Biochemical and Biophysical Research Communications 441 (2013) 262–265

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Creating a TALE protein with unbiased 5′-T binding

Shogo Tsuji, Shiroh Futaki, Miki Imanishi *

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

ARTICLE INFO

Article history:
Received 7 October 2013
Available online 19 October 2013

Keywords:
Transcriptional activator-like effector (TALE)
DNA binding
Directed evolution

ABSTRACT

Transcription activator-like effectors (TALEs) are convenient tools for genome engineering at specific genomic sites. However, their use is constrained because most TALE binding sites are preceded by a highly conserved 5′ terminal T nucleotide (5′-T). To remove the 5′-T constraint, we substituted tryptophan 232 in the repeat-1 loop region of the dHax3 N-terminal domain for other amino acids. Furthermore, we randomized four amino acid residues of the hairpin loop region of repeat-1. Although point mutation was insufficient to remove the 5′-T constraint, directed evolution from the randomized library yielded repeat-1 mutants with unbiased targeting sites for 5′-bases. Our result indicates that the repeat-1 loop region of dHax3 is important for 5′-base accommodation, and that molecular evolution of repeat-1 of TALEs is an efficient strategy to remove the 5′-T constraint and thus allow targeting of any DNA sequences.

1. Introduction

Designing sequence-specific DNA-binding proteins is important for the development of artificial nucleases, transcription factors, and DNA-modifying enzymes to achieve site-specific modifications at desired genomic loci. In addition to C2H2-type zinc finger domains, transcription activator-like effectors (TALEs) are receiving increasing attention in genome engineering [1–3]. TALEs are sequence-specific DNA binding proteins that are secreted by the bacterial pathogen *Xanthomonas*. Their DNA binding specificity is determined by a series of tandem repeats of typically 34 highly conserved amino acids. These repeats contain variable di-residues at positions 12 and 13, called RVDs, that define the base preference of a repeat [4,5]. Because of the simple one-to-one base recognition of each repeat, TALEs can be readily designed to target specific DNA sequences by simply modifying the RVDs.

Although TALEs have target sequence versatility that is determined by tandem repeat regions, almost all TALE binding sites are preceded by a highly conserved 5′-terminal T nucleotide (5′-T) [4,5]. The 5′-T is critical for the efficient DNA binding of TALEs [4,6], which restricts the sites that can be targeted by designed TALEs for genome editing, gene regulation, or synthetic biology. Therefore, to target any DNA sequences in the genome, it is necessary to remove the 5′-T constraint. The N-terminal domain immediately preceding the canonical repeat domain has been assumed to interact specifically with the 5′-T [1,4,7–9]. Recent structural data indicate that there are at least two (0 and −1) noncanonical repeats N-terminal to the canonical repeats, which form two α-helices with an adjoining loop, similar to the canonical repeats [9,10]. Although the mechanism of the binding of TALEs to a 5′-T remains unclear, one report suggested that the side chain of W232 of the repeat-1 hairpin loop of the PthXo1 TALE protein is positioned close to the 5′-T in the DNA complex [8,9]. In this study, we aimed to design a TALE backbone with the ability to target unbiased 5′-bases. The dHax3 TALE protein was used as a scaffold because dHax3 is widely investigated to design TALE-transcription factors or -nucleases [11–13]. We created point mutants of W232. Furthermore, we randomized four amino acid residues of the hairpin loop region of repeat-1. We then successfully removed the 5′-T constraint using a bacterial one-hybrid screening system.

2. Materials and methods

2.1. Chemicals

The modification and restriction enzymes were purchased from New England BioLabs, and the synthesized oligonucleotides were purchased from Operon Biotechnologies. All other chemicals were of commercial reagent grade.

