Available online at www.sciencedirect.com
ScienceDirect
Plant Physiology and Biochemistry 46 (2008) 394–401
Plant Physiology and Biochemistry
www.elsevier.com/locate/plaphy

Review

# Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains

Kazuhiko Yamasaki [a,b,*], Takanori Kigawa [b,c], Makoto Inoue [b], Satoru Watanabe [b], Masaru Tateno [a,d], Motoaki Seki [e], Kazuo Shinozaki [f], Shigeyuki Yokoyama [b,g]

[a] Age Dimension Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba 305-8566, Japan
[b] Protein Research Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Yokohama 230-0045, Japan
[c] Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, Japan
[d] Institute of Physics and Center for Computational Sciences, Tsukuba University, Tsukuba 305-8571, Japan
[e] Plant Functional Genomics Research Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
[f] RIKEN Plant Science Center (PSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
[g] Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Received 26 September 2007
Available online 3 January 2008

**Abstract**

Plant-specific transcription factors are classified according to DNA-binding domains (DBDs) that were believed to be distinct from those of prokaryotes or other lineages of eukaryotes. Recently, structures of the DBDs including WRKY, NAC, B3, and SBP, which comprise major families of transcription factors, were determined by NMR spectroscopy or X-ray crystallography. In this review, we summarize the recent progress of structural biology in this field, especially on their DNA-binding mechanism and structural similarity to DBDs from other kingdoms. Unexpected structural relationships, together with recent identifications of homologous sequences in a variety of genomes, indicated that majority of the ''plant-specific'' DBDs originated from non-plant species, and that they largely expanded along with the evolution of higher plants.


*Keywords:* B3; Crystallography; DNA-binding domain; Evolution; NAC; NMR; SBP; Structure; Transcription factor; WRKY

## 1. Introduction

Although plants share fundamental mechanisms of life with other eukaryotic kingdoms, they acquired indeed different systems and organs after their arrival on land 0.4–0.5 billion years ago. Relating to their immobility, they possess distinct systems of adaptation to the environment, defense against pathogens, growth of specific organs such as flowers, etc. Transcription factors are involved in the control of all such plant-specific reactions, and, very interestingly, many of them showed no obvious sequence similarity to those of other eukaryotes or bacteria. In ~1500 probable specific transcription factors identified in the *Arabidopsis thaliana* genome, ~45% are classified into those plant-specific transcription factors [1]. It should be noted that these transcription factors were identified and classified according to their DNA-binding domains (DBDs) and that, therefore, the names of the DBDs, e.g., AP2/ERF (or EREBP), WRKY, NAC, are also used as the names of the transcription factor families (Table 1).

In order to elucidate their functions and evolutionary relationships, structures of the DBDs should be revealed. However, until recently, their structures were unknown, except for the AP2/ERF domain of *Arabidopsis* AtERF1 [2]. In recent years, our group and others have revealed structures of DBDs of other plant-specific transcription factor families [3–10].

* Corresponding author. Age Dimension Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba 305-8566, Japan.
*E-mail address:* k-yamasaki@aist.go.jp (K. Yamasaki).

0981-9428/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.plaphy.2007.12.015

Table 1
Representative DBDs of plant-specific transcription factor families[a]

| DBD | AP2/ERF | NAC | WRKY | B3 | SBP |
|---|---|---|---|---|---|
| Transcription factor family | AP2, ERF, RAV[c] | NAC | WRKY | ARF, ABI3/VP1, RAV[c] | SBP |
| Number in *Arabidopsis* genome[b] | 144 | 109 | 72 | 43 | 16 |
| Protein (PDB ID/method) | AtERF1 (1gcc, 2gcc, 3gcc/NMR) | NAC1 (1ut4, 1ut7/X-ray) | AtWRKY4 (1wj2/NMR), AtWRKY1 (2ayd/X-ray) | RAV1 (1wid/NMR), At1g16640 (1yel/NMR) | AtSPL4 (1ul4/NMR), AtSPL7 (1ul5/NMR), AtSPL12 (1wj0/NMR) |
| DNA-binding interface[d] | β-sheet | β-sheet | β-sheet | β-sheet | α-helix |
| Reference | [2] | [3] | [7,10] | [4,6] | [5,9] |

[a] From 11 large families of plant-specific transcription factors classified by Riechman et al. [1], i.e., AP2/ERF, NAC, WRKY, GARP, GRAS, Trihelix, TCP, ARF, SBP, Nin-like, ABI3/VP1, those possessing domains with revealed DNA-binding functions were picked up, and those possessing Myb-related or helix-loop-helix DNA-binding domains were excluded.
[b] According to Riechman et al. [1].
[c] RAV-like transcription factors contain both an AP2/ERF domain and a B3 domain.
[d] Secondary structure elements involved in the direct contact with DNA bases, which were observed in the complex structure or predicted based on experiments.

