Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry
Bioorganic & Medicinal Chemistry 13 (2005) 3107–3116

Dual DNA recognition codes of a short peptide derived from the basic leucine zipper protein EmBP1

Akiyoshi Hirata,[a] Masaru Ueno,[b] Yasunori Aizawa,[a] Katsutoshi Ohkubo,[a] Takashi Morii[a,*] and Susumu Yoshikawa[a]

[a]*Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan*
[b]*Department of Chemistry, Shizuoka University, Oya, Shizuoka 422-8529, Japan*

Received 10 February 2005; revised 24 February 2005; accepted 25 February 2005

Abstract—Sequence-specific DNA binding of short peptide dimers derived from a plant basic leucine zipper protein EmBP1 was studied. A homodimer of the EmBP1 basic region peptide recognized a palindromic DNA sequence, and a heterodimer of EmBP1 and GCN4 basic region peptides targets a non-palindromic DNA sequence when a β-cyclodextrin/adamantane complex is utilized as a dimerization domain. A homodimer of the EmBP1 basic region peptide binds the native EmBP1 binding 5′-GCCACGTGGC-3′ and the native GCN4 binding 5′-ATGACGTCAT-3′ sequences with almost equal affinity in the α-helical conformation, indicating that the basic region of EmBP1 by itself has a dual recognition codes for the DNA sequences. The GCN4 basic region peptide binds 5′-ATGAC-3′ in the α-helical conformation, but it neither shows affinity nor helix formation with 5′-GCCAC-3′. Because native EmBP1 forms 100 times more stable complex with 5′-GCCACGTGGC-3′ over 5′-ATGACGTCAT-3′, our results suggest that the sequence-selectivity of native EmBP1 is dictated by the structure of leucine zipper dimerization domain including the hinge region spanning between the basic region and the leucine zipper.

1. Introduction

The X-ray crystallographic and NMR studies provide detailed information on the stereo-specific interactions between the amino acid residues in the protein and the functional group in DNA. Such high resolution structures of sequence-specific protein–DNA complexes provide a far greater understanding of how proteins and DNA assemble and function within these complexes to perform essential cellular processes and how the proteins recognize specific DNA sequences.[1–4] Native sequence-specific DNA-binding proteins mostly bind in the major groove by using a simple secondary structure, usually an α-helix, which is complementary to the structure of B-DNA major groove. Structural studies and sequence comparisons revealed that many DNA-binding proteins could be grouped into classes that use related structural motifs for recognition. It seems reasonable to assume that the helical peptides are convenient scaffolds for the design of new DNA-binding peptides and/or proteins. However, a universal code for the recognition between short helices and nucleic acids has yet to be generalized.[5]

Basic leucine zipper protein (bZIP)[6] binds in the major groove by using an α-helix, which is complementary to the shape of DNA major groove. DNA-binding affinity of bZIP protein is amplified by dimerization that is mediated through a coiled-coil structure at the leucine-zipper domain located at the C-terminal of the basic region. One of the best structurally characterized bZIP family of proteins is the yeast transcription activator GCN4. GCN4 is known to bind DNA as a homodimer with each basic region directly contacting the major groove of DNA. The native GCN4 specifically binds 5′-ATGACTCAT-3′ (AP1) and 5′-ATGACGTCAT-3′ (CRE) sequences in a similar affinity.[7–10] Structural studies and sequence comparison analyses revealed that bZIP family proteins could be grouped into several classes depending on the usage of amino acids motifs for recognition, and mutation studies of basic region amino acid residues revealed that the sequence–preference of GCN4 changed to various DNA sequence.[11–17] However, no

Keywords: DNA recognition; Peptides.
* Corresponding author. Tel.: +81 774 38 3585; fax: +81 774 38 3516; e-mail: t-morii@iae.kyoto-u.ac.jp

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.02.052

rule has been reported for a systematic recognition of the DNA sequence by bZIP proteins. The basic region peptides derived from GCN4 can serve as the simplest short peptides that can target DNA sequences four base pairs in size, but the DNA-binding affinity of a short basic region peptide monomer is inherently low. The basic region peptides of bZIP is sufficient for the sequence-specific DNA binding when covalently[18–26] or non-covalently dimerized.[27–31] A host–guest inclusion complex of β-cyclodextrin and the adamantly group is very useful because this module enables a cooperative DNA binding and a recognition of non-palindromic DNA sequences by short basic region peptide dimers.[32,33] In the previous study, we used short basic region peptides derived from GCN4 and C/EBP.[34] A plant bZIP protein of EmBP1 specifically binds a 5′-GCCACGTGGC-3′ sequence[35–37] different from the AP1 or CRE sequences. The bZIP factors from plants bind the so called G box and C box, possessing 5′-CACGTG-3′ and 5′-GACGTC-3′ sequences, respectively, although little is known for these plant bZIP factors on the mechanisms of sequence-selective DNA binding. We herein describe sequence-specific DNA binding of peptide dimers derived from the basic region of GCN4 and EmBP1. The EmBP1 basic region peptide dimer binds both 5′-CACGTG-3′ and 5′-GACGTC-3′ sequences in the helical conformation, indicating that the EmBP1 basic region peptide possesses dual recognition codes.

2. Materials and methods

2.1. Materials

Protected Fmoc (9-fluorenylmethoxycarbonyl) amino acids and PyBOP (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate) were obtained from Novabiochem. 1-Hydroxybenzotriazole (HOBt) was from Nakaraitesque. Fmoc-PAL-PEG resin (0.2 mmol/g) was from PerSeptive Biosystems. Dimethylformamide was dried over CaH_2, distilled from ninhydrine at 55 °C under reduced pressure, and stored over molecular sieves 4 Å. Protected nucleoside phosphoramidites were obtained from PerSeptive Biosystems. T4 polynucleotide kinase was obtained from New England Biolab. [γ-^{32}P]adenosine-5′-triphosphate was from Amersham. HPLC grade acetonitrile (Wako Chemicals) was used for both analytical and preparative HPLC. Reagent grade water was used throughout the experiments. Gel electrophoresis grade acrylamide and bisacrylamide were obtained from Wako Chemicals. All other chemicals were reagent grade and used without further purification. Sephadex G-10 and G-25 were obtained from Pharmacia. A reversed-phase C18 column (20 × 250 mm, Ultron VX-Peptide, Shinwa Chemical Industry) was used for purification of peptides for preparative purpose. Analytical HPLC were carried out on a reversed-phase C18 column (4.6 × 150 mm, Ultron VX-Peptide, Shinwa Chemical Industry). Oligonucleotides were purified on a reversed-phase C18 column (6 × 150 mm, Ultron VX-Nucleotide, Shinwa Chemical Industry). Amino acid analyses were performed with an AccQ Tag Chemistry Package (Waters) according to a company protocol. Molecular weights were confirmed by MALDI-TOF mass spectrometer Voyager DE-STR (PerSeptive Biosystems).