2.2. Plasmid construction

The mammalian cell expression vector of dHax3$_{(N241/C183)}$, with a nuclear localization signal (NLS) and synthetic transcription activation domain (VP64) [14] driven by a CMV promoter, was constructed as follows. The coding region of a truncation mutant (N1-C0) of Hax3, the NLS and VP64 from pLenti-EF1a-Backbone(NG) (Addgene plasmid 27963) [13] was amplified by PCR using

* Corresponding author. Fax: +81 774 32 3038.
E-mail address: imiki@scl.kyoto-u.ac.jp (M. Imanishi).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.10.060

following primers: 5′-GCG GGA TCC GCC ACC ATG CCC GGG TCG CGG ACC CGG CTC CC-3′ and 5′-TCT GCC CTC TCC ACT GC-3′. N1 and C0 contained amino acids 48–288 and 678–860 of dHax3, respectively [12,13]. We named our designed TALE dHax3$_{(N241/C183)}$. The amplified fragment was inserted into the multiple cloning site of a CMV-driven vector based on pCMV-AD (Stratagene). Following the method described by Zhang et al. [13], 11.5 tandem repeats were assembled to correspond to the dHax3 RVDs [12], creating dHax3$_{(N241/C183)}$/pCMVt. Site-directed mutagenesis of W232 was performed by QuikChange using Phusion Hi-Fidelity DNA polymerase (New England Biolabs). Luciferase reporter vectors were constructed by inserting a dHax3 binding sequence, beginning 5′-A/C/G/T, into pGL3 (Promega).

To create *Escherichia coli* expression vectors, pET42b (Novagen) was digested using *Nde*I and *Hin*dIII, and a DNA fragment containing a His-tagged coding sequence followed by a multiple cloning site (dsDNA; 5′-T ATG CAC CAT CAC CAT CAC CAC CCC GGG GTA GGA GCT AGC TGA A-3′) was inserted. The coding region of dHax3$_{(N241/C183)}$ without NLS and VP64 was then ligated into this vector to create dHax3$_{(N241/C183)}$/pET42.

The omega-dHax3$_{(N241/C183)}$ expression plasmid, dHax3$_{(N241/C183)}$/pB1H2W5, was constructed as follows. The bacterial one-hybrid (B1H) expression vector pB1H2W5-Prd (Addgene plasmid 18039) [15] was digested by *Kpn*I and *Xba*I, and a DNA fragment containing a His-tag coding sequence with *Xma*I and *Nhe*I sites (5′-GGT ACC CAC CAT CAT CAC CAC CAT CCC GGG TGT CGC GCT AGC TCT AGA-3′) was inserted. The coding region of dHax3$_{(N241/C183)}$ was inserted into the prepared pB1H2W5 vector, so that the omega-subunit of RNA polymerase was fused to the N-terminus of dHax3$_{(N241/C183)}$. To create the B1H reporter plasmids, a dHax3 target sequence starting with a 5'-A or 5'-C (dsDNA; 5′-GGC GCA CCC TTT ATC TCT TGG TAT CTG AGG-3′ or 5′- GGC GCC CCC TTT ATC TCT TGG TAT CTG AGG-3′) was inserted into *Not*I and *Eco*RI sites of pH3U3-mcs (Addgene plasmid 12609) [16].

The sequences of the constructed plasmids were confirmed using the ABI PRISM 3130 genetic analyzer (Applied Biosystems). The amino acid sequences of all proteins are described in the Supporting information.

2.3. Repeat-1 library construction

*Bsa*I and *Asc*I sites were introduced in dHax3$_{(N241/C183)}$/pB1H2W5 by QuikChange. To generate the repeat-1 library of the region K230-S233, DNA fragments were PCR-amplified from dHax3$_{(N241/C183)}$/pCMVt using the primers F (5′-AAG CCA GGT CTC AGG CGA TCG TCG GTG TGG GGN NSN NSN NSN NSG GAG CCC GAG CGC TTG AG-3′; $S = C/G$) and R (5′-GTT GAG GGG CGC GCC CGT GAG CGC ATT G-3′), and ligated into *Bsa*I/*Asc*I sites of dHax3$_{(N241/C183)}$/pB1H2W5. Vectors were electroporated into NEB 10β Electrocompetent *E. coli* (New England BioLabs), which were plated on LB agar with ampicillin and incubated at 37 °C for 14 h. Approximately 5.0×10^6 transformants were then isolated and the plasmids were purified, yielding a library of dHax3$_{(N241/C183)}$/pB1H2W5.

