J Korean Soc Appl Biol Chem (2012) 55, 819–822
DOI 10.1007/s13765-012-2186-2

SHORT COMMUNICATION

Quadruple 9-mer-based Protein Binding Microarray Analysis of the Arabidopsis Transcription Factor AtMYB77

Nam Iee Oh · Yeon-Ki Kim · Baek Hie Nahm · Jong-Joo Cheong

Received: 24 August 2012 / Accepted: 10 September 2012 / Published Online: 31 December 2012
© The Korean Society for Applied Biological Chemistry and Springer 2012

Abstract Binding sequence specificity of the Arabidopsis transcription factor AtMYB77 was determined by the quadruple 9-mer-based protein binding microarray (Q9-PBM) analysis. The position weight matrix and Wilcoxon-Mann-Whitney test with total 1843 clustered signals revealed that the full size AtMYB77 protein binds specifically to the consensus sequence pAACnG, where p represents T or C; n, A, G, C or T. This sequence is known as the type I binding site for MYB transcription factors. This result indicates that functional diversity among the type I MYB transcription factors is not due to their binding specificity on a particular promoter sequence.

Keywords *Arabidopsis* · AtMYB77 · conserved DNA sequence · MYB transcription factor · protein binding microarray

A transcription factor binds to a specific DNA sequence in promoter regions of target genes to regulate RNA polymerase activity for the gene transcription. MYB transcription factors contain a conserved DNA-binding domain, which consists of two or three (R1, R2, and R3) imperfect repeats of 50–53-amino acid (Rosinsky and Atchley, 1998). A total of 126 R2R3 MYB genes have been identified in the Arabidopsis genome (Yanhui et al., 2006). R2R3 MYB protein-binding sites have been classified into three types: I, CnGTTr (= pAACnG); II, GkTwGTTr; and IIG, GkTwGGTr, where n indicates A, G, C or T; K, G or T; R, A or G; P, T or C; W, A or T (Romero et al., 1998). Although members of a class of MYB transcription factors share high degree of homology in R2R3 DNA-binding domain, individual proteins play unique roles in diverse plant processes.

N. I. Oh · J.-J. Cheong (✉)
Center for Food and Bioconvergence, Seoul National University, Seoul 151-921, Republic of Korea
E-mail: cheongjj@snu.ac.kr

Y.-K. Kim · B. H. Nahm
Genomics Genetics Institute, GreenGene Biotech Inc., Yongin 449-728, Republic of Korea

B. H. Nahm
Division of Bioscience and Bioinformatics, Myongji University, Yongin 449-728, Republic of Korea

The Arabidopsis AtMYB77 (AGI number AT3G50060) belongs to the subgroup 22 R2R3-type MYB transcription factors, together with AtMYB44, AtMYB70, and AtMYB73 (Kranz et al., 1998; Stracke et al., 2001). These proteins share two conserved motifs (Fig. 1A): TGLYMSPxSP (motif 22.1) and GxFMxVVQEMIxx EVRSYM (motif 22.1) (Stracke et al., 2001). AtMYB77 was reported to mediate auxin responses to control lateral root growth and development under changing environmental conditions (Shin et al., 2007).

We determined the DNA-binding sequence specificity of AtMYB77 by a comprehensive genome-wide method. The quadruple 9-mer-based protein binding microarray (Q9-PBM) we used was designed to contain quadruples of all possible 9-mer combinations of oligonucleotides, permitting unequivocal interpretation of DNA sequences bound with a protein (Kim et al., 2009). The full-length cDNAs of *AtMYB77* (TAIR clone G14459) was N-terminal fused with DsRed fluorescent protein gene and a polyhistidine-tag (Fig. 1B), and expressed in *E. coli* strain BL21-ColonPLus. Purified AtMYB44-DsRed-polyHis fusion proteins (200 nM) was incubated with the Q9-PBM at 25°C for 1 h. Fluorescence images were obtained with a 4000B microarray scanner (Molecular Devices, USA). The consensus binding sequence was determined based on signal strength, as described by Kim et al. (2012).

