BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **224,** 219–223 (1996)
ARTICLE NO. 1010

A Single Amino Acid Determines the Specificity for the Target Sequence of Two Zinc-Finger Proteins in Plants

Hiroshi Takatsuji

National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba, Ibaraki, 305 Japan

Received May 22, 1996

The EPF family is a group of DNA-binding proteins with two canonical Cys_2/His_2 zinc-finger motifs in *Petunia.* These proteins are unique in terms of structure in that (i) the two zinc fingers are separated by spacers of various lengths and (ii) the sequence QALGGH is strongly conserved in the zinc-finger motifs of members of the family. In this study, domain-swapping and site-directed mutagenesis experiments with two members of the protein family, EPF2-5 and EPF2-7, which have different target sequences, revealed that only a single amino acid in the second zinc finger is responsible for the difference in target specificity. The position of this amino acid is different from those of determinants of target-sequence specificity in other zinc-finger proteins. Thus, the EPF family recognizes target sequences in a unique manner, together with the recognition of spacings in the target sequence that we demonstrated recently. © 1996 Academic Press, Inc.

The EPF family is a group of DNA-binding proteins with two canonical Cys_2/His_2 zinc-finger motifs that we isolated from *Petunia hybrida* (1,2). The genes for these proteins are expressed in a floral organ-specific manner that is correlated with the expression of genes that control the identity of individual floral organs. This pattern of expression, when considered together with the results of transgenic studies (unpublished results), suggests that the EPFs might be floral organ-specific patterns of expression (unpublished results). Thus, the EPF family seems to be a large family of transcription factors that might possibly be involved in the control of organ development in *Petunia.*

The members of the EPF family are unique in that they include two zinc fingers that are separated by spacers of various lengths (from 19 to 232 amino acids) and in that the sequence QALGGH is strongly conserved in the zinc-finger motifs. We have demonstrated that EPF2-5 binds to two separated AGT sequences, with each zinc-finger making contact with one core sequence. We have also demonstrated that each member of the family has a different preference for the distance between the core nucleotide sequences. On the basis of these observations, we have proposed a model wherein each member of the EPF family distinguishes its target gene by two features: the primary nucleotide sequences of the core sites and the distance between the core sequences (submitted).

In some of Cys_2/His_2 zinc-finger proteins, such as Sp1, Krox20, and Zif268, three positions in the α-helical region ($CXXCXXXFX\underline{X}^{a}XX\underline{X}^{b}LX\underline{X}^{c}HXXXH$) have been shown to participate in major interactions with nucleotides in target sequences (3-7). These amino acids recognize one nucleotide in a triplet of the target sequence, as demonstrated by mutational and X-ray crystallographic studies. The importance of these positions in the recognition of target sequences was also revealed by statistical analysis (8) and by phage-display selection of randomized fingers (9). However, this rule is not applicable to all Cys_2/His_2-type zinc-finger proteins (10,11). In the EPF proteins, positions with superscripts b and c correspond to Ala and the second Gly residue in the QALGGH motif, respectively. If these positions were determinants of base recognition, the diversity of target sequences that could be recognized by the members

0006-291X/96 $18.00

of EPF protein family would be very limited because these amino acids are strongly conserved at these positions in more than 30 members. Moreover, the Gly residue, which has no side chain, is unlikely to make contact with a base. Thus, we can assume that amino acids at other positions are more likely to be determinants in the recognition of specific bases.

Our previous results suggest that EPF2-5 and EPF2-7 recognize different target sequences in spite of the very similar structures of their zinc-finger motifs (1). These two proteins provide a very good system for attempts to characterize the protein-DNA interaction. In this study, we have performed domain-swapping and site-directed mutagenesis experiments to identify the amino acids that determine the different specificity required for recognition of different target sequences by the two proteins.

EXPERIMENTAL PROCEDURES

Generation of truncated forms of EPF2-5 and EPF2-7. A truncated form of the gene for EPF2-5 (encoding EPF2-5ZF; residues 66-210) was created for expression in a rabbit reticulocyte lysate system as follows. A DNA fragment was amplified with an upstream primer (GGCCTCTAGAACAATGGGAACTACACCCGGTTCAACTGATACTACT) that contained a restriction site for Xba I and an initiation codon and a downstream primer (GGCCCTCGAGTCAACT-TCCACTATGACCACCGCCGTCACGGTG) that contained a restriction site for Xho I and a stop codon. The amplified DNA was digested with Xba I and Xho I and then inserted into the vector pBC SK$^-$ (Stratagene, La Jolla, CA) to yield pBC EPF2-5ZF. A truncated form of the gene for EPF2-7 (encoding EPF2-7ZF; residues 62-198) was generated by a similar procedure (pBC EPF2-7ZF).

