J|A|C|S
COMMUNICATIONS

Published on Web 12/08/2001

A General Strategy To Determine a Target DNA Sequence of a Short Peptide: Application to a D-Peptide

Takashi Morii,*[†,‡] Tomohisa Tanaka,[†] Shin-ichi Sato,[†] Masaki Hagihara,[†] Yasunori Aizawa,[†] and Keisuke Makino[†]

Institute of Advanced Energy, Kyoto University and PRESTO, Japan Science and Technology Corporation, Uji, Kyoto 611-0011, Japan

Received September 14, 2001

The design of molecules that target desired DNA sequences has been one of the major challenges in the field of molecular recognition. The most successful strategy to date uses polyamides containing *N*-methylpyrrole and *N*-methylimidazole recognition elements to systematically read a wide variety of sequences from the DNA minor groove.[1] Native sequence-specific DNA binding proteins, in contrast, recognize the rich information in the major groove of DNA.[2] The major groove is specifically targeted by multifinger proteins prepared from the Cys_2His_2 zinc finger motif with the phage display technique[3] and by the triple helix formation.[4] However, no example has been reported for systematic recognition of DNA sequences by short peptides despite the fact that native DNA-binding proteins mostly utilize short peptide regions for the direct contact to three or four DNA base pairs.[2] While short peptides could potentially provide an alternative element to read-out DNA sequences from the major groove, it is difficult to determine sequence-preference of de novo designed monomeric short peptides. Because DNA-binding affinity and specificity of short peptides are usually much lower than those of native DNA-binding proteins, determining sequence-preference of short peptides by conventional methods utilized to deduce the target sequence of proteins, such as the DNase footprinting,[5] often produces unclear outcomes. We report here a general strategy for defining the sequence-preference of a DNA-binding short peptide using the heterodimers. Our method successfully identified specific sequences of short peptides derived from native DNA-binding proteins. The usefulness of this approach has been demonstrated by identifying preferred DNA targets for a peptide composed only of D-amino acids.

Short peptides derived from the basic region of leucine zipper protein GCN4 cooperatively bind the major groove of native GCN4 binding sequence as a dimer upon simple modifications with β-cyclodextrin and adamantane.[6] The cyclodextrin-adamantane dimerization is also effective in forming a heterodimer of two different peptides. Because the heterodimer binds DNA more strongly and specifically than the peptide monomer does,[7] sequence preference of the peptide would be easily determined in its heterodimeric form.

The specific DNA sequence for the heterodimer is determined by the selected and amplified binding site (SAAB) imprinting technique[8] that has been widely used to determine the optimal DNA-binding sequence for native protein from a pool of DNA sequences (Figure 1). A DNA pool containing five randomized base-pairs (N5), which provide enough binding site size for a short peptide, adjacent to 5′-ATGAC-3′ is chemically synthesized to cover all the possible binding sequences for the heterodimer. The basic region peptide of GCN4 modified by the adamantyl group at its C-terminal cysteine (GAd)[6] anchors at the 5′-ATGAC-3′ sequence in a heterodimer,[6,7] thereby facilitating the location of the other peptide in the major groove of the randomized N5 sequence. A binding mixture of the 5′-^{32}P-end-labeled DNA pool and the peptide dimer is separated into the peptide dimer-bound and the unbound DNA by polyacrylamide gel electrophoresis (PAGE). The peptide dimer-bound DNA is recovered from the gel, and is amplified by the polymerase chain reaction (PCR). The resulting DNA pool should contain a higher population of the specific DNA for the heterodimer, implying that the fraction of dimer-bound DNA would increase with the progress of the SAAB cycle. The SAAB cycle is repeated until no more increase in the binding efficiency is observed, then the resulting DNA pool is cloned into a vector, followed by sequencing to determine the selected nucleotide sequence at N5. In the whole scheme of the experiment, the binding condition for the peptide dimer with the pool of DNA is critical for successful selection of the target sequence. We first optimized the binding condition by using a peptide dimer consisting of GAd and its cyclodextrin derivative GCd, for which the target DNA sequence already has been determined.[6]

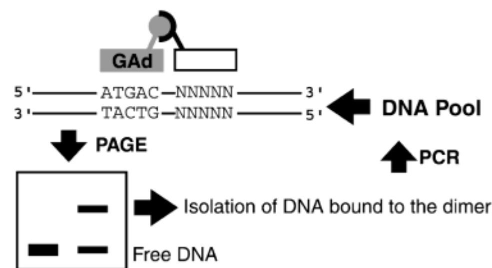

Figure 1. A SAAB cycle to determine a target sequence of a short peptide (open box) by using its heterodimer with GAd (hatched box). Amino acid sequences for GAd (GCd)[8] and CCd[9a] are Ac-DPAALKRARN-TEAARRSRARKLQC-NH_2 and Ac-NEYRYRRERNNIAVRKSRDKAKQC-NH_2, respectively.