2.2. Synthesis of the basic region peptide E24

E24 was synthesized by using a solid-phase method (0.2 mmol scale). *N*-α-Fmoc-protected amino acid pentafluorophenylesters (0.5 mmol) and *N*-hydroxybenzotriazole (0.5 mmol) were used for coupling the amino acids. The peptide was cleaved from the resin by using an ice-cold TMSBr (trimethylsilyl bromide)/thioanisole/TFA (trifluoroacetic acid) cleavage mixture containing 1.35 mL TMSBr, 1.2 mL thioanisole, 0.6 mL EDT (1,2-ethanedithiol), 0.2 mL *m*-cresol, and 7.48 mL TFA. The solvent was removed and the residue was redissolved in 30 mL of water and washed with diethyl ether. The mixture was passed through a Sephadex G-10 column with a 5% acetic acid solution. The peptide was purified by a reversed-phase HPLC. HPLC condition: eluent A, 0.1% TFA–water; eluent B, 0.1% TFA–water containing 50% CH_3CN; flow rate 6 mL/min. E24: MW calculated 3240.70, found 3240.38. The amino acid sequence of E24 is, Ac-ELKRERRKQSNRESARRSRLRKQQC-NH_2. Abbreviations for the amino acid are: A, Ala; C, Cys; E, Glu; K, Lys; L, Leu; N, Asn; Q, Gln; R, Arg; S, Ser; T, Thr.

2.3. Synthesis of E24Ad

E24Ad was synthesized by a reaction of E24 (10 mg, 3.1 mmol) with *N*-(bromoacetyl)-1-adamantanemethylamine (17 mg, 60 mmol) in an aqueous DMF solution (pH 9, water:DMF = 3:2) at 0 °C under nitrogen as described.[31] After 1 h, the reaction was quenched by an addition of acetic acid, and the mixture was passed through a Sephadex G-10 column with a 5% acetic acid solution. Successive purification with a reversed-phase HPLC and lyophilization yielded pure E24Ad (6 mg, 56%). E24Ad: MW calculated 3445.99, found 3446.15. HPLC conditions: eluent A, 0.1% TFA–water, eluent B, 0.1% TFA–water containing 50% CH_3CN; flow rate 6 mL/min. Stock solution containing E24Ad peptide was prepared by dissolving the peptide in TE buffer (1 mM Tris–HCl, 0.1 mM EDTA, pH 8.0). Peptide concentrations were determined by quantitative amino acid analyses.

2.4. Synthesis of E24Cd

E24Cd was synthesized by a reaction of E24 (20 mg, 6.17 mmol) with mono-6-deoxy-6-iodo-β-cyclodextrin (124 mg, 100 mmol) in an aqueous solution (pH 9) at 0 °C under nitrogen for 5 h as described.[31] The reaction was quenched by an addition of acetic acid, and the mixture was passed through a Sephadex G-10 column with a 5% acetic acid solution. E24Cd was purified by a reversed-phase HPLC and lyophilized to yield pure E24Cd (10 mg, 37%). E24Cd: MW calculated 4357.68, found 4358.08. HPLC conditions: eluent A, 0.05% TFA–water, eluent B, 0.05% TFA–water containing 50% CH_3CN; flow rate 5 mL/min. After HPLC purification, the mobile phase solution containing E24Cd was

neutralized by an addition of triethylamine, and resulting triethylammonium trifluoroacetate was removed by passing through a Sep-pak C18 cartridge. A stock solution containing E24Cd was prepared by dissolving the peptide in TE buffer (1 mM Tris–HCl, 0.1 mM EDTA, pH 8.0). Peptide concentration was determined by a quantitative amino acid analysis.

2.5. Synthesis and 5′-end labeling of oligonucleotides CRE21, GBX21, CRGB21, NON, and CRGB46

Oligonucleotides were synthesized on a perseptive DNA synthesizer with standard method and purified by reversed-phase HPLC on an Ultron VX-Nucleotide column (Shinwa Chemical Industry; 6×150 mm) with 0.1 M triethylammonium acetate-acetonitrile as an eluent. The oligonucleotides were labeled by kinase reaction using [γ-^{32}P]ATP (5000 Ci/mmol) and T4 polynucleotide kinase, followed by removal of unincorporated label over Sep-pak. The oligonucleotide was then denatured and annealed to a 2-fold molar excess of the opposite strand in 25 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA at pH 8.0. Nucleotide sequences of the oligonucleotides used in the present study are CRE21(5′-CGGATGACGTCATTTTTTTTC-3′), GBX21 (5′-CGGGCCACGTGGCTTTTTTTC-3′), CRGB21 (5′-CGGATGACGTGGCTTTTTTTC-3′), NON (5′-GATCCCCCCAACACCTGCTGCCTGA-3′), and CRGB46 (5′-TCGAATTCCACAGTGAGAAATGACGTGGCCCAGACTTAGGATCCGC-3′). Sequences of only single strand are shown.

2.6. Electrophoretic mobility shift assay

Typical binding reactions were carried out in the presence of the indicated peptide with ~20 pM 5′-^{32}P-labeled oligonucleotide (double-stranded) in a binding mixture containing 20 mM Tris–HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM EDTA, and 6% sucrose. The binding mixtures were incubated at the indicated temperature for 30 min, and an aliquot (8 μL) of each binding mixture was directly loaded onto an 8% nondenaturing acrylamide gel (29:1 acrylamide/bis-acrylamide), run in TBE buffer (20 mM Tris, 20 mM boric acid, and 0.1 mM EDTA) at the indicated temperature, and analyzed by autoradiography. Gel electrophoresis was performed in ATTO AE-6410 gel boxes cooled by a temperature-regulated circulating methanol bath. The temperature of the buffer was maintained at the indicated temperature and the gels were thermally equilibrated and pre-electrophoresed for 45 min in an apparatus prior to loading the samples. The increase of the mobility-shifted band was quantitated by the densitometry of the autoradiogram. Concentration of the peptide was determined by quantitative amino acid analysis with α-aminobutyric acid as an internal standard.

2.7. Analysis of energetics by titration of electrophoretic mobility shift

All binding reactions were performed as described above at the indicated temperature, salt concentrations, and peptide concentrations. The equilibriums involved in the overall DNA binding reaction of Ad- and Cd-peptides are:

$$AC \overset{K_{D1}}{\rightleftarrows} A + C \tag{1}$$

$$ACD \overset{K_{D2}}{\rightleftarrows} AC + D \tag{2}$$

$$[A]_t = [A] + [AC] + [ACD] \tag{3}$$

$$[C]_t = [C] + [AC] + [ACD] \tag{4}$$

A, C and D are, respectively, the free concentrations of Ad-peptide, Cd-peptide, and DNA. ACD represents a concentration of the dimer–DNA complex and D_t represents a total concentration of DNA. K_{D1} and K_{D2} represent dissociation constants of the Ad-peptid/Cd-peptide dimer and dimer–DNA complex, respectively. Because $[A]_t$ is equal to $[C]_t$ in the present work, the following equations are derived from Eqs. 1–4 and were used to analyze the fraction of bound DNA (Fb) data.

$$\begin{aligned} \mathrm{Fb} &= [ACD]/([D] + [ACD]) \\ &= [A][C]/(K_{D1} * K_{D2} + [A][C]) \\ &= 1/[1 + 2K_{D2}/\{K_{D1} + 2[A]_t \\ &\quad - (K_{D1}^2 + 4K_{D1} * [A]_t)^{1/2}\}] \end{aligned} \tag{5}$$

2.8. DNase I footprinting[38]

A 46-mer double-stranded oligonucleotide CRGB46 was prepared as described above. Reaction mixture contained 20 mM Tris–HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 20,000 cpm singly 5′-^{32}P-end-labeled 46-mer DNA, and peptide dimer where indicated, in 20 μL total volume. Nuclease digestion was initiated on addition of one unit of DNase I on ice and was quenched by addition of 20 μL of 0.6 M NaOAc, 10 mM EDTA, and 0.4 mM calf thymus DNA after 30 s. Samples were purified by phenol–chloroform extraction and ethanol precipitation, suspended in 80% formamide loading dye, run on a 12% sequencing acrylamide gel, and analyzed by autoradiography.