Generation of mutant forms of the EPF protein. Chimeric genes composed of three region derived from either EPF2-5ZF or EPF2-7ZF, were generated by the recombinant PCR method (12). Two sets of site-specific complementary primers (F1/R1 and F2/R2) were designed, corresponding to two crossover sites that are similar in the DNA sequences of EPF2-5ZF and EPF2-7ZF. The DNA fragment corresponding to each portion of EPF2-5ZF or EPF2-7ZF was amplified by PCR with Pfu DNA polymerase (Stratagene) using a particular combination of site-specific and/or vector-specific (M13 reverse and M13 -21) primers. To generate chimeric genes composed of two or three regions, one-tenth each of the products of the first PCRs were combined and a second PCR was carried out with these templates, which were complimentary to each other at one end, using the M13 reverse primer and the M13 -21 primer.

Point mutations that encoded substitutions of amino acids were introduced into the coding regions of the wild-type or chimeric genes by the recombinant PCR method, as described for the generation of chimeric genes but with mutated primers.

All the reaction products were digested with Xba I and Xho I and then inserted into the pBC SK$^-$ vector. All constructs were confirmed by DNA sequencing.

Transcription and translation in vitro. The plasmids that contained wild-type and mutant derivatives of the genes for EPF2-5ZF and EPF2-7ZF were digested by Xho I, treated with proteinase K, extracted with phenol/chloroform (1:1, v/v), and then purified with a Wizard DNA cleanup kit (Promega, Madison, WI). The templates were transcribed with T3 RNA polymerase from the T3 promoter in the pBC SK$^-$ vector and the transcripts were purified by extraction with phenol/chloroform (1:1, v/v) and ethanol precipitation. One microgram of transcripts was translated in a rabbit reticulocyte lysate in the presence of [^{35}S] methionine (1,000 Ci/mmol; Amersham, Tokyo). The sizes and amounts of the products were determined by electrophoresis in a 12.5% SDS-polyacrylamide gel with subsequent autoradiography. Equivalent amounts of products were used for gel-shift experiments.

DNA-binding assays. All DNA-binding reactions were carried out in 25 mM HEPES-KOH (pH 7.6), 40 mM KCl, 0.1% NP-40, 0.01 mM $ZnCl_2$, 10 μg/ml poly(dI-dC), and 0.1 mM DTT. Gel-shift assays were performed with 10,000 cpm of ^{32}P-end-labeled probe A (tandem repeat of TTGACAGTGTCACTTGACAGTGTCAC) and about 1 μl of the product of *in vitro* translation. After incubation for 20 min at room temperature, the mixtures were subjected to electrophoresis in a 0.7% agarose / 3% polyacrylamide gel as described previously (2).

RESULTS AND DISCUSSION

EPF2-5 and EPF2-7 are similar in terms of amino acid sequences, in particular in each of their two zinc-finger motifs (Fig. 1). However, their patterns of expression are very different. The gene for EPF2-5 is expressed preferentially in epidermal cells of petals and stamens, whereas the gene for EPF2-7 is expressed in phloem cells, mostly in sepals and petals. The different patterns of expression suggest that the two putative transcription factors might play different roles, thereby regulating the expression of different target genes. In support of this possibility, the sequences of their target DNA sequences appear to be different in spite of the

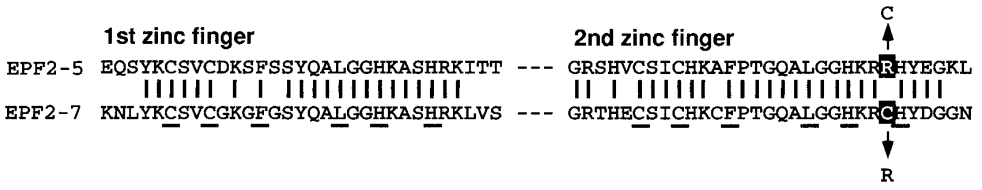

FIG. 1. Alignment of amino acid sequences in the zinc-finger regions of EPF2-5 and EPF2-7. Identical amino acids are linked by vertical lines. Six amino acids that are conserved in Cys_2/His_2-type zinc-finger proteins are underlined. The amino acids that were mutated in the experiments for which results are shown in Figure 3 are highlighted.

strong conservation of sequences in the zinc fingers. EPF2-5 binds strongly to two separated AGT core sites, with each zinc finger making contact with one core site. By contrast, EPF2-7 binds to this sequence only very weakly and we have not yet determined the optimal binding sequence for EPF2-7.