The rank-ordered signal distribution curve showed a steep leftward slope and an extended right tail region (Fig. 2). Two independent linear models, y=ax+b, were applied to the left steep (b1=50455.4, slope = –27.0) and right tail (b1=674.1, slope =

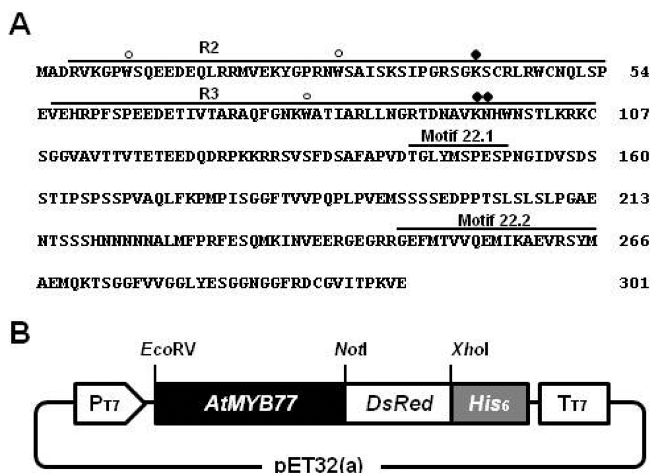

Fig. 1 DNA construct for the Q9-PBM (protein-binding microarray). (A) Amino acid sequence of the AtMYB77 (GenBank accession no. AY124828). Sequences under the lines represent amino acid residues that are conserved in the *Arabidopsis* subgroup 22 MYB transcription factors, R2, R3, motif 22.1, and motif 22.2. Tryptophan residues conserved in typical R2R3 MYB transcription factors are indicated by open dots. Two lysine and one asparagine residues essential for specific DNA binding are indicated by closed square. (B) DNA construct for expression of AtMYB77-DsRed-polyHis fusion protein in *Escherichia coli* strain BL21-ColonPLus.

–0.0027) regions. For motif extraction, 1843 total signals in the steep left region were clustered. These groups were denoted with SEQLOGO 'Visualize information content of patterns' [http://www.bioinf.ebc.ee/EP/EP/SEQLOGO/]. Based on the best alignment using the highest-ranked 9-mers among the clusters as seeds, the position weight matrix (PWM) of the cluster was obtained (Fig. 3A and 3B). The Wilcoxon-Mann-Whitney test was performed to yield a consensus sequence, 'pAACnG', where p represents T or C and n represents A, G, C or T. *P*-value calculated from the test was 0 (p =0).

It appears that T prefers to C as the first nucleotide in the consensus sequence. When considering the rank of signal strength (affinity), the best eight oligomers contain TAACnG (Table 1). The last G is sometimes replaced by A without significant weakening of the binding strength. The PBM analysis revealed that AtMYB77 binds specifically to the type I site for R2R3 MYB transcription factors.

The binding site selection assay using a pool of synthetic oligonucleotides with degenerate sequences performed previously by Romero et al. (1998) suggested that the AtMYB77-binding consensus sequence was pAACpGpC (= GrCrGTTr, where r represents A or G). The last C appeared with 82% probability at this position. In the experiment, a truncated form of the AtMYB77 protein containing the R2R3 binding domain was used, because the full-size protein AtMYB77 exhibited binding affinity too low to be analyzed. In the present study, when we used a full size AtMYB77, a significant appearance such the last C was not observed at this nucleotide position (Fig. 3).