2.4. Luciferase reporter assay

One hundred and fifty nanograms of the expression vector, 200 ng of the reporter vector, and 50 ng of the control vector (pRL-TK; Promega) were transiently co-transfected into HeLa cells using Lipofectamine LTX (Life Technologies). Cells were incubated for 24 h, and luciferase activity was measured using the dual luciferase reporter system (Promega). The luminescence was obtained by normalizing to the transfection control.

2.5. Expression and purification of TALE constructs

The dHax3$_{(N241/C183)}$ protein was expressed in *E. coli* BL21(DE3) cells by induction with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 12 h at 18 °C. Cells were resuspended in buffer A (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 2 mM sodium azide, pH 8.0). The soluble fraction was loaded onto HisTrap HP (GE Healthcare), and purified using buffers B (20 mM NaH_2PO_4, 500 mM NaCl, 5 mM imidazole, 2 mM sodium azide, pH 8.0) and C (20 mM NaH_2PO_4, 500 mM NaCl, 500 mM imidazole, 2 mM sodium azide, pH 8.0). Fractions were further purified using HiTrap Heparin HP (GE Healthcare) with buffers D (10 mM NaH_2PO_4, 2 mM sodium azide, pH 7.4) and E (10 mM NaH_2PO_4, 2 M NaCl, 2 mM sodium azide, pH 7.4). Finally, gel filtration was performed using Superdex75 (GE Healthcare) with buffer F [17] [480 mM KCl, 1.6 mM ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT), 12 mM Tris–HCl, pH 7.5]. The purity of dHax3$_{(N241/C183)}$ was confirmed by SDS–PAGE.

2.6. Electrophoretic mobility shift assay (EMSA)

EMSA was performed based on the method reported by Meckler et al. [17], except that direct detection of FITC-labeled DNA was carried out using Typhoon FLA 9500 (GE Healthcare). The equilibrium dissociation constant (K_d) of each protein–DNA fragment complex was evaluated, as described [18].

2.7. Bacterial one-hybrid screening

The library-dHax3$_{(N241/C183)}$/pB1H2W5 and pH3U3 reporter plasmid were co-transformed into the omega knockout bacteria hybrid selection strain USO hisB- pyrF- rpoZ- (Addgene 18049) [15], and screening was performed as described by Noyes et al. [19]. Cells were plated on NM media plates containing kanamycin (25 mg/ml), carbenicillin (100 mg/ml), 3-aminotriazole (2.5 or 5 mM), and IPTG (10 μM). They were then incubated at 37 °C for 48–72 h. Individual surviving colonies on 5 mM 3-aminotriazole plates were isolated, and the repeat-1 sequences were analyzed using a sequencing primer (5′-CAG GAA AAG ATC AAG CCC-3′) and displayed as a sequence logo [20].

3. Results and discussion

3.1. Requirement of 5′-T for the binding site of dHax3

To confirm the requirement of a 5′-T for dHax3$_{(N241/C183)}$, a luciferase reporter assay was performed using reporter vectors containing a dHax3 binding sequence beginning with 5′-T/A/G/C. The dHax3$_{(N241/C183)}$ protein fused with a NLS and an activation domain, VP64, was expressed in HeLa cells. As shown in Fig. 1A, the wild-type construct activated the reporter containing a 5′-T >10- to 300-fold more than reporters containing a 5′-A/G/C, suggesting that 5′-T is critical for DNA binding of the dHax3$_{(N241/C183)}$ TALE protein.