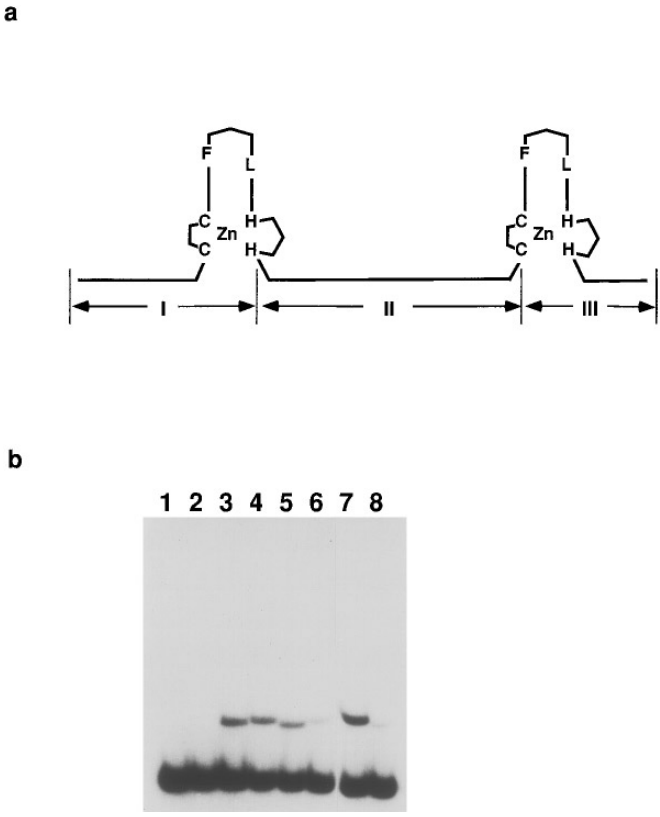

FIG. 2. The second zinc finger is responsible for the different DNA-binding activities of EPF2-5 and EPF2-7. (a) A scheme for the generation of chimeric proteins that consist of various combinations of regions of EPF2-5ZF and EPF2-7ZF. The three regions of the EPF protein that were used to generate chimeric proteins are indicated: the amino-terminal region, including the first zinc finger (I); the spacer region (II); and the carboxy-terminal region, including the second zinc finger (III). The two crossover points are located adjacent to the second conserved His residue in the first zinc finger and upstream of the first Cys residue in the second zinc finger. (b) All six combinations of the above three regions of either EPF2-5ZF or EPF2-7ZF were examined for binding to probe A. The chimeric proteins are defined by three numbers, n1-n2-n3, in which the numbers are either 5 (EPF2-5) or 7 (EPF2-7) to indicate the three regions derived from the two proteins. Lanes: 1, 5-5-7; 2, 5-7-7; 3, 5-7-5; 4, 7-5-5; 5, 7-7-5; 6, 7-5-7; 7, 5-5-5 (EPF2-5ZF); and 8, 7-7-7 (EPF2-7ZF).

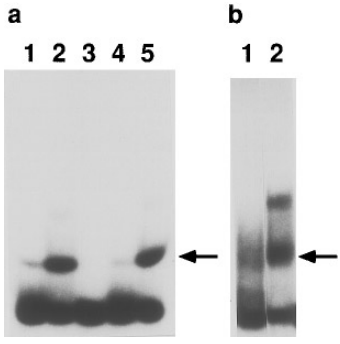

FIG. 3. A single amino acid in the second zinc finger is responsible for the difference in binding activities between EPF2-5 and EPF2-7. (a) Cys in the second zinc finger in EPF2-7ZF and in the chimeric protein 5-5-7 was replaced by Arg (see Figure 1). The mutated proteins were examined for binding to probe A. Lanes: 1, EPF2-7ZF; 2, Cys-to-Arg mutant of EPF2-7ZF; 3, chimeric protein 5-5-7; 4, Cys-to-Arg mutant of 5-5-7; and 5, EPF2-5ZF. (b) Arg in EPF2-5ZF was replaced by Cys and the resultant protein was examined for binding to probe A. Lanes: 1, Arg-to-Cys mutant of EPF2-5ZF; and 2, EPF2-5ZF. Positions of shifted bands are indicated by arrows. The upper bands in some lanes are due to the binding of two protein molecules to one molecule of probe.

