

Transcription factors

*

I Introduction

II Initiation of transcription

III Transcription factors family

[pdf version](#)

III.1 Helix-Turn-Helix proteins

III.2 Zinc finger proteins

III.3 Leucine zipper proteins

III.4 Helix-Loop-Helix proteins

*

I Introduction

In eukaryotic cells, there are three different RNA polymerases (RNA Pol). Each RNA Pol is responsible for a different class of transcription : PolI transcribes rRNA (ribosomal RNA), PolII mRNA (messenger RNA), and PolIII tRNA (transfer RNA) and other small RNAs. Any protein that is needed for the initiation of transcription is defined as a transcription factor. Many transcription factors act by recognizing *cis*-acting sites that are parts of promoters or enhancers. However, binding to DNA is not the only means of action for a transcription factor. A factor may recognize another factor, or may recognize RNA Polymerases. In Eukaryotes, transcription factors, rather than the enzymes themselves, are principally responsible for recognizing the promoter.

Transcription factors are able to bind to specific sets of short conserved sequences contained in each promoter. Some of these elements and factors are common, and are found in a variety of promoters and used constitutively; others are specific and their use is regulated.

The factors that assist RNA polIII can be divided into 3 general groups:

- The **general factors**, which are required for the initiation of RNA synthesis at all class II promoters (coding genes). With RNA PolII, they form a complex surrounding the transcription startpoint, and they determine the site of initiation ; this complex constitute **the basal transcription apparatus**.
- The **upstream factors**, which are DNA-binding proteins that recognize specific short consensus elements located upstream the transcription startpoint (e.g. Sp1, which binds the GC box). These factors are ubiquitous

and act upon any promoter that contain the appropriate binding site on DNA. They increase the efficiency of initiation.

- The **inducible factors**, which function in the same general way as the upstream factors, but have a regulatory role. They are synthesized or activated at specific times and in specific tissues. The sequences that they bind are called response elements.

II Initiation of transcription

RNA pol II enzyme cannot initiate transcription itself, but is absolutely dependent on auxiliary transcription factors (called TFII_X, where "X" is a letter that identifies the individual factor). The enzyme together with these factors constitutes the basal (or minimal) transcriptional apparatus that is needed to transcribe any class II promoter. The efficiency and specificity with which a promoter is recognized depend upon short sequences, farther upstream the TATA box, which are recognized by upstream and inducible factors. Examples of these sequences are the CAAT box, which plays a strong role in determining the efficiency of the promoter, and is recognized in different promoters by different factors, such as factors of the CTF family, the factors CP1 and CP2, and the factors C/EBP and ACF, and the GC box, which is recognized by the factor Sp1. These factors have the ability to interact with one another by protein-protein interactions. The main purpose of the elements is to bring the factors they bind into the vicinity of the initiation complex, where protein-protein interactions determine the efficiency of the initiation reaction.

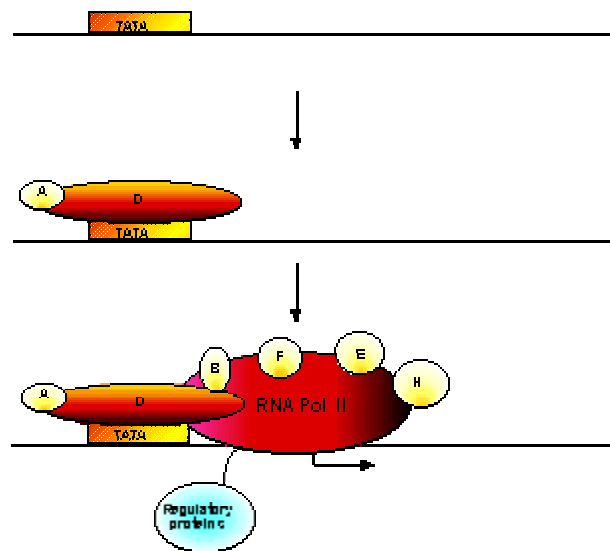


Figure 1:
Schematic model
for the assembly of
the basal
transcriptional
apparatus

III Transcription factors family

Common types of motifs that are responsible for binding to DNA can be found in different transcription factors. There are several groups of proteins that regulate transcription by using particular motifs to bind DNA :

III.1 Helix-turn-helix proteins

The helix-turn-helix motif was originally identified as the DNA-binding domain of phage repressors; one α -helix lies in the wide groove of DNA, the other lies at an angle across DNA. A related form of the motif is present in the homeodomain, a sequence first characterized in several proteins encoded by genes concerned with developmental regulation in *Drosophila* ; it is also present in genes coding for mammalian transcription factors. The homeobox is a sequence that codes for a domain of 60 amino-acids. The homeodomain is responsible for binding to DNA; the specificity of DNA recognition lies within the homeodomain. Its C-ter region shows homology with the helix-turn-helix motif of procaryotic repressors.