3.2. Point mutation of W232 of repeat-1 did not allow binding to non-5′-T

The crystal structure of the DNA-bound TAL effector PthXo1 indicates that the hairpin loop region of repeat-1 (amino acids 230–233, KQWS) faces the major groove of the DNA, and that W232 plays a role in the contact with 5′-T (Fig. 1B) [8,9]. To assess the possibility that W232 could be substituted in dHax3$_{(N241/C183)}$ to accommodate non-5′-T, W232 was mutated to all other 19 amino acids and their DNA binding ability was evaluated using

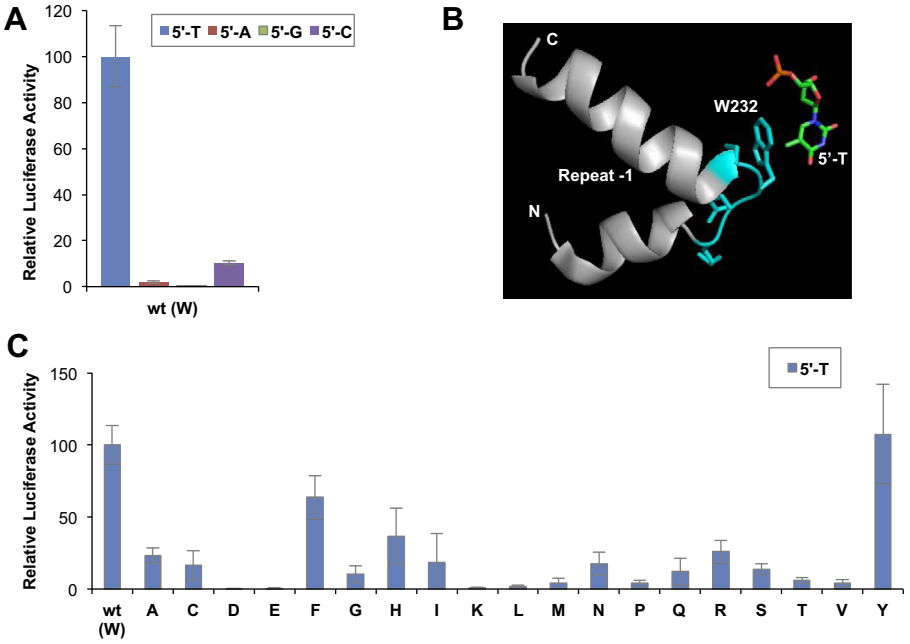

Fig. 1. The 5′-terminal base specificity of the wild type and W232 mutants. (A) Luciferase reporter activity of wild-type dHax3$_{(N241/C183)}$ in reporter vectors with binding sites beginning 5′-T/A/G/C (blue, red, green, and purple, respectively). (B) Recognition of a 5′-T by W232 in repeat-1 [9]. (C) Luciferase reporter activity of W232 mutants for a reporter vector with a dHax3 binding site beginning 5′-T. The reporter activities for sites beginning 5′-A/G/C were <2% of wild-type for 5′-T (see Fig. S2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

luciferase reporter assays. In the 5′-T reporter, all of the W232 mutant proteins except W232Y showed significantly decreased luciferase activity (Fig. 1C). This observation highlights the importance of W232 for the DNA binding of dHax3$_{(N241/C183)}$. The data obtained from EMSAs supported the results of the luciferase assay. Specifically, the wild type and W232Y showed comparable DNA binding affinities, whereas W232F and W232A showed significantly reduced affinities compared with the wild type (Fig. S1). This suggests that not simply aromatic rings, but the specific side-chains of tryptophan and tyrosine interact with target DNA containing 5′-T. Importantly, the luciferase activities of W232 mutants in luciferase reporters containing a 5′-A/G/C were <2% of wild-type for the reporter containing a 5′-T, meaning that none of the substituted mutants of W232 allowed binding to 5'-A/C/G (Figs. 1C and S2). These results suggest that a single mutation of W232 just decreased DNA binding ability at target sites with a 5′-T, and was insufficient to accommodate sites lacking 5′-T.