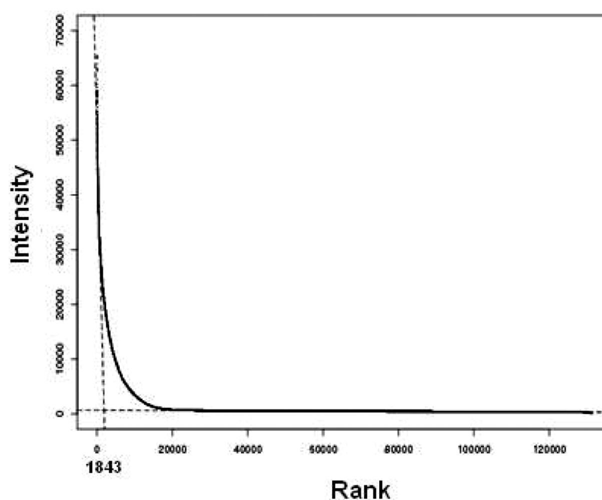

Fig. 2 Rank analysis of AtMyb77 PBM binding. From the rank-ordered signal distribution, two independent linear models, y=ax+b, were applied in the steep (b1=50455.4, slope= –27.0) and the heavy right (b1=674.1, slope = –0.0027) tail regions. Rank extrapolated for motif extraction was 1843.

A

	1	2	3	4	5	6	7	8	9
A	547	167	1041	989	1	21	120	352	404
C	234	262	4	62	1058	297	30	270	439
G	155	10	7	5	0	291	891	45	38
T	123	620	7	3	0	450	18	392	178
	n	T	A	A	C	n	G	n	n

B

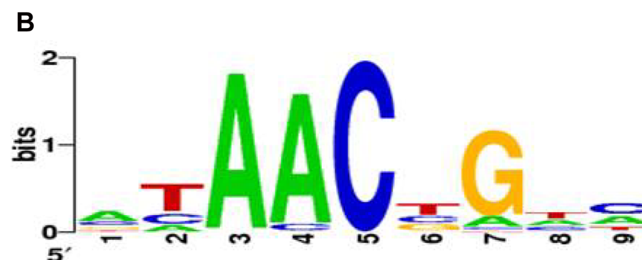

Fig. 3 Schematic representation of protein binding microarray analysis. (A) Total position weight matrix. Spots that exhibited strong intensity and high enrichment were clustered. Based on the best alignment using the highest-ranked 9-mers among the clusters as seeds, position weight matrix (PWM) of the cluster was obtained. (B) Sequence logo for the determined consensus binding sequence of AtMYB77.

It has been demonstrated that the first A, third C, and fifth G in the type I MYB-binding consensus sequence AACTG are involved in very specific interactions with corresponding MYB proteins (Tanikawa et al., 1993). In the MYB proteins, both R2 and R3 contain three helices, and the third helix in each constitutes a recognition helix. Heteronuclear multidimensional NMR analysis demonstrated the solution structure of a complex between R2R3 domain and AACTG oligomer (Ogata et al., 1994). R2 and R3 are

Table 1 The best top 50 nine-mer oligonucleotides exhibiting strong affinity to AtMYB77[1)]