III.2 Zinc finger proteins

The zinc-finger motif comprises a DNA-binding domain. It was originally found in the factor TFIIIA, which is required for RNA PolIII to transcribe 5S rRNA genes. These proteins take their name from their structure, in which a small group of conserved aminoacids binds a zinc ion. Two types of DNA-binding proteins have structures of this type: the classic " zinc finger " proteins, and the steroid receptors.

A " **finger protein** " typically has a series of zinc fingers; the consensus sequence of a single finger is:



The motif takes its name from the loop of aminoacids that protrudes from the zinc-binding site and is described as the Cys₂/His₂ finger.

The fingers are usually organized as a single series of tandem repeats ; the stretch of fingers ranges from 9 repeats that occupy almost the entire protein (as in TFIIIA), to providing just one small domain consisting on 2 fingers ; the general transcription factor Sp1 has a DNA-binding domain that consists of 3 zinc fingers. The C-terminal part of each finger forms α -helices that bind DNA ; the N-terminal part form β -sheets. The non-conserved aminoacids in the C-terminal side of each finger are responsible for recognizing specific target sites.

Steroid receptors, which are activated by binding a particular steroid (e.g. glucocorticoids, thyroid hormone, retinoic acid), and some other proteins, have another type of finger. The structure is based on a sequence with the zinc-binding consensus :



These are called Cys₂/Cys₂ fingers. Proteins with Cys₂/Cys₂ fingers often have non-repetitive fingers, in contrast with the tandem repetition of the Cys₂/His₂ type. Binding sites on DNA are usually short and palindromic. The glucocorticoid and estrogen receptors each have 2 fingers, that form α -helices that fold together to form a large globular domain.

III. 3 Leucine zipper proteins

The leucine zipper is a stretch of aminoacids rich in leucine residues that provide a dimerization motif. Dimerization allows the juxtaposition of the DNA-binding regions of each subunit. A leucine zipper forms an amphipathic helix in which the leucines of the zipper on one protein could protrude from the α -helix and interdigitate with the leucines of the zipper of another protein in parallel to form a coiled coil domain. The region adjacent to the leucine repeats is highly basic in each of the zipper proteins, and could comprise a DNA-binding site. The 2 leucine zippers in effect form a Y-shaped structure, in which the zippers comprise the stem, and the 2 basic regions bifurcate simmetrically to form the arms that bind to DNA. This is known as the bZIP structural motif. It explains why the target sequences for such proteins are inverted repeats with no separation. Zippers may be used to sponsor the formation of homodimers or heterodimers. There are 4 repeats in the protein C/EBP (a factor that binds as a dimer to both the CAAT box and the SV40 core enhancer), and 5 repeats in the factors and (which form the heterodimeric transcription factor AP1).

III.4 Helix-loop-helix proteins

The amphipathic helix-loop-helix (HLH) motif has been identified in some developmental regulators and in genes coding for eukaryotic DNA-binding proteins. The proteins that have this motif have both the ability to bind DNA and to dimerize. They share a common type of sequence motif: a stretch of 40-50 aminoacids contains 2 amphipathic α -helices separated by a linker region (the loop) of varying length. The proteins in this group form both homodimers and heterodimers by means of interactions between the hydrophobic residues on the corresponding faces of the 2 helices. The ability to form dimers resides with these amphipathic helices, and is common to all HLH proteins.

Most HLH proteins contain a region adjacent to the HLH motif itself that is highly basic, and which is needed for binding to DNA. Members of the group with such a region are called bHLH proteins. A dimer in which both subunits have the basic region can bind to DNA. The bHLH proteins fall into 2 general groups. Class A consists of proteins that are ubiquitously expressed, including mammalian E12/E47. Class B consists of proteins that are expressed in a tissue-specific manner, including mammalian MyoD, Myf5, myogenin and MRF4 (a group of transcription factors that are involved in myogenesis, called myogenic regulatory factors, MRFs). A common *modus operandi* for a tissue-specific bHLH protein may be to form a heterodimer with a ubiquitous partner. There is also a group of gene products that specify development of the nervous system in *Drosophila melanogaster* (where Ac-S is the tissue-specific component, and da is the ubiquitous component). The proteins form a separate class of bHLH proteins.

Contributor(s)

Written 03-2004 Valentina GUASCONI, Hakima YAHYI, Slimane AIT-SI-ALI

Citation

This paper should be referenced as such :

Guasconi V, Yahy H, Ait-Si-Ali S . Transcription factors. Atlas Genet Cytogenet Oncol Haematol. March 2004 .

URL : <http://AtlasGeneticsOncology.org/Deep/TranscripFactorsID20043.html>