3.3. Molecular evolution of repeat-1 yielded TALE backbones with unconstrained target 5′-bases

To alter the recognition pattern of the 5′ terminal nucleotide, we hypothesized that it may be necessary to mutate the entire loop region of repeat-1 due to the limitations of point mutations of W232. Therefore, a randomized library of the region K230-S233 in repeat-1 of dHax3$_{(N241/C183)}$ was created (Figs. 1B and 2A). Bacterial one-hybrid screening [15,19] was performed using reporter vectors with a dHax3 binding site beginning with 5′-A or -C (Fig. 2A). We did not screen a vector with a site beginning with 5′-G because *Ralstonia* TALE protein was reported to target DNA sequences starting with 5′-G [21]. Fig. 2B shows the variant populations obtained after selection for 5′-A or 5′-C. Similar amino acid sequence patterns were obtained from both selections, where small amino acid residues (including Gly, Ala, Ser, and Pro) predominantly occupied the sites of K230 and W232. It is notable that the Q231R and S233A mutations were highly conserved in both selections.

To assess the DNA binding preference of the selected mutant proteins, the transcriptional activities of representative mutants (amino acids 230–233: SRGA, SRPA, ARGA, and ARPA) were examined by luciferase assays using four reporter vectors containing a dHax3 binding site beginning with a 5′-T/A/G/C (Fig. 2C). Compared with the wild type (amino acids 230–233: KQWS), all mutants showed transcriptional activities in the 5′-A/G/C reporters that were increased significantly. Consequently, all mutants accommodated any 5′ base, although the activation levels were decreased slightly for a 5′-T reporter. This suggests that these mutants acquired tolerance for any 5′ nucleotide, unlike the wild type (KQWS), which has a strong preference for a 5′-T. These data are consistent with the two independent selections for 5′-A and -C, which yielded similar mutation patterns.

These results indicate that the mutated loop region of repeat-1 nonspecifically interacts with the DNA phosphate backbone through R231, a residue that is conserved during directed evolution. To assess whether a point mutation of Q231R allows the accommodation of all 5′ bases, a loop mutant with a KRWS sequence was created. However, this mutant retained a strong preference for a 5′-T (Fig. S3). This suggests that the overall loop region determines the specificity of the interaction with DNA.

In this study, we successfully removed the 5′-T constraint of the dHax3 TALE protein by bacterial one-hybrid screening using a library of K230-S233, which is the hairpin loop region of repeat-1 of dHax3. Our results are consistent with a recent report by Lamb et al. [22], in which the directed evolution of the AvrXa7 TALE protein was performed using a TALE recombinase (TALE-Rs) screening system [23]. Our data strongly suggest that the hairpin loop region of repeat-1 of TALE proteins plays a role in DNA binding, particularly to the 5′-terminal nucleotide, and that directed evolution of the repeat-1 loop region is an efficient strategy to modify the DNA binding preference for 5′-T. Therefore, our results provide information useful for the design of TALE proteins that can target any DNA sequence for use in genome engineering and synthetic biology.