Rank	Oligomer	P-value (Median)	Median Intensity	Max Intensity
1	**taactg**tca	$1.325e^{-198}$	49,113.0	65,535.0
2	**taacgg**taa	$1.314e^{-194}$	47,048.0	59,055.0
3	**taacgg**tca	$3.773e^{-192}$	45,811.0	60,770.0
4	**taactg**taa	$1.462e^{-185}$	42,617.0	53,007.0
5	**taaccg**tca	$1.792e^{-176}$	38,490.0	55,541.0
6	**taaccg**taa	4.357e-^{175}	37,888.0	52,508.0
7	**taacgg**tta	$3.313e^{-174}$	37,509.0	47,820.0
8	**taactg**cca	$7.964e^{-174}$	37,346.0	65,535.0
9	**taacta**cca	$4.054e^{-173}$	37,045.0	57,254.0
10	**caaccg**tca	$3.842e^{-172}$	36,632.0	65,535.0
11	aa**aacgg**ta	$6.517e^{-171}$	36,117.0	49,088.0
12	**caactg**tca	$9.636e^{-170}$	35,632.0	65,535.0
13	**aaccg**aaat	$6.329e^{-169}$	35,296.0	42,811.0
14	at**aacgg**cc	$2.352e^{-168}$	35,063.0	50,648.0
15	**taactg**aaa	$6.818e^{-168}$	34,875.0	65,535.0
16	**taaccg**tcg	$5.448e^{-167}$	34,510.0	48,843.0
17	**taaccg**aat	$1.886e^{-165}$	33,894.0	51,113.0
18	aaat**aacgg**	$1.897e^{-165}$	33,893.0	46,309.0
19	**taaccg**cca	$6.057e^{-164}$	33,299.0	49,150.0
20	t**taactg**aa	$7.337e^{-163}$	32,876.0	57,313.0
21	**taacta**ccc	$1.684e^{-162}$	32,736.0	55,630.0
22	**taactg**ccg	$2.809e^{-162}$	32,650.0	50,553.0
23	**taacgg**tcg	$1.560e^{-161}$	32,363.0	64,243.0
24	**taactg**tcg	$6.828e^{-160}$	31,737.0	55,245.0
25	a**aaccg**tta	$4.500e^{-159}$	31,428.0	43,637.0
26	aa**aacgg**tt	$4.527e^{-159}$	31,427.0	46,531.0
27	**taaccg**ccg	$1.474e^{-157}$	30,862.0	56,318.0
28	**caaccg**tcc	$7.722e^{-156}$	30,229.0	53,075.0
29	**taacta**aca	$1.626e^{-155}$	30,111.0	50,854.0
30	aa**aactg**tc	$2.407e^{-155}$	30,049.0	59,877.0
31	**taaccg**cta	$1.063e^{-154}$	29,815.0	52,481.0
32	**caacgg**taa	$9.518e^{-154}$	29,472.0	65,535.0
33	at**accgg**ta	$1.364e^{-153}$	29,416.0	36,018.0
34	aa**aaccg**tc	$2.154e^{-152}$	28,989.0	54,196.0
35	**caacgg**tca	$2.168e^{-152}$	28,988.0	58,135.0
36	**taacga**cca	$2.225e^{-152}$	28,984.0	49,097.0
37	**taaccg**aca	$1.418e^{-151}$	28,700.0	48,681.0
38	**taaccg**ctc	$5.623e^{-151}$	28,490.0	49,559.0
39	**taactg**tcc	$1.462e^{-150}$	28,345.0	43,049.0
40	**taacgg**aat	$1.831e^{-150}$	28,311.0	53,317.0
41	**taacta**ccg	$1.905e^{-150}$	28,305.0	49,705.0
42	taaa**aactg**	$1.083e^{-149}$	28,043.0	40,686.0
43	**taactg**ccc	$1.135e^{-149}$	28,036.0	43,718.0
44	**taactg**aca	$1.785e^{-149}$	27,968.0	53,870.0
45	**taaccg**ctg	$8.146e^{-149}$	27,741.0	47,653.0
46	**taactg**aac	$1.117e^{-148}$	27,694.0	38,115.0
47	**taaccg**tcc	$2.217e^{-148}$	27,592.0	44,293.0
48	**aaccg**aact	$2.324e^{-148}$	27,585.0	41,545.0
49	**taacgg**tcc	$3.802e^{-148}$	27,512.0	39,836.0
50	ta**aacta**ac	$4.323e^{-148}$	27,493.0	51,462.0

[1)]The consensus sequence AACnG in the oligomers is indicated in bold. Nucleotides not fit to the consensus sequence are underlined.

closely packed in the major groove, and thus the two recognition helices contact with each other directly to cooperatively bind to the specific base sequence AACTG. The three key base pairs (the first A, third C, and fifth G) in this sequence are specifically recognized by an asparagine residue in R3 domain, a lysine in R3 domain, and another lysine in R2 domain. These amino acid residues are well conserved in the AtMYB77, at Asn-97 (R3), Lys-96 (R3), and Lys-42 (R2) (Fig. 1A).

