HOPX: The Unusual Homeodomain-Containing Protein

Anita Mariotto[1], Olesya Pavlova[1], Hyun-Sook Park[1], Marcel Huber[1] and Daniel Hohl[1]

The homeodomain-only protein homeobox (HOPX) is the smallest known member of the homeodomain-containing protein family, atypically unable to bind DNA. HOPX is widely expressed in diverse tissues, where it is critically involved in the regulation of proliferation and differentiation. In human skin, HOPX controls epidermal formation through the regulation of late differentiation markers, and HOPX expression correlates with the level of differentiation in cutaneous pathologies. In mouse skin, Hopx was additionally identified as a lineage tracing marker of quiescent hair follicle stem cells. This review discusses current knowledge of HOPX structure and function in normal and pathological conditions.

Journal of Investigative Dermatology (2016) **136,** 905–911; doi:10.1016/j.jid.2016.01.032

THE HOMEODOMAIN AND HOMEODOMAIN-CONTAINING PROTEINS

Homeotic genes have been largely studied in the fruit fly *Drosophila melanogaster*, focusing on the Antennapedia and Bithorax gene clusters (Kaufman et al., 1980; Lewis, 1978). The 180-base pair DNA sequence contained in these genes and named the homeobox encodes for the homeodomain (HD), which can both bind DNA and mediate protein-protein interactions (McGinnis et al., 1984a, 1984b; Scott and Weiner, 1984). Most HDs are 60 amino acids (AAs) in length and composed of three α-helices, arranged around a hydrophobic core, and a flexible amino-terminal (N-terminal) region. Helices 2 and 3 fold into the typical helix-turn-helix motif used for contacting DNA in the major grove (Kissinger et al., 1990; Laughon and Scott, 1984; Otting et al., 1990; Shepherd et al., 1984).

HDs differ in their DNA-binding properties because of significant variation in their AA composition, thus determining the specific DNA sequence recognized, although they share a common DNA docking (Berger et al., 2008; Noyes et al., 2008). Generally, Arg5 within the N-terminal region and Asn51 and Lys/Arg55 in helix 3 are almost always conserved among different HDs (Noyes et al., 2008), and many HDs choose the conserved TAAT consensus motifs as high DNA affinity binding sites (Kissinger et al., 1990; Qian et al., 1989) (Figure 1).

Homeobox genes have been differently subdivided into superclasses, classes, subclasses, or groups (Holland et al., 2007). Generally, the most described groups are the ANTP, PRD, LIM, POU, HNF, SINE, TALE, CUT, PROS, and ZF groups (or alternative names), although there is inconsistency in the terminology used to categorize them (Holland and Takahashi, 2005; Holland et al., 2007; Ryan et al., 2006). HD-containing proteins are transcription factors involved in developmental programs and cell fate, and their target genes are varied and complex (Akam, 1987; Hayashi and Scott, 1990).

HOMEODOMAIN-ONLY PROTEIN HOMEOBOX

The homeodomain-only protein homeobox (HOPX, formerly HOP) was identified in 2002 by two independent groups (Chen et al., 2002; Shin et al., 2002) and has been subsequently named not expressed in choriocarcinoma clone 1 (NECC1), lung cancer-associated gene Y (LAGY), odd homeobox 1 protein (OB1), and cardiac HD (Cameo) (Asanoma et al., 2007; Chen et al., 2003; Yamashita et al., 2008). Phylogenetic analysis does not allow precise inference of the evolutionary origins and affinities of the HOPX gene family (Zhong et al., 2008). Holland et al. (2007) placed *HOPX* in the PRD class for three reasons: it possesses the same residue combinations conserved among human PRD HDs, it shares the intron position 46/47 observed in several PRDs, and it is highly similar in terms of sequence identity with GSC and PAX6 (38% and 36% of identity, respectively) PRD-class HDs. However, HOPX protein is unique because it does not bind to DNA, it has a single AA insertion between helices 1 and 2, and it can only approximately be assigned to the PRD homeobox gene class (Burglin, 2011; Holland et al., 2007).