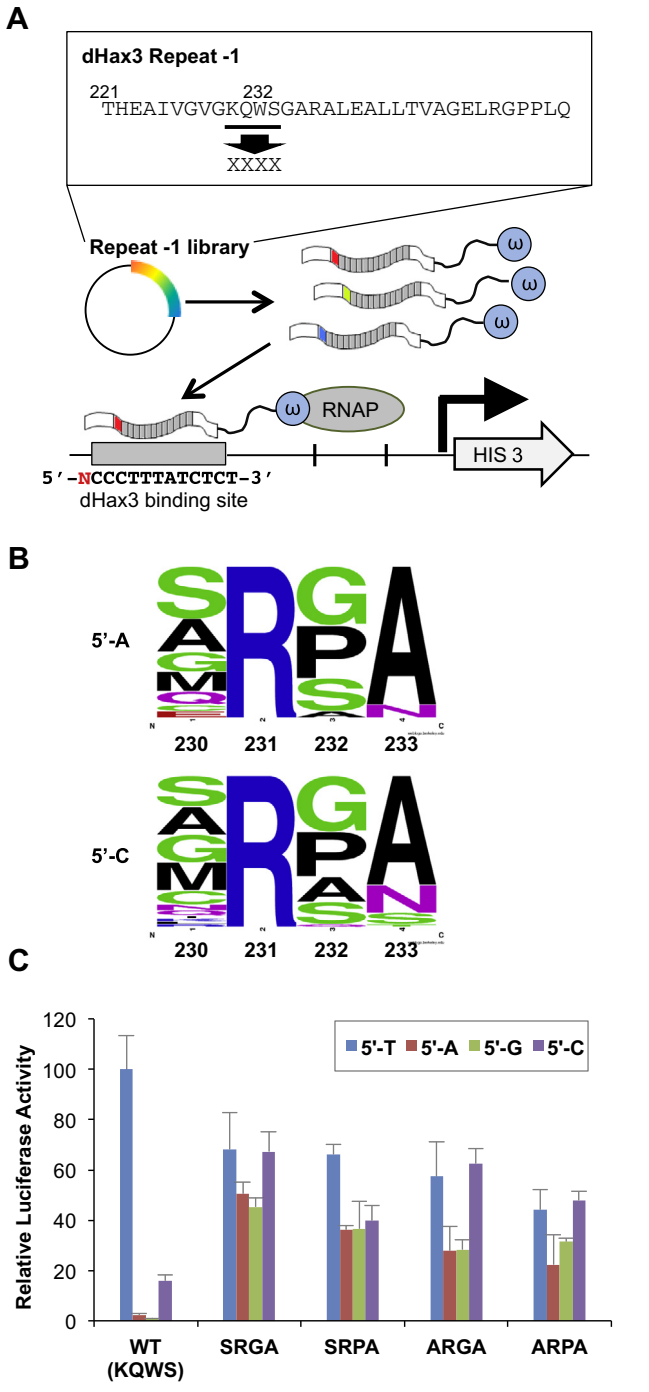

Fig. 2. Bacterial one-hybrid (B1H) screening of the repeat-1 loop region for unconstrained targeting sequences. (A) Schematic representation of the B1H system used in this study. Four amino acid residues, K230-S233, were randomized shown as XXXX. (B) Sequences obtained from the dHax3$_{(N241/C183)}$ library with a randomized loop region are displayed as a sequence logo [20]. The upper and lower panels show the results using the 5′-A and 5′-C reporters, respectively. (C) The luciferase reporter activities of the mutants SRGA, SRPA, ARGA, and ARPA.

Acknowledgments

We kindly thank Feng Zhang for plasmids to construct TALEs and Scot Wolfe for plasmids and cells for B1H screening. We acknowledge T. Morii for providing a DNA sequencer and H. Kono for helpful discussions. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (24510295, 24119507); The Naito Foundation; the ZE Research Program, IAE, Kyoto University (ZE25 B-1), and Platform for Drug Discovery, Informatics, and Structural Life Science.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2013.10.060.