Although our PBM data revealed nucleotide p (T or C) at the first position of the consensus sequence in addition to AACnG, significance of the additional nucleotide in the binding between the AtMYB77 and promoter region is yet unknown, because PBM and the binding site selection assay are *in vitro* binding assays with a pool of synthetic oligonucleotides. On the other hand, AtMYB77 appears to bind to promoters containing the sequence AATnG, without exerting any specificity in activating the target gene transcription.

In many cases, the functional specificity of transcription factors can be determined by their binding sequences. However, our data indicates that functional diversity among the type I MYB transcription factors is not due to their binding specificity on a particular promoter sequence. A group of MYB transcription factors were found to interact with other transcription factors to carry out their specific roles (Grotewold et al., 2000; Zimmermann et al., 2004; Quattrocchio et al., 2006). Shin et al. (2007) reported that C-terminus activation domain of AtMYB77 interacts with C terminus of ARFs auxin response factors (ARFs) *in vitro*, resulting in a strong reduction in lateral root numbers. Therefore, functional specificity of AtMYB77 appears to come from the interaction with other protein(s), and not from the specificity of binding sequence on the promoter regions of target genes.

Acknowledgments This work was supported by the Basic Science Research Program through the National Research Foundation (grant number 2012-0007030), by the Next Generation BioGreen 21 Research Program through the National Center for GM Crops of the Rural Development Administration (grant number PJ008073), and by the Technology Development Program for Life Industry through the Korea Institute of Planning and Evaluation for Technology of Food, Agriculture, Forestry and Fisheries (grant number 111076-5).

References

Grotewold E, Sainz MB, Tagliani L, Hernandez, JM, Bowen B, and Chandler VL (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc Natl Acad Sci USA* **97**, 13579–84.

Kim M-J, Chung PJ, Lee T-H, Kim T-H, Nahm BH, and Kim Y-K (2012) Convenient determination of protein-binding DNA sequences using quadruple 9-mer-based microarray and DsRed fusion protein. *BMC Mol Biol* **10**, 91.

Kim M-J, Lee T-H, Pahk Y-M, Kim Y-H, Park H-M, Choi YD et al. (2009) Quadruple 9-mer-based protein binding microarray with DsRed-monomer fusion protein. *Methods Mol Biol* **786**, 65–77.

Kranz HD, Denekamp M, Greco R, Jin H, Leyva A, Meissner RC et al. (1998) Towards functional characterization of the members of the *R2R3-*

MYB gene family from *Arabidopsis thaliana*. *Plant J* **16**, 263–76.

Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H et al. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* **79**, 639–48.

Quattrocchio F, Verweij W, Kroon A, Spelt C, Mol J, and Koes R (2006) PH4 of petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* **18**, 1274–91.

Romero I, Fuertes A, Benito MJ, Malpica JM, Leyva A, and Paz-Ares J (1998) More than 80 *R2R3-MYB* regulatory genes in the genome of *Arabidopsis thaliana*. *Plant J* **14**, 273–84.

Rosinsky JA and Atchley WR (1998) Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin. *J Mol Evol* **46**, 74–83.

Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ et al. (2007) The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction. *Plant Cell* **19**, 2440–53.

Stracke R, Werber M, and Weisshaar B (2001) The *R2R3-MYB* gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* **4**, 447–56.

Tanikawa J, Yasukawa T, Enari M, Ogata K, Nishimura Y, Ishii S et al. (1993) Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain. *Proc Natl Acad Sci USA* **90**, 9320–4.

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G et al. (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol* **60**, 107–24.

Zimmermann IM, Heim MA, Weisshaar B, and Uhrig JF (2004) Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J* **40**, 22–34.