it is likely that the DNA structures for