HOPX presents some typical characteristics of others HD-containing proteins, such as AAs Trp and Phe in position 50 and 51 (Duboule, 1994), and Lys52 found in goosecoid and bicoid HDs, mediating specific DNA contacts (Chen et al., 2002; Shin et al., 2002; Trelsman et al., 1989) (reviewed in Kook and Epstein, 2003). The HD-like structure of HOPX is also similar the HD fold of others HD-containing proteins, characterized by three α-helices connected by short loops (Kook et al., 2006). However, HOPX possesses unique characteristics, which are different from other HD-containing proteins (Kook and Epstein, 2003). Indeed, although all previously described HD-containing proteins are more than 200 AAs long and have other sequence elements, HOPX is almost entirely composed of an HD-like sequence. DNA-binding

[1]*Laboratory of Cutaneous Biology, Service of Dermatology and Venereology, University Hospital of Lausanne (CHUV), Lausanne, Switzerland*

Correspondence: Daniel Hohl, Service de Dermatologie et Vénéréologie, CHUV, Hopital de Beaumont, BT-437, Avenue de Beaumont 29, CH-1011 Lausanne, Switzerland. E-mail: daniel.hohl@chuv.ch

Abbreviations: AA, amino acid; HD, homeodomain; Hdac, histone deacetylase; HOPX, homeodomain-only protein homeobox; mRNA, messenger RNA

Received 30 August 2015; revised 23 December 2015; accepted 4 January 2016; corrected proof published online 23 March 2016