References

[1] A.J. Bogdanove, D.F. Voytas, TAL effectors: customizable proteins for DNA targeting, Science 333 (2011) 1843–1846.
[2] E.L. Doyle, B.L. Stoddard, D.F. Voytas, A.J. Bogdanove, TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins, Trends Cell Biol. 23 (2013) 390–398.
[3] T. Gaj, C.A. Gersbach, C.F. Barbas 3rd, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering, Trends Biotechnol. 31 (2013) 397–405.
[4] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, U. Bonas, Breaking the code of DNA binding specificity of TAL-type III effectors, Science 326 (2009) 1509–1512.
[5] M.J. Moscou, A.J. Bogdanove, A simple cipher governs DNA recognition by TAL effectors, Science 326 (2009) 1501.
[6] P. Romer, S. Recht, T. Strauss, J. Elsaesser, S. Schornack, J. Boch, S. Wang, T. Lahaye, Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. oryzae, New Phytol. 187 (2010) 1048–1057.
[7] A.J. Bogdanove, S. Schornack, T. Lahaye, TAL effectors: finding plant genes for disease and defense, Curr. Opin. Plant Biol. 13 (2010) 394–401.
[8] A.N. Mak, P. Bradley, A.J. Bogdanove, B.L. Stoddard, TAL effectors: function, structure, engineering and applications, Curr. Opin. Struct. Biol. 23 (2013) 93–99.
[9] A.N. Mak, P. Bradley, R.A. Cernadas, A.J. Bogdanove, B.L. Stoddard, The crystal structure of TAL effector PthXo1 bound to its DNA target, Science 335 (2012) 716–719.
[10] H. Gao, X. Wu, J. Chai, Z. Han, Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region, Cell Res. 22 (2012) 1716–1720.
[11] D. Deng, C. Yan, X. Pan, M. Mahfouz, J. Wang, J.K. Zhu, Y. Shi, N. Yan, Structural basis for sequence-specific recognition of DNA by TAL effectors, Science 335 (2012) 720–723.
[12] M.M. Mahfouz, L. Li, M. Shamimuzzaman, A. Wibowo, X. Fang, J.K. Zhu, De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks, Proc. Natl. Acad. Sci. USA 108 (2011) 2623–2628.
[13] F. Zhang, L. Cong, S. Lodato, S. Kosuri, G.M. Church, P. Arlotta, Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription, Nat. Biotechnol. 29 (2011) 149–153.
[14] R.R. Beerli, D.J. Segal, B. Dreier, C.F. Barbas 3rd, Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks, Proc. Natl. Acad. Sci. USA 95 (1998) 14628–14633.
[15] M.B. Noyes, X. Meng, A. Wakabayashi, S. Sinha, M.H. Brodsky, S.A. Wolfe, A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system, Nucleic Acids Res. 36 (2008) 2547–2560.
[16] X. Meng, M.H. Brodsky, S.A. Wolfe, A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors, Nat. Biotechnol. 23 (2005) 988–994.
[17] J.F. Meckler, M.S. Bhakta, M.S. Kim, R. Ovadia, C.H. Habrian, A. Zykovich, A. Yu, S.H. Lockwood, R. Morbitzer, J. Elsaesser, T. Lahaye, D.J. Segal, E.P. Baldwin, Quantitative analysis of TALE-DNA interactions suggests polarity effects, Nucleic Acids Res. 41 (2013) 4118–4128.
[18] M. Imanishi, K. Matsumura, S. Tsuji, T. Nakaya, S. Negi, S. Futaki, Y. Sugiura, Zn(II) binding and DNA binding properties of ligand-substituted CXHH-type zinc finger proteins, Biochemistry 51 (2012) 3342–3348.
[19] M.B. Noyes, Analysis of specific protein–DNA interactions by bacterial one-hybrid assay, Methods Mol. Biol. 786 (2012) 79–95.
[20] T.D. Schneider, R.M. Stephens, Sequence logos: a new way to display consensus sequences, Nucleic Acids Res. 18 (1990) 6097–6100.
[21] O. de Lange, T. Schreiber, N. Schandry, J. Radeck, K.H. Braun, J. Koszinowski, H. Heuer, A. Strauss, T. Lahaye, Breaking the DNA-binding code of *Ralstonia solanacearum* TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease, New Phytol. 199 (2013) 773–786.
[22] B.M. Lamb, A.C. Mercer, C.F. Barbas 3rd, Directed evolution of the TALE N-terminal domain for recognition of all 5′ bases, Nucleic Acids Res. 2013, http://dx.doi.org/10.1093/nar/gkt754.
[23] A.C. Mercer, T. Gaj, R.P. Fuller, C.F. Barbas 3rd, Chimeric TALE recombinases with programmable DNA sequence specificity, Nucleic Acids Res. 40 (2012) 11163–11172.