2.9. Measurement of CD spectra

Spectra of the peptide in the presence of oligonucleotides were calculated as the difference between the bound spectrum and a spectrum of the respective free oligonucleotide CD spectra were obtained with a Jasco J-720 CD spectrometer equipped with at 4 °C in a 1-mm cell. Samples contained 20 mM Tris–HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM EDTA, 4 μM peptide, and 5 μM oligonucleotide duplex when present. Spectra were the average of 32 scans and were corrected with a spectrum of buffer alone but not smoothed.

3. Results

3.1. Modification of basic region short peptides derived from EmBP1 and GCN4 with adamanthyl or cyclodextrin group

Oligopeptides G23 and E24 were derived from the basic region of GCN4 and the G box binding plant bZIP protein EmBP1, respectively. G23 and E24 were modified at the C-terminal unique cysteine residue with 6-iodo-β-cyclodextrin (Cd)[27] to give G23Cd and E24Cd, respectively. Peptides with adamanthyl group, G23Ad and E24Ad, were obtained by the reaction of *N*-bromoacetyl-1-adamantanemethylamine (Ad)[27] with G23 and E24 peptides, respectively (Fig. 1). GCN4 recognizes a palindromic DNA sequence with a half-site of 5′-ATGAC-3′, while EmBP1 recognizes that with a half-site of 5′-GCCAC-3′. Homopeptide dimers G23Ad/G23Cd and E24Ad/E24Cd would recognize palindromic 5′-ATGACGTCAT-3′ (CRE) and 5′-GCCACGTGGC-3′ (GBX) sequences, respectively. On the other hand, heteropeptide dimers G23Ad/E24Cd and E24Ad/G23Cd were expected to recognize a non-palindromic sequence containing both the CRE and GBX half-sites, 5′-ATGACGTGGC-3′ (CRGB).

3.2. Evaluation of sequence-selective DNA binding of homo- and heterodimers by electrophoretic mobility shift assay

DNA binding of homo-(G23Ad/G23Cd and E24Ad/E24Cd) and heterodimers (G23Ad/E24Cd and E24Ad/G23Cd) to the palindromic (CRE and GBX) and non-palindromic (CRGB) DNA sequences were compared by electrophoretic mobility shift assay in the binding buffer containing 20 mM Tris–HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM EDTA, and 6% sucrose at 4 °C (Fig. 2). By using the dissociation constant for the inclusion complexes between the G23Ad and G23Cd peptides (K_{D1}),[31] dissociation constants for the complexes between peptide dimer and DNA (K_{D2}) were obtained by fitting the titration data to Eq. 5. Heterodimers of E24 and G23 peptides preferentially bound the non-palindromic CRGB DNA sequence. Binding isotherms for heterodimers of E24/G23 are shown in Figure 3A. The E24 peptide homodimer almost equally bound GBX, CRE, and CRGB DNA sequences (Fig. 3B), and the G23 peptide homodimer preferentially bound the CRE DNA sequence (Fig. 3C).

Stabilities of the dimer–DNA complexes were evaluated by using the apparent dissociation constants ($K_D = K_{D1} * K_{D2}$) in Tables 1 and 2. The homodimer E24Ad/E24Cd bound with almost equal affinities to CRE21, GBX21, and CRGB21 (Fig. 2A, Table 1). The G23Ad/G23Cd–CRE21 complex was 0.9 and 2.8 kcal mol^{-1} more efficiently than the G23Ad/G23Cd–CRGB21 and the G23Ad/G23Cd–GBX21 complexes, respectively (Fig. 2B, Table 1). The E24Ad/G23Cd–CRGB21 complex was 0.7 and 1.9 kcal mol^{-1} more stable than E24Ad/G23Cd–CRE21 and E24Ad/G23Cd–GBX21 complexes, respectively (Fig. 2C, Table 2). G23Ad/E24Cd did not show a cooperative formation of a dimer complex with CRE21 or GBX21. Several mobility-shifted bands were observed in addition to the bands corresponding to the free DNA and the dimer–DNA complex (Fig. 2C). The G23Ad/E24–CRGB21 complex was as stable as the E24Ad/G23Cd–CRGB21 complex (Table 2). Heterodimers G23Ad/E24Cd and E24Ad/G23Cd bound the CRGB sequence

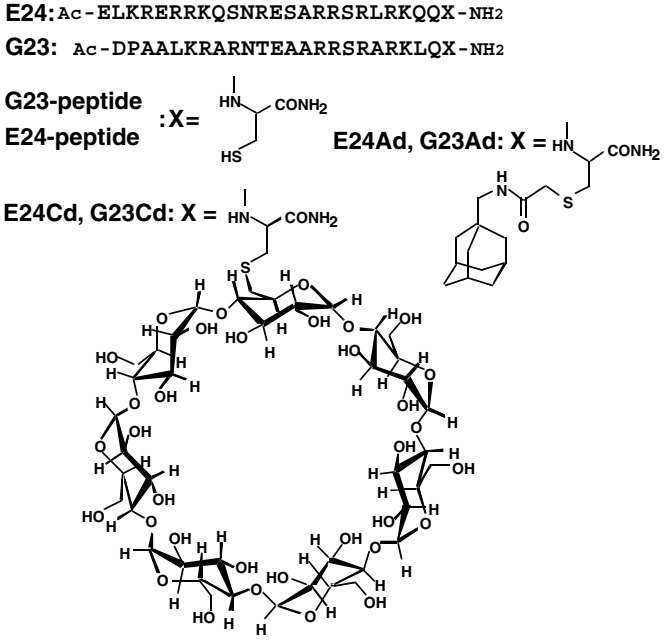

Figure 1. (A) Amino acid sequences for the E and G peptides and structures of the modified C-terminal Cys residue (X). The cysteine residues were modified with 6-deoxy-6-iodo-β-cyclodextrin for E24Cd and G23Cd, and with *N*-bromoacetyl-1-aminoadamantane for E24Ad and G23Ad. Abbreviations for the amino acid are: A, Ala; C, Cys; D, Asp; E, Glu; K, Lys; L, Leu; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr.