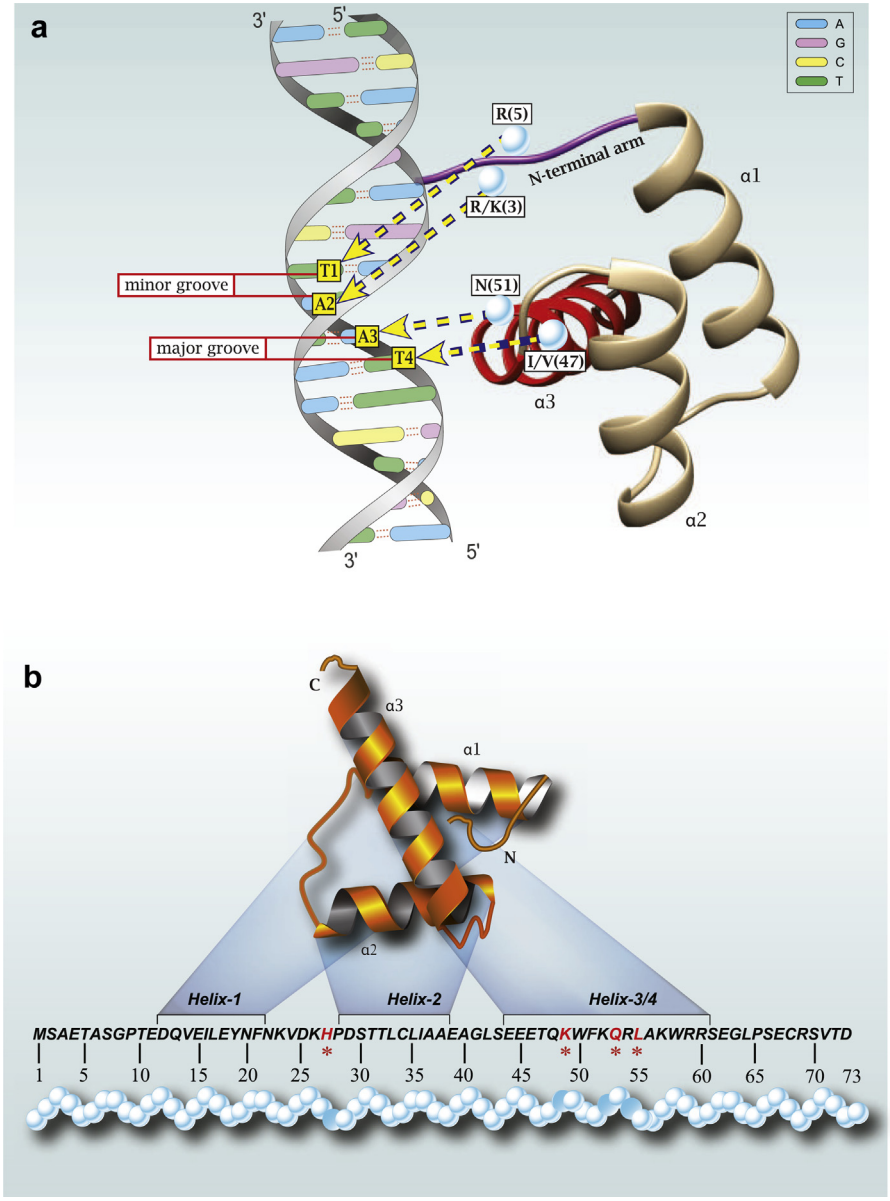

Figure 1. Three-dimensional reconstruction of ordinary homeodomain (HD)-binding DNA and atypical homeodomain-only protein homeobox (HOPX) HD. (**a**) Co-crystal structure of ordinary HD in complex with DNA (PDB ID 3HDD). The C-terminal part of helix α3 is responsible for direct interaction with the major groove of DNA. The N-terminal tail (depicted in violet) and recognition helix α3 (shown in red) of HD interact with the TAAT-binding consensus motif of DNA. Amino acids R/K(3) and R(5) (light blue pearls) of the N-terminal arm of helix α1 is involved in shape readout of the minor groove and bind nucleotides T1 and A2, whereas amino acids I/V(47) and N(51) of helix α3 (light blue pearls) contact with the major groove by binding the nucleotides T4 and A3, respectively (all interactions indicated by dotted arrows). T1, A2, A3, and T4 indicate base pairs making up the TAAT motif. I, isoleucine; K, lysine; N, asparagine; R, arginine; V, valine. (**b**) The crystal structure of the human HOPX protein containing an atypical HD. Top: Three α-helices. Bottom: HOPX protein amino acid sequence with the mutations in key residues (27, 49, 53, and 55, depicted in red) responsible for DNA binding. HOPX contains histidine (H) and leucine (L) in positions 27 and 55, respectively. HDs contain at these positions tyrosine (Y) and arginine (R), which are important for contacts with the sugar-phosphate backbone of DNA. HOPX also contains lysine (K) and glutamine (Q) in positions 49 and 53, respectively, instead of the highly conserved isoleucine (I) and asparagine (N) normally present in HDs for contacts with the major groove of DNA.

site selection assays followed by gel mobility shift assays demonstrated that HOPX is unable to bind DNA because of the lack of critical residues important for mediating DNA binding (Burglin, 2011; Chen et al., 2002; Shin et al., 2002) (Figure 1).

The human *HOPX* gene is located on chromosome 4 (4q11–q12) (National Center for Biotechnology Information data) and is composed of seven exons. Only exons 1, 5, 6, and 7 participate entirely or partially to the messenger RNA (mRNA) structures of the five mRNA transcripts existing, originated by alternative splicing (National Center for Biotechnology Information data). mRNA variant 1 (γ) encodes for a 91-AA-long protein named isoform a; mRNA variants 2 (β), 3 (α), and 4 encode for a 73-AA-long protein known as isoform b; and mRNA variant 5 encodes for a 112-AA-long protein recognized as isoform c showing a highly different C-terminal sequence compared to the other protein isoforms (National Center for Biotechnology Information data) (Ooki et al., 2010) (Figure 2). In contrast, the mouse *Hopx* gene is on chromosome 5, and there are three *Hopx* spliced variants, all encoding for a 73-AA-long protein (National Center for Biotechnology Information data) (Figure 2). Human and murine HOPX sequences are similar, sharing 92% identity at the AA level (Figure 2); however, the precise function and tissue distribution of the different transcript variants have not been defined yet (Kook and Epstein, 2003).

HOPX IN FOLLICULAR STEM CELLS AND EPIDERMAL DIFFERENTIATION

Hopx was identified as a marker of mouse hair follicle stem cells residing in the telogen basal bulge (Fuchs, 2009; Takeda et al., 2013). Previously identified in the intestinal epithelium, Hopx marks a population of quiescent, slow cycling,