Some of these structures showed unexpected similarity to DBDs originating from prokaryotes or other eukaryotic lineages. In this review, we summarize the progress in this field, and, together with other findings, discuss the evolutionary origins of the plant-specific DBDs.

## 2. WRKY DNA-binding domain

Approximately 70 WRKY transcription factors are identified from the *Arabidopsis* genome (Table 1) [1,11]. The representative functions revealed for WRKY proteins are related to pathogen responses [12,13], although many other functions such as senescence [14], morphogenesis [15], cold tolerance [16], and responses to drought and high salinity stresses [17] are increasingly identified.

The WRKY DNA-binding domain is ~60 amino acids in length and contains a conserved WRKYGQK sequence, from which the domain was termed, and a $CX_{4-5}CX_{22-23}HXH$ zinc-binding motif. A majority of the WRKY proteins are classified in group I and contain two WRKY domains, while group II or group III WRKY proteins possess a single WRKY domain [11]. For the group I WRKY proteins, the C-terminal WRKY domain, but not the N-terminal one, is responsible for the DNA-binding activity [13,18,19]. The target recognition sequence of the WRKY domains, termed W-box sequence, is (T)TTGACY, where Y is C or T [12,13,19].

As the first structure of the WRKY domain, the NMR solution structure of the C-terminal WRKY domain of *Arabidopsis* WRKY4 was determined by our group (Fig. 1A) [7]. The structure consists of four β-strands forming an anti-parallel β-sheet, where the most N-terminal β-strand includes the conserved WRKYGQK sequence. At one of the ends of the β-sheet, a zinc-binding pocket is formed by the conserved Cys/His residues. It should be noted that the N-terminal strand is largely kinked in the middle by an ''insertion'' of the Gly residue of the motif. Namely, without this residue, the β-strand forms hydrogen-bonding connections with the adjacent strand in a manner typical of an antiparallel β-sheet.

A recent crystal structure of the C-terminal WRKY domain of *Arabidopsis* WRKY1 consists of five β-strands [10], adding an N-terminal strand to the four-stranded solution structure (Fig. 1B). The difference may be caused by the difference in the peptide length, in that the peptide used for the NMR analysis starts at the middle of this additional strand. However, this region is not well conserved among the WRKY domains, and the four-stranded structure is stable in solution. Therefore, the common WRKY structure is likely to be a four-stranded core, although it may possess an additional N-terminal strand.

In order to reveal the residues involved in the DNA binding, an NMR titration experiment was performed, where changes in NMR spectra upon adding a W-box DNA oligomer were analyzed [7]. Based on the result, a computational docking model of the complex of the WRKY domain and DNA was produced (Fig. 1C) [7]. It is clearly seen in the model that the most N-terminal β-strand that contains the WRKYGQK sequence enters the major groove of the DNA in such a way that the β-sheet plane is nearly perpendicular to the DNA axis. The Arg, two Lys, Tyr, and Gln residues in the motif are shown to be the most important residues in the sequence-specific recognition. It should be noted that the concave curvature of the β-strand induced by the Gly residue of the motif allows the deep entrance of this strand into the DNA groove.

The mutational experiments showed that the replacement of Trp, Arg, two Lys, Tyr, and Gly residues of the WRKYGQK motif results in a decrease in DNA-binding activity [10,20]. The Trp residue shows extensive hydrophobic interactions with other residues and is of particular importance in stabilizing the structure of the β-sheet. The Arg and Lys are typical DNA-contacting residues and make it possible to directly bind to the DNA, according to the docking model of the complex. Also, the Tyr residue is in a location that enables contact with the DNA in the complex model. For the Gly residue, the mutational effect of a bulky side chain, i.e., Phe [10], appears to be much larger than that of Ala [20], showing that the lack of side-chains is important in this position in order to avoid steric hindrance. Therefore, the model is largely consistent with the mutational experiments. However, to understand the mechanism of the specific recognition of the W-box sequence, the structure of the complex with DNA should be determined.

As a zinc-binding motif, folding of the WRKY domain is quite distinct from others, except for the *Drosophila* GCM domain (Fig. 1D) [21]. The GCM domain, identified from insects