For the initial three SAAB cycles, a low salt condition (4 mM KCl) has been applied to form the DNA−peptide dimer complex. The fraction of DNA bound for the GAd/GCd dimer (200 nM) was increased to 49% at the third round, but the resulting DNA pool did not reveal any characteristic sequence at the N5 region. From the fourth SAAB round, binding reactions were performed with 100 mM KCl and a nonspecific competitor DNA poly(dI-dC). In the ninth round, 61% of DNA was bound to the dimer (200 nM), suggesting that the DNA pool has been enriched with the specific sequence for GAd/GCd. Clones from the ninth round DNA pool dominantly afforded the 5′-GTCA(T/C)-3′ sequences at the N5 region (Table S1). Native GCN4 dimer binds the 5′-TGACTCA-3′ or 5′-TGAC-GTCA-3′, with the outer two base pairs being either A/T or G/C,[9] indicating that the condition applied here (Table 1) is sufficient to deduce the specific binding sequence of the GCN4 basic region peptide. The SAAB condition was next applied to select a target sequence of a 23 amino acid residue peptide modified with

* To whom correspondence should be addressed. E-mail: t-morii@iae.kyoto-u.ac.jp.
[†] Kyoto University.
[‡] PRESTO, JST.

10.1021/ja017078f CCC: $22.00 © 2002 American Chemical Society

Table 1. Binding Conditions for the Peptide Dimer and DNA Used in SAAB[a]

SAAB round	KCl	dI-dC
1	4 mM	
2	4 mM	20 μM
3	4 mM	30 μM
4	100 mM	20 μM
5−12	100 mM	30 μM

[a] The binding mixture contains 20 mM Tris-HCl (pH 7.6), 2 mM $MgCl_2$, 1 mM EDTA, and 6% sucrose.

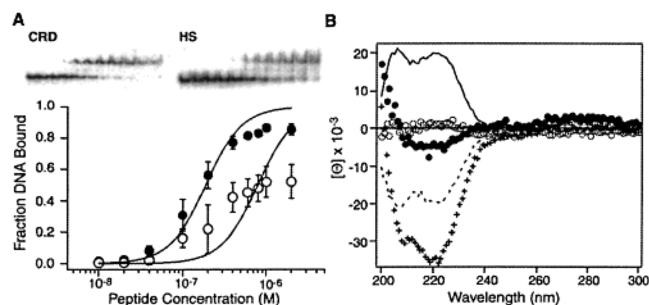

Figure 2. (A) An autoradiogram of the gel shift assay shows GAd/DGCd binds more efficiently to CRD than HS. Peptide concentrations from left to right were 0, 10, 20, 40, 100, 200, 400, 600, 800, 1000, and 2000 nM for both CRD and HS. Semilogarithmic plots show the fraction of ^{32}P-labeled CRD (filled circles) and HS (open circles) bound to GAd/DGCd. (B) Circular dichroism differential spectra of GAd/DGCd in the absence (open circles) or presence (filled circles) of CRD show an increase of the right-handed helix upon binding to CRD. Spectra of monomeric free GAd (dotted line) and DGCd (solid line) indicate both peptides are in the partially helical conformation but with an opposite handedness. GAd forms a full-helical structure (crosses) upon binding to 5′-ATGAC-3′.

cyclodextrin (CCd)[7a] derived from the DNA binding region of transcription factor C/EBP.[10] The fraction of DNA bound to the heterodimer GAd/CCd raised to 55% in the tenth round SAAB, and the resulting DNA pool was sequenced to give predominantly 5′-GCAA(T/C)-3′ at the N5 region (Table S1), which exactly matched with the preferential binding sequence of native C/EBP.[11] These results demonstrate that the target DNA sequences of short peptides are determined by the SAAB system with use of the heterodimer.

A peptide composed of D-amino acids (D-peptide) shares an exact mirror image structure of the corresponding L-peptide, and has been demonstrated as a potential drug.[12] A short D-peptide with DNA binding ability would be an excellent candidate for the major groove-binding element. To ascertain the DNA binding of D-peptide, the SAAB procedure was next applied to a heterodimer of the D-amino acid analogue of the GCN4 basic region peptide (DGCd). The fraction of DNA bound to the GAd/DGCd dimer was 28% in the third round, where no distinct pattern was observed for the sequences at the N5 site. By repeating SAAB to the ninth round at high salt conditions, the fraction of DNA bound to GAd/DGCd was recovered to 28%, and sequencing of the DNA pool revealed an increase in the population of 5′-ACNNN-3′ at the N5 site (Figure S1). SAAB was repeated further, but the DNA binding efficiency moderately increased to 38% in the twelfth round. Out of eighteen clones sequenced, eight clones possessed 5′-ACACN-3′, three 5′-ACGNN-3′, and two 5′-ACCNN-3′ at N5.