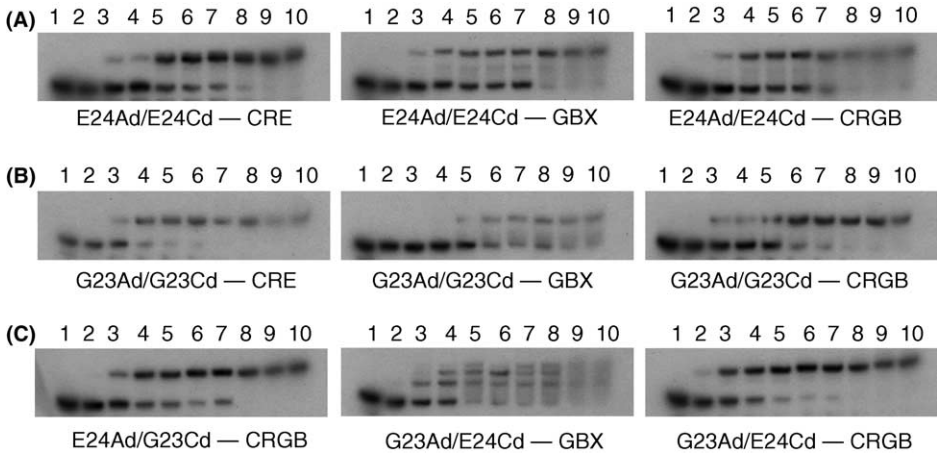

Figure 2. Autoradiograms showing the electrophoretic mobility shift titrations of (A) E24Ad/E24Cd to CRE, GBX, and CRGB DNAs, (B) G23Ad/G23Cd to CRE, GBX, and CRGB DNAs, and (C) E24Ad/G23Cd to CRGB and G23Ad/E24Cd to GBX and CRGB DNAs. Binding reactions were carried out as described in the Methods. No peptide was added to the reaction in lane 1. The peptide concentrations in lanes 2–10, respectively, were 0.5, 1, 2, 3, 4, 6, 10, 20, and 30 nM.

with comparable affinities. The orientation of the G23Ad/E24Cd dimerization domain affects the binding of the G23/E24 heterodimers to the CRE and GBX sequences but not to the CRGB sequence.

3.3. Evaluation of sequence-specific DNA binding of homo- and heterodimers to a non-palindromic CRGB sequence by DNase I footprinting

The sequence-specific DNA binding of G23Ad/E24Cd and E24Ad/G23Cd to the non-palindromic CRGB sequence was further characterized by DNase I footprinting[38] using a duplex of 46-mer oligonucleotide containing the CRGB sequence (Fig. 4). Distinct footprinting patterns were observed at the whole CRGB sequence with G23Ad/E24Cd and G23Cd/E24Ad for both DNA strands.

3.4. CD spectral analyses of DNA bound peptides

Structures of the peptide in dimer–DNA complexes were analyzed by circular dichroism (CD) spectroscopy in the presence of CRE21, GBX21, and CRGB21 (Fig. 5). The basic region of GCN4 has disordered structure in the absence of specific DNA sequence, but is structured upon binding to a specific DNA sequence. As already reported, the CD signal at 222 nm, a measure of α-helicity, was increased significantly on addition of CRE21 to G23Ad/G23Cd[27,28] (Fig. 5A). G23 in the binding complexes with CRGB21 was less helical than that with CRE21. No obvious change in the intensity of the CD signal at 222 nm was observed on addition of GBX21. E24Ad/E24Cd also showed an increase in the intensity of helical band in the presence of GBX21 (Fig. 5B). However, E24Ad/E24Cd exhibited even higher intensity of the helical band in the complexes with CRE21 or CRGB21. Because no such a change in the difference CD spectrum was observed on addition of NON (not shown), which contained neither CRE nor GBX half-site, the E24 basic region peptide changed its structure upon binding to the specific DNA sequences containing the CRE or GBX half-site. The heterodimer E24Ad/G23Cd revealed similar characteristics on the increase of the helical band (Fig. 5C). Intensities of the helical band for E24Ad/G23Cd–CRE21 and –CRGB21 complexes were almost equal, while that of the E24Cd/G23Ad–GBX21 complex was less than those for the former two complexes. However, such a helix formation was not observed in the presence of NON. The same tendency was observed for G23Ad/E24Cd–CRE21, –CRGB21 and –GBX21 complexes (Fig. 5D).

4. Discussion

4.1. Sequence-selectivity of the E24 homodimer

The preferential binding sequences of bZIP proteins are usually either a pseudopalindromic site where the two half-sites overlap at a central base pair or a palindromic site. The best studied bZIP is the yeast transcriptional activator GCN4, which binds the pseudopalindromic AP1 sequence (5′-ATGACTCAT-3′) and palindromic CRE sequence (5′-ATGACGTCAT-3′).[7–10] The CAAT/enhancer binding protein (C/EBP) recognizes the DNA sequences where the two 5′-ATTGC-3′ half-sites about one another.[34] The bZIP proteins from plants exhibit a relaxed DNA-binding specificity for three classes of DNA sequences: G box (5′-CACGTG-3′), C box (5′-GACGTC-3′), and A box (TACGTA-3′).[36]

Native EmBP1 dimer has been shown to bind G box 100 times more preferentially over C Box.[36] The EmBP1 derived basic region peptide dimer E24Ad/E24Cd cooperatively binds GBX21 containing the G box sequence (Fig. 2A). The homodimer E24Ad/E24Cd binds the CRE21 containing the C box sequence as efficiently as GBX21. E24Ad/E24Cd shows slightly higher affinity (0.1 kcal mol^{-1}) for CRE21 over GBX21, and G23Ad/G23Cd shows 2.8 kcal mol^{-1} higher affinity for CRE21 over GBX21. Because E24Ad/E24Cd did not form a