These two proteins provide us with a very good system for identifying some of the amino acids that participate in the recognition of their target DNA sequences. To examine which parts of the zinc-finger regions are responsible for the difference in specificity, we first generated chimeric proteins with all possible combinations of the three portions of the two proteins: the first zinc finger, the spacer region, and the second zinc finger, as shown in Figure 2a. The chimeric proteins were examined for their ability to bind to probe A that contained two tandemly repeated AGT core sites, separated by 13 bp. These experiments revealed that only those chimeric proteins that contained the second zinc finger from EPF2-5 could bind to probe A with high affinity, comparable to the affinity of binding of wild-type EPF2-5ZF (Fig. 2-b). These results indicate that differences between second zinc fingers are responsible for the difference in DNA-binding activity.

There are only few positions in the second zinc fingers at which the amino acids differ between EPF2-5 and EPF2-7. Among them, we focused on one position between two conserved histidines (Fig. 1) within a putative α-helical region, which is known to face the surface of DNA when certain other zinc finger proteins bind to it. This position is occupied by Arg in EPF2-5 and Cys in EPF2-7, respectively. As shown in Figure 3-a, the replacement of Cys by Arg in EPF2-7ZF dramatically enhanced the binding to probe A. By contrast, the reverse replacement, of Arg by Cys, in EPF2-5ZF reduced the binding affinity (Fig. 3-b). These results indicate that the difference in amino acids at this position is responsible for the difference in DNA-binding activity and, thus, this position is a strong candidate for one of the determinants of DNA-binding specificity. It should be noted that this position is different from those identified in Sp1, Krox20 and Zif 268, which were mentioned above (3-7).

When the Cys-to-Arg substitution was made in the chimeric protein that had the first zinc finger and the spacer region of EPF2-5 and the second zinc finger of EPF2-7, the binding activity was clearly enhanced. However, the binding activity did not reach the level achieved with wild-type EPF2-5. These results suggest that there must be some compatibility between the second zinc finger and the first zinc finger or, more likely, the spacer region.

Our results suggest that the members of the EPF family make contact with DNA at very different positions within the zinc finger from those of the Cys_2/His_2-type zinc-finger proteins that were characterized previously. This difference is presumably associated with a structural feature of the EPF family, namely, the strongly conserved QALGGH motif at the base-

recognition surface. Together with the recognition of the spacing in the target DNA, this mode of DNA recognition seems to be unique to this family of transcription factors. We are in the process of conducting X-ray crystallographic and two-dimensional NMR analyses for further characterization of the protein-DNA interactions. An understanding of the mechanism for the recognition of target-DNA sequences should lead to prediction of the target sequences of other members of the EPF family and, furthermore, to the design of artificial transcription factors that recognize specifically selected sequences.

ACKNOWLEDGMENTS

This work was supported in part by the Enhancement of Center of Excellence, Special Coordination Funds for Promoting Science and Technology, Science and Technology Agency, Japan. The author is grateful to Ms. Yoshiko Ouchi for her technical assistance.

REFERENCES

1. Takatsuji, H., Nakamura, N., and Katsumoto, Y. (1994) *Plant Cell* **6,** 947–958.
2. Takatsuji, H., Mori, M., Benfey, P. N., Ren, L., and Chua, N.-H. (1992) *EMBO J.* **11,** 241–249.
3. Nardelli, J., Gibson, T. J., Vesque, C., and Charnay, P. (1991) *Nature* **349,** 175–178.
4. Nardelli, J., Gibson, T., and Charnay, P. (1992) *Nucl. Acids. Res.* **20,** 4137–4144.
5. Desjarlais, J. R., and Berg, J. M. (1992) *Proteins* **12,** 101–104.
6. Desjarlais, J. R., and Berg, J. M. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89,** 7345–7349.
7. Pavletich, N., and Pabo, C. (1991) *Science* **252,** 809–817.
8. Jacobs, G. H. (1992) *EMBO J.* **11,** 4507–4517.
9. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91,** 11163–11167.
10. Pavletich, N. P., and Pabo, C. O. (1993) *Science* **261,** 1701–1707.
11. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Nature* **366,** 483–487.
12. Higuchi, R. (1990) *in* PCR Protocols; A Guide to Methods and Applications (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., Eds.), pp. 177–183, Academic Press, San Diego.