Specific DNA binding of GAd/DGCd was verified by using a DNA containing the most abundant 5′-ATGAC-ACACA-3′ (CRD) sequence (Figure 2A). The GAd/DGCd−CRD complex was formed in a cooperative manner, and titration of the gel shift afforded an equilibrium dissociation constant of 3.51×10^{-14} (M^2). GAd/DGCd bound less efficiently to a DNA containing 5′-ATGAC-ACTGC-3′ (HS) (Figure 2A) with a dissociation constant of 6.24×10^{-13} (M^2), and poorly to a DNA containing 5′-ATGAC-GCAAT-3′ without a cooperative formation of dimer−DNA complex (Figure S2). GAd/DGCd discriminated the selected CRD sequence from HS even though 5′-ATGAC-AC-3′ is common for both sequences (Figure S3). Circular dichroism differential spectra revealed a moderate increase in the intensity of the negative band corresponding to the right-handed helical structure in the complex of GAd/DGCd−CRD (Figure 2B). GAd composed of L-amino acids forms a full-helical structure in the complex with 5′-ATGAC-3′.[8,9,13] Judging from the intensity of the slightly negative CD signal, DGCd bound to 5′-ACACA-3′ not in a full-helical structure, although its helicity seemed to be higher than that in the free from. Examples of the nonhelical DNA binding motif exist for native DNA binding proteins;[14] however, the structure by which the D-peptide binds its specific DNA sequence remains to be identified.

In summary, our strategy of using the heterodimer of short peptides with the SAAB technique provides a general method to determine target DNA sequences of short peptides bound in the major groove. The method is applicable not only to artificial peptides, but also to other synthetic ligands because the cyclodextrin-adamantane dimerization domain is easily introduced to various molecules. The result that a D-peptide binds a specific DNA sequence encourages further design efforts using D-peptides with DNA binding ability. Moreover, it would be interesting to apply the heterodimer system to a library of D-peptides, which enables one to select D-peptides specifically targeting the desired DNA sequences.

Acknowledgment. This work was supported in part by the Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan to T.M. (No. 12680590) and to K.M. (No. 11101001).

Supporting Information Available: Experimental details for the SAAB protocol using the heterodimers and the DNA binding of the D-peptide (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature*, **1998**, *391*, 468. (b) Dervan, P. B.; Burli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688.
(2) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053.
(3) (a) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2256. (b) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11168. (c) Rebar, E. J.; Pabo, C. O. *Science* **1994**, *263*, 671. (d) Wu, H.; Yang, W. P.; Barbas, C. F., III *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 344. (e) Greisman, H. A.; Pabo, C. O. *Science* **1997**, *275*, 657.
(4) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.
(5) Gallas, D. J.; Schmitz, A. *Nucleic Acids Res.* **1978**, *5*, 3157.
(6) Ueno, M.; Murakami, A.; Makino, K.; Morii, T. *J. Am. Chem. Soc.* **1993**, *115*, 12575.
(7) (a) Ueno, M.; Sawada, M.; Makino, K.; Morii, T. *J. Am. Chem. Soc.* **1994**, *116*, 11137. (b) Morii, T.; Yamane, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* **1996**, *118*, 10011.
(8) Blackwell, T. K.; Weintraub, H. *Science* **1990**, *250*, 1104.
(9) Sellers, J. W.; Vincent, A. C.; Struhl, K. *Mol. Cell. Biol.* **1990**, *10*, 5077.
(10) Landschulz, W. H.; Johnson, P. F.; Adashi, E. Y.; Graves, B. J.; McKnight, S. L. *Genes Dev.* **1988**, *2*, 786.
(11) Johnson, P. F. *Mol. Cell. Biol.* **1993**, *13*, 6919.
(12) (a) Schumacher, T. N. M.; Mayr, L. M.; Minor, D. L., Jr.; Milhollen, M. A.; Burgess, M. W.; Kim, P. S. *Science* **1996**, *271*, 1854. (b) Eckert, D. M.; Malashkevich, V. N.; Hong, L. H.; Kim, P. S. *Cell* **1999**, *99*, 103.
(13) (a) O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F. *Biochemistry* **1991**, *30*, 9030. (b) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223. (c) König, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139.
(14) (a) Breg, J. N.; van Opheusden, J. H. J.; Burgering, M. J. M.; Boelens, R.; Kaptein, R. *Nature* **1990**, *346*, 586. (b) Rafferty, J. B.; Somers, W. S.; Saint-Girons, I.; Phillips, S. E. V. *Nature* **1989**, *341*, 705. (c) Kim, J. L.; Nikolov, D. B.; Burley, S. K. *Nature* **1993**, *365*, 520. (d) Ghosh, G.; van Duyne, G.; Ghosh, S.; Sigler, P. B. *Nature* **1995**, *373*, 303. (e) Muller, C. W.; Rey, F. A.; Sodeoka, M.; Verdine, G. L.; Harrison, S. C. *Nature* **1995**, *373*, 311.

JA017078F