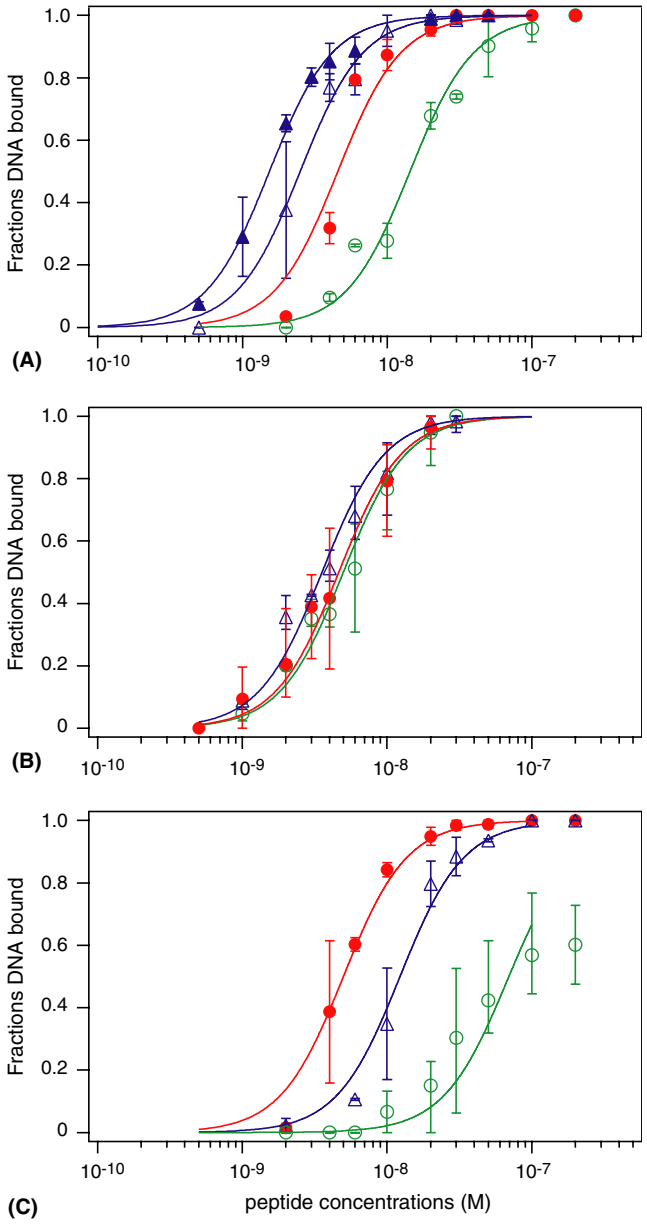

Figure 3. (A) Heterodimers of E24 and G23 peptides preferentially bound the non-palindromic CRGB DNA sequence. Semilogarithmic plots show the fractions of ^{32}P-labeled GBX (green open circles), CRE (red filled circles), CRGB (blue open triangles) bound to E24Ad/G23Cd, and the fraction of ^{32}P-labeled CRGB bound to G23Ad/E24Cd (blue filled triangles) as a function of peptide concentrations. (B) The E24 peptide homodimer almost equally bound GBX, CRE, and CRGB DNA sequences. Semilogarithmic plots show the fraction of ^{32}P-labeled GBX (green open circles), CRE (red filled circles), CRGB (blue open triangles) bound to E24Ad/E24Cd as a function of peptide concentrations. (C) The G23 peptide homodimer preferentially bound the CRE DNA sequence. Semilogarithmic plots show the fraction of ^{32}P-labeled GBX (green open circles), CRE (red filled circles), CRGB (blue open triangles) bound to G23Ad/G23Cd as a function of peptide concentrations. The data points represent the average of three experiments. The sigmoidal curves represent the titration binding isotherms calculated from Eq. 5 using the mean dissociation constants obtained from individual experiments.

stable dimer complex with the unrelated sequence NON, the observed recognition mode reflects intrinsic binding character of the E24 basic region peptide, not indicating

Table 1. Equilibrium dissociation constants K_D and the binding free energy for the G23Ad/G23Cd and E24Ad/E24Cd complexes with CRE21, GBX21, and CRGB21

	K_D (M^2)[a]	ΔG (kcal mol^{-1})[b]	$\Delta\Delta G$[c]
G23Ad/G23Cd–CRE	$2.53 \pm 0.30 \times 10^{-17}$	-21.0 ± 0.1	—
G23Ad/G23Cd–GBX	$4.48 \pm 1.37 \times 10^{-15}$	-18.2 ± 0.2	+2.8
G23Ad/G23Cd–CRGB	$1.51 \pm 0.20 \times 10^{-16}$	-20.1 ± 0.1	+0.9
E24Ad/E24Cd–CRE	$2.18 \pm 0.91 \times 10^{-17}$	-21.1 ± 0.2	—
E24Ad/E24Cd–GBX	$2.52 \pm 0.30 \times 10^{-17}$	-21.0 ± 0.1	+0.1
E24Ad/E24Cd–CRGB	$1.28 \pm 0.16 \times 10^{-17}$	-21.4 ± 0.1	−0.3

[a] Determined by electrophoretic mobility shift titration data (K_{D1} was fixed to 1.34×10^{-6}).
[b] Calculated for 277 K.
[c] Calculated from the relationships, $\Delta\Delta G = \Delta G_{GBX}$ (or ΔG_{CRGB}) − ΔG_{CRE}.

Table 2. Equilibrium dissociation constants K_D and the binding free energy for the G23Ad/E24Cd and E24Ad/G23Cd complexes with CRE21, GBX 21, and CRGB21

	K_D (M^2)[a]	ΔG (kcal mol^{-1})[b]	$\Delta\Delta G$[c]
G23Ad/E24Cd–CRE	[d]		
G23Ad/E24Cd–GBX	[d]		
G23Ad/E24Cd–CRGB	$2.35 \pm 0.14 \times 10^{-18}$	-22.3 ± 0.0	
E24Ad/G23Cd–CRE	$2.06 \pm 0.45 \times 10^{-17}$	-21.1 ± 0.1	—
E24Ad/G23Cd–GBX	$2.02 \pm 0.27 \times 10^{-16}$	-19.9 ± 0.1	+1.2
E24Ad/G23Cd–CRGB	$6.25 \pm 0.65 \times 10^{-18}$	-21.8 ± 0.1	−0.7

[a] Determined by electrophoretic mobility shift titration data (K_{D1} was fixed as 1.34×10^{-6}).
[b] Calculated for 277 K.
[c] Calculated from the relationships, $\Delta\Delta G = \Delta G_{GBX}$ (or ΔG_{CRGB}) − ΔG_{CRE}.
[d] Accurate quantification of band intensity was impossible because several kinds of subbands were observed except free DNA and the dimer-DNA complex bands.

a non-specific binding of E24Ad/E24Cd. Preferential DNA sequences of the basic region peptide dimers are changed from the parent proteins by using dimerization modules different from the native one. Subtle differences in the structures of leucine-zipper domain and the hinge region, which connects the basic region and the leucine-zipper domain, are critical for the sequence-selectivity of bZIP proteins.

4.2. Sequence-selectivity of the E24 and G23 heterodimers

The heterodimers E24Ad/G23Cd and E24Cd/G23Ad bind the non-palindromic CRGB sequence more efficiently than the CRE and GBX sequence. The footprinting data confirmed that both E24Ad/G23Cd and E24Cd/G23Ad specifically bound at the CRGB sequence.

