ELSEVIER

FEMS Microbiology Reviews 29 (2005) 231–262

FEMS MICROBIOLOGY Reviews

www.fems-microbiology.org

The many faces of the helix-turn-helix domain: Transcription regulation and beyond ☆

L. Aravind *, Vivek Anantharaman, Santhanam Balaji, M. Mohan Babu, Lakshminarayan M. Iyer

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, Room 5N50, Bethesda, MD 20894, USA

Received 20 November 2004; received in revised form 22 December 2004; accepted 23 December 2004

First published online 28 January 2005

Abstract

The helix-turn-helix (HTH) domain is a common denominator in basal and specific transcription factors from the three superkingdoms of life. At its core, the domain comprises of an open tri-helical bundle, which typically binds DNA with the 3rd helix. Drawing on the wealth of data that has accumulated over two decades since the discovery of the domain, we present an overview of the natural history of the HTH domain from the viewpoint of structural analysis and comparative genomics. In structural terms, the HTH domains have developed several elaborations on the basic 3-helical core, such as the tetra-helical bundle, the winged-helix and the ribbon-helix–helix type configurations. In functional terms, the HTH domains are present in the most prevalent transcription factors of all prokaryotic genomes and some eukaryotic genomes. They have been recruited to a wide range of functions beyond transcription regulation, which include DNA repair and replication, RNA metabolism and protein–protein interactions in diverse signaling contexts. Beyond their basic role in mediating macromolecular interactions, the HTH domains have also been incorporated into the catalytic domains of diverse enzymes. We discuss the general domain architectural themes that have arisen amongst the HTH domains as a result of their recruitment to these diverse functions. We present a natural classification, higher-order relationships and phyletic pattern analysis of all the major families of HTH domains. This reconstruction suggests that there were at least 6–11 different HTH domains in the last universal common ancestor of all life forms, which covered much of the structural diversity and part of the functional versatility of the extant representatives of this domain. In prokaryotes the total number of HTH domains per genome shows a strong power-equation type scaling with the gene number per genome. However, the HTH domains in two-component signaling pathways show a linear scaling with gene number, in contrast to the non-linear scaling of HTH domains in single-component systems and sigma factors. These observations point to distinct evolutionary forces in the emergence of different signaling systems with HTH transcription factors. The archaea and bacteria share a number of ancient families of specific HTH transcription factors. However, they do not share any orthologous HTH proteins in the basal transcription apparatus. This differential relationship of their basal and specific transcriptional machinery poses an apparent conundrum regarding the origins of their transcription apparatus.

Keywords: Helix-turn-helix; DNA-binding; Transcription regulation; Evolution; Two-component systems; Structure

☆ Edited by Mark J. Pallen.

* Corresponding author. Tel.: +1 301 594 2445; fax: +1 301 435 7794.
E-mail address: aravind@ncbi.nlm.nih.gov (L Aravind).

0168-6445/$22.00
doi:10.1016/j.femsre.2004.12.008

Contents

1. Introduction

The general paradigms for the processes of transcription initiation and regulation first emerged from the pioneering studies on gene expression in bacteria and phages [1]. Transcription in bacteria was found to be catalyzed by a single multi-subunit RNA polymerase, which is recruited to promoters by means of a DNA-binding protein, the sigma factor that recognizes specific sequences upstream of genes [2–4]. The sigma factor and the RNA polymerase, together, constitute the basal transcription apparatus that is required for the baseline transcription of all genes. Early studies, especially in the *Bacillus subtilis* sporulation model, suggested that there may be more than one sigma factor that might recruit the catalytic core of the RNA polymerase to alternative sets of genes. This provided a mechanism for regulating the broad changes in gene expression, which correlate with the different developmental or differentiation states of a bacterium [5,6]. A number of early studies on metabolic regulation in bacteria also indicated that there are other regulatory DNA-binding proteins that act as switches to control the expression of specific smaller sets of genes. These sets of genes are often collinear on the chromosome, and encode components of a common pathway for the utilization of a particular environmental metabolite (e.g. lactose), or constitute interacting components of a developmental pathway (e.g. lytic or lysogenic development of phages). These regulatory proteins, termed the specific transcription factors, were found to belong to two distinct functional types: (1) those which negatively regulated transcription of their target gene (repressors) and (2) those which positively regulated transcription of their target genes (activators) [1]. The affinities of the specific transcription factors for their target sequences on DNA were often found to be dependent on their interaction with low-molecular weight compounds (effectors), which bound to them, or phosphorylation and other post-transcriptional modifications [1]. When transcription in eukaryotes was first investigated, several differences in the subunit composition and architecture of the basal transcription machinery and specific transcription factors were noted [7,8]. However, the basic regulatory mechanisms in bacterial transcription, which were brought to light in early studies, remained applicable in a generic sense across the entire Tree of Life [1].

The pioneering investigations of Matthews, Ohlendorf, Sauer, Doolittle and co-workers in 1982 provided the first glimpse of the features unifying diverse transcription regulators [9–13]. They showed that the phage lambda transcription regulators, cro and the cI repressor, and lacI, the lactose operon repressor, shared a similar tri-helical DNA-binding domain. The 2nd and the 3rd helices of this tri-helical domain constituted a Helix-Turn-Helix motif, and this motif was shown to be a critical determinant of their interaction with DNA. Thus, these DNA-binding domains came to be