G23Ad/E24Cd formed mobility-shifted bands corresponding not only to the dimer–GBX21 complex but also to the peptide monomer–GBX21 complex (Fig. 2C). Such a non-cooperative binding mode was also observed for the G23Ad/E24Cd–CRE21 complex formation. The monomer–GBX21 complex is likely to be the E24Cd–GBX21 complex because G23Ad/G23Cd barely binds GBX21 and the affinity of monomeric E24 to

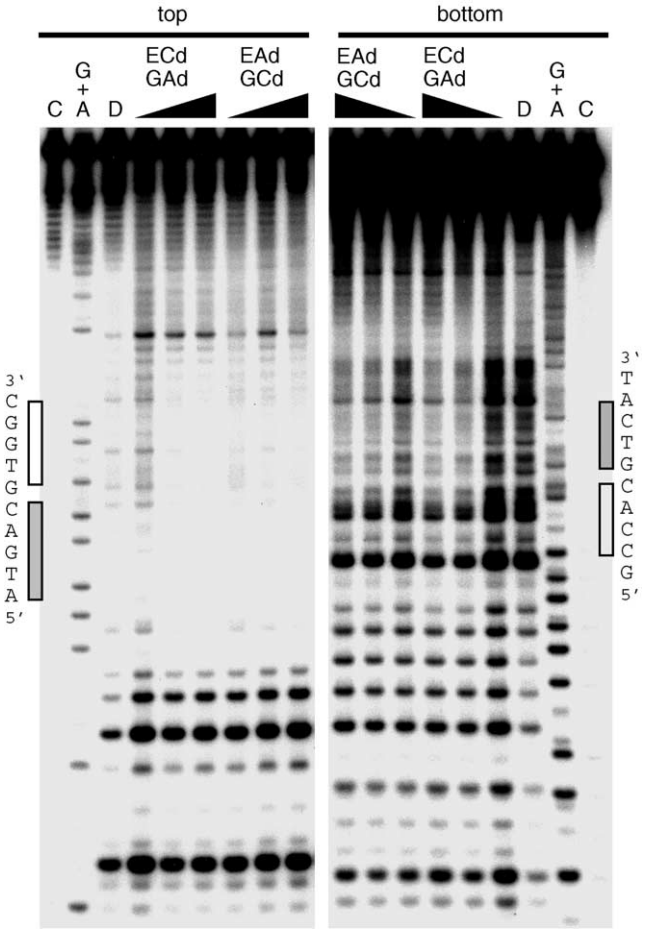

Figure 4. DNase I footprinting experiments with heterodimers E24Ad/G23Cd and G23Ad/E24Cd on the 5′-^{32}P-labeled CRGB46 at the top and at the bottom strand: left half lanes labeled at top strand; right half lanes labeled at bottom strand; lanes C, intact DNA; lanes G+A, G+A chemical reaction marker; lanes D, DNase I standard; lanes G23Ad/E24Cd, 100, 200, and 300 nM G23Ad/E24Cd, respectively; lanes E24Ad/G23Cd, 100, 200, and 300 nM E24Ad/G23Cd, respectively. Sequences for the binding site, 5′-ATGACGTGGC-3′ and 5′-GCCACGTCAT-3′, are shown on the left and right sides of the autoradiogram.

GBX21 is higher than that of G23 to CRE21. Such a monomer–DNA complex was not observed for the binding mixture containing G23Ad/G23Cd and CRE21 where only the cooperative formation of the G23Ad/G23Cd–CRE21 ternary complex was observed. Thus the binding reaction of G23Ad/E24Cd to GBX21 is considered as a stepwise formation of the dimer on the DNA.[39] Alternatively, the stability of the monomer–GBX21 complex may be comparable to that of the half-specific heterodimer–GBX21 complex.

4.3. Structure of the E24 peptide with the CRE and GBX sequences

The basic region peptides are folded into stable α-helices only upon binding to the specific DNA sequence. Such coupling of the α-helix formation to the site-specific protein–DNA interaction can be represented by the example of the GCN4 that has been proposed as a model for the induced-fit binding.[40–43] The half-specific binding complex of the basic region peptide dimer was distinguished from its specific binding complex by comparing the secondary structure of peptides by CD spectral analysis of the binding complexes.[28] In the case of the heterodimer of GCN4- and C/EBP-derived basic region peptides, a matched and a half-matched complex are different in the structure.[28] Two basic region peptides are in the α-helical structure in the matched complex, whereas only one of the basic region peptides formed an α-helical structure in the half-matched complex.

The difference CD spectrum indicates that both G23 peptides in G23Ad/G23Cd become almost 100% helical upon binding to CRE21 (Fig. 5A), as has been observed for native GCN4 and GCN4 basic region peptide dimers.[27,28,41] E24Ad/E24Cd binds in the helical conformation to GBX21, but it also binds in the helical conformation to CRE21 and CRGB21 (Fig. 5B). This is not the case for G23Ad/G23Cd in the presence of GBX21or CRGB21, as the fully helical structure was only observed in the matched G23Ad/G23Cd–CRE21 complex (Fig. 5A). In addition, the results that E24Ad/E24Cd shows comparable affinities to CRE21, GBX21 and CRGB21 DNAs indicate that the E24Ad/E24Cd–CRE21, –GBX21 and –CRGB21 complexes are all considered to be specific complexes.

Increase in the intensity of negative CD band at 222 nm indicates that E24Ad/G23Cd and E24Cd/G23Ad bind in the α-helical conformation to CRGB21. The α-helical contents of E24Ad/G23Cd and E24Cd/G23Ad in the presence of CRE21 are almost equal to that of E24Ad/G23Cd– or E24Cd/G23Ad–CRGB21 (Fig. 5C and D). These CD results indicate that the E24 peptide fits into the CRE and G box half-sites in the helical conformation. Interestingly, intensity of the negative CD band at 222 nm of E24Ad/E24Cd increases slightly upon addition of the oligonucleotide containing C/EBP sequence, even though E24Ad/E24Cd binds to the C/EBP sequence inefficiently (not shown). Such an increase in the helix contents induced by a non-specific DNA has been reported for the Lys-Ala containing peptides.[44] The helix-forming tendency of the E24 peptide induced by the cancellation of positive charges of the peptide by the negatively charged phosphate backbone is higher than that of the G23 peptide. It is possible that the higher helix-forming character accounts for the observed relaxed selectivity of the E24 peptide dimers.

Recognition helices are known to bind unique DNA sequences, as have been demonstrated for the engrailed homeodomain to 5′-TAAT-3′[45,46] GCN4, c-fos, c-jun to AP-1, and CRE sequences containing 5′-TGAC-3′.[8–10,47] The dual recognition mode of the EmBP1 recognition helix targets 5′-TGAC-3′/5′-GTCA-3′ and 5′-CCAC-3′/5′-GTGG-3′ duplexes, indicating that a set of amino acid sequences specifically matches to two kinds of DNA sequences. Possible interactions between the amino acid residues of EmBP1 basic region and the nucleic acid bases of GBX (Fig. 6A) and CRE (Fig. 6B) sequences are summarized in Figure 6. Our results confirm the notion that there exists no absolute code

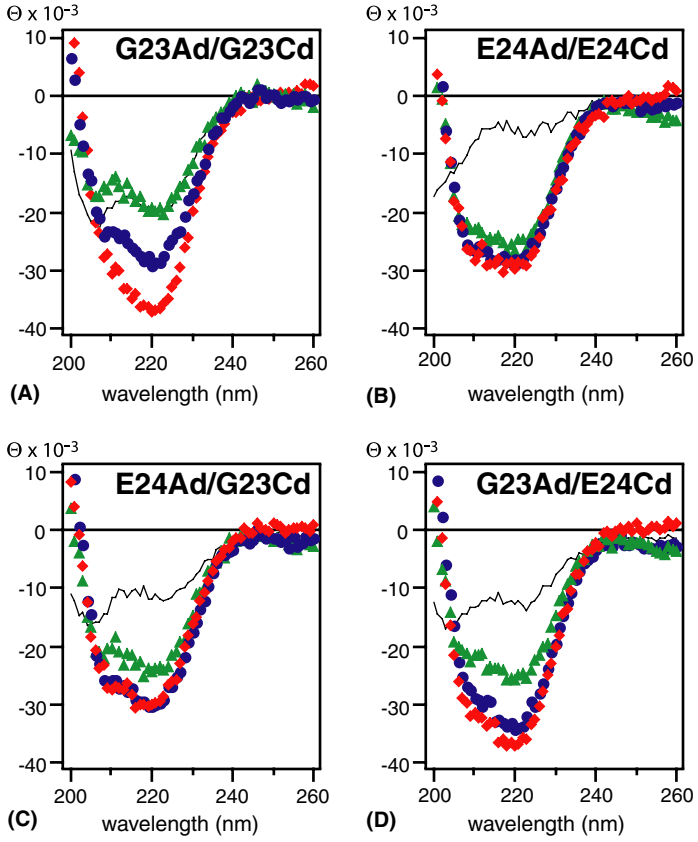

Figure 5. Circular dichroism difference spectra for homo- and heterodimers in the absence or presence of various DNA indicate that the helical content of each peptide varies with the DNA sequence for G23Ad/G23Cd (A), E24Ad/E24Cd (B), E24Ad/G23Cd (C), and G23Ad/E24Cd (D). Spectra in the absence of DNA (solid line), and in the presence of CRE (filled red squares), GBX (filled green triangles), and CRGB (filled blue circles) were calculated as the difference between the bound spectrum and a spectrum of the respective free oligonucleotide CD spectra were obtained as described in Methods.

between the amino acid sequences and nucleotide sequences, and suggest that even a surface defined by a short α-helix will not specifically match to a unique major groove surface provided by 3 or 4 base pairs of DNA. In native bZIP proteins, both the basic region and the leucine zipper dimerization region contribute to recognize specific DNA sequences. Importance of the shape of dimerization domain is highlighted in the case of short peptide dimers. In our previous report, a short peptide dimer with a chiral bridged-biphenyl dimerization domain effectively marked reduction in the affinity of the peptide dimer to non-specific sequences.[26] In other report, it has been suggested that the structural adaptability of the hinge region is important for the discrimination between the AP1 and CRE sequences (half-site spacing specificity) by GCN4.[20,21,48–54] Studies correlating specific DNA binding sites with the function of the basic, the hinge, and the leucine zipper regions, by using chimeric bZIP proteins[12,17,55,56] and the result of in vivo selection[57] also indicate the importance of the hinge region for the specific DNA binding by the bZIP domain. With the hinge region spanning the basic region and leucine-zipper domain, the native EmBP1 homodimer effectively discriminate the GBX sequence over the CRE sequence. The E24 peptide, as a monomeric recognition helix, likely possesses dual recognition codes for two sets of 4-bp DNA sequences.

5. Conclusions

The basic region peptide derived from a bZIP protein EmBP1 recognized palindromic or non-palindromic DNA sequences upon formation of the homo- or heterodimers by the β-Cd/Ad dimerization domain. The EmBP1 basic region peptide dimers bind the DNA sequences containing not only the 5′-GCCAC-3′ half-site but also the 5′-ATGAC-3′ half-site in the α-helical conformation. The GCN4 basic region peptide binds the 5′-ATGAC-3′ half-site in the α-helical conformation, but it neither shows high affinity nor helix formation with the 5′-GCCAC-3′ half-site. The homodimer of EmBP1 peptide shows almost equal affinity to palindromic DNA sequences comprising of 5′-ATGAC-3′ and 5′-GCCAC-3′. Because native EmBP1 forms 100 times more stable complex with the 5′-GCCAC-3′ palindrome (GBX) over the 5′-ATGAC-3′ palindrome (CRE) sequence, the sequence-selectivity of bZIP proteins seems to be highly dictated by the structure of the zipper domain and the amino acids locate in between the basic region and the leucine-zipper domain. A helical short peptide possesses potential advantage as a scaffold for the sequence-specific contacts of amino acid side chains in the major groove. In order to generate a helical short peptide that binds a specific DNA sequence, it is necessary to define its target DNA sequence by the selected and amplified binding site imprinting technique.[58,59] Alternatively, a

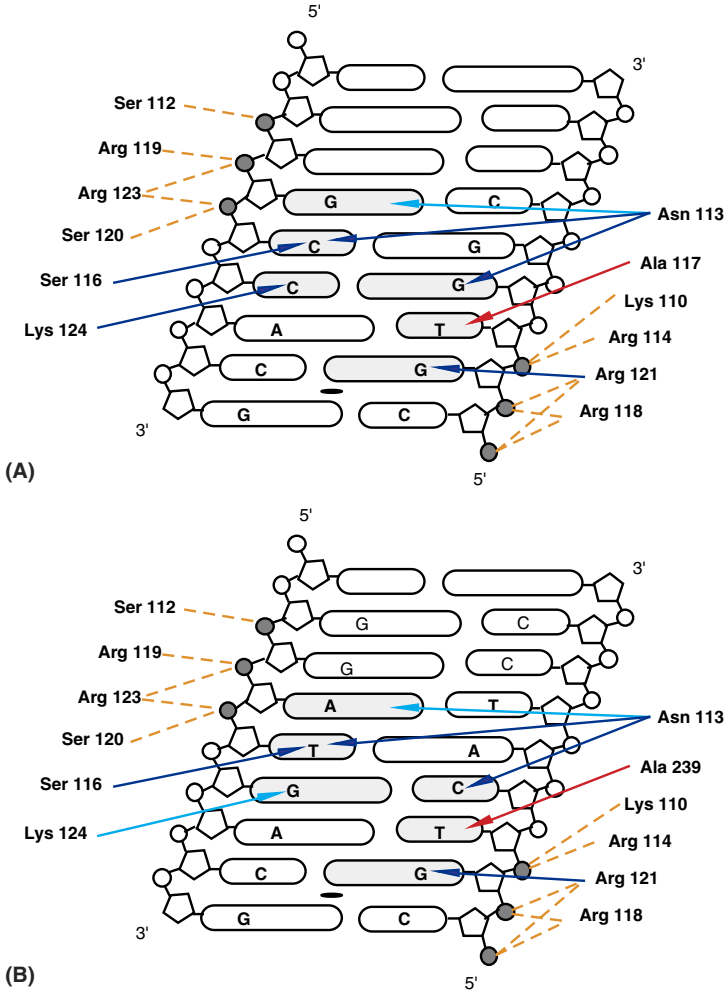

Figure 6. Possible protein–DNA interactions for the basic domain of EmBP1 and the GBX sequence (A) and CRE sequence (B) in the major groove. The interactions between the amino acid residues of the basic domain of EmBP1 and nucleic acids are speculated based on the interactions observed in the X-ray study of GCN4 and the CRE sequence.[8,9]

combined approach using structurally well-defined dimerization module and a peptide library to select specific amino acid residues would be necessary for the design of short peptide dimers targeting specific DNA sequences.

Acknowledgements

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. 15030220 and No. 16651108).

References and notes

1. Harrison, S. C.; Aggarwal, A. K. *Annu. Rev. Biochem.* **1990**, *59*, 933–969.
2. Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053–1095.
3. Aggarwal, A. K. *Curr. Opin. Struct. Biol.* **1997**, *5*, 11–19.
4. Patikoglou, G.; Burley, S. K. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 289–325.
5. Pabo, C. O.; Nekludova, L. *J. Mol. Biol.* **2000**, *301*, 597–624.
6. Hurst, H. C. *Protein Profile* **1994**, *1*, 123–168.
7. Hinnebusch, A. G. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6442–6446.
8. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223–1237.
9. König, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139–154.
10. Keller, W.; König, P.; Richmond, T. J. *J. Mol. Biol.* **1995**, *254*, 657–667.
11. Tzamarias, D.; Pu, W. T.; Struhl, K. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2007–2011.
12. Suckow, M.; von Wilcken-Bergmann, B.; Müller-Hill, B. *EMBO J.* **1993**, *12*, 1193–1200.
13. Suckow, M.; von Wilcken-Bergman, B.; Müller-Hill, B. *Nucleic Acids Res.* **1993**, *21*, 2081–2086.
14. Suckow, M.; Madan, A.; Kisters-Woike, B.; von Wilcken-Bergmann, B.; Müller-Hill, B. *Nucleic Acids Res.* **1994**, *22*, 2198–2208.
15. Suckow, M.; Schwamborn, K.; Kisters-Woike, B.; von Wilcken-Bergmann, B.; Müller-Hill, B. *Nucleic Acids Res.* **1994**, *22*, 4395–4404.

16. Koldin, B.; Suckow, M.; Seydel, A.; von Wilcken-Bergmann, B.; Müller-Hill, B. *Nucleic Acids Res.* **1995**, *23*, 4162–4169.
17. Niu, X.; Renshaw-Gegg, L.; Miller, L.; Guiltinan, M. J. *Plant Mol. Biol.* **1999**, *41*, 1–13.
18. Talanian, R. V.; McKnight, C. J.; Kim, P. S. *Science* **1990**, *249*, 769–771.
19. Talanian, R. V.; McKnight, C. J.; Rutkowski, R.; Kim, P. S. *Biochemistry* **1992**, *31*, 6871–6875.
20. Cuenoud, B.; Schepartz, A. *Science* **1993**, *259*, 510–513.
21. Cuenoud, B.; Schepartz, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1154–1159.
22. Park, C.; Campbell, J. L.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9094–9096.
23. Park, C.; Campbell, J. L.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4892–4896.
24. Morii, T.; Shimomura, M.; Morimoto, M.; Saito, I. *J. Am. Chem. Soc.* **1993**, *115*, 1150–1151.
25. Okagami, M.; Ueno, M.; Makino, K.; Shimomura, M.; Saito, I.; Morii, T.; Sugiura, Y. *Bioorg. Med. Chem.* **1995**, *3*, 777–784.
26. Morii, T.; Saimei, Y.; Okagami, M.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* **1997**, *119*, 3649–3655.
27. Ueno, M.; Murakami, A.; Makino, K.; Morii, T. *J. Am. Chem. Soc.* **1993**, *115*, 12575–12576.
28. Ueno, M.; Sawada, M.; Makino, K.; Morii, T. *J. Am. Chem. Soc.* **1994**, *116*, 11137–11138.
29. Morii, T.; Yamane, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* **1996**, *118*, 10011–10017.
30. Aizawa, Y.; Sugiura, Y.; Ueno, M.; Mori, Y.; Imoto, K.; Makino, K.; Morii, T. *Biochemistry* **1999**, *38*, 4008–4017.
31. Aizawa, Y.; Sugiura, Y.; Morii, T. *Biochemistry* **1999**, *38*, 1626–1632.
32. Sato, S.; Hagihara, M.; Sugimoto, K.; Morii, T. *Chem.-Eur. J.* **2002**, *22*, 5067–5071.
33. Morii, T.; Makino, K. *Adv. DNA Sequence Specific Agents* **2002**, *4*, 105–137.
34. Landschultz, W. H.; Johnson, P. F.; Adashi, E. Y.; Graves, B. J.; McKnight, S. L. *Genes Dev.* **1988**, *2*, 786–800.
35. Guiltinan, M. J.; Marcotte, W. R.; Quatrano, R. S. *Science* **1990**, *250*, 267–271.
36. Izawa, T.; Foster, R.; Chua, N.-H. *J. Mol. Biol.* **1993**, *230*, 1131–1144.
37. Foster, R.; Izawa, T.; Chua, N.-H. *FASEB J.* **1994**, *8*, 192–200.
38. Galas, D. J.; Schmitz, A. *Nucleic Acids Res.* **1978**, *5*, 3157–3170.
39. Park, C.; Campbell, J. L.; Goddard, W. A., III. *J. Am. Chem. Soc.* **1996**, *118*, 4235–4239.
40. Frankel, A. D.; Kim, P. S. *Cell* **1991**, *65*, 717–719.
41. Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature* **1990**, *347*, 575–578.
42. Alber, T. *Curr. Biol.* **1993**, *3*, 182–184.
43. Spolar, R. S.; Record, M. T., Jr. *Science* **1994**, *263*, 777–784.
44. Johnson, N. P.; Lindstrom, J.; Baase, W. A.; von Hippel, P. H. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4840–4844.
45. Kissinger, C. R.; Liu, B. S.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579–590.
46. Fraenkel, E.; Rould, M. A.; Chambers, K. A.; Pabo, C. O. *J. Mol. Biol.* **1998**, *284*, 351–361.
47. Glover, J. N.; Harrison, S. C. *Nature* **1995**, *373*, 257–261.
48. Johnson, P. F. *Mol. Cell. Biol.* **1993**, *13*, 6919–6930.
49. Kim, J.; Tzamarias, D.; Ellenberger, T.; Harrison, S. C.; Struhl, K. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4513–4517.
50. Paolella, D. N.; Palmer, C. R.; Schepartz, A. *Science* **1994**, *264*, 1130–1133.
51. Metallo, S. J.; Schepartz, A. *Chem. Biol.* **1994**, *1*, 143–151.
52. Metallo, S. J.; Paolella, D. N.; Schepartz, A. *Nucleic Acids Res.* **1997**, *25*, 2967–2972.
53. Paolella, D. N.; Liu, Y.; Fabian, M. A.; Schepartz, A. *Biochemistry* **1997**, *36*, 10033–10038.
54. Kim, J.; Struhl, K. *Nucleic Acids Res.* **1995**, *23*, 2531–2537.
55. Agre, P.; Johnson, P. F.; McKnight, S. L. *Science* **1989**, *246*, 922–926.
56. Haas, N. B.; Cantwell, C. A.; Johnson, P. F.; Burch, J. B. E. *Mol. Cell. Biol.* **1995**, *15*, 1923–1932.
57. Sera, T.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2920–2925.
58. Blackwell, T. K.; Weintraub, H. *Science* **1990**, *250*, 1104–1110.
59. Morii, T.; Tanaka, T.; Sato, S.; Hagihara, M.; Aizawa, Y.; Makino, K. *J. Am. Chem. Soc.* **2002**, *124*, 180–181.