Synthesis of a new zinc finger peptide; comparison of its 'code' deduced and 'CASTing' derived binding sites

Nicoletta Corbi[a], Marie Perez[a], Rossella Maione[a], Claudio Passananti[b,c,*]

[a] *Dipartimento di Biotecnologie Cellulari ed Ematologia, Università di Roma La Sapienza, Viale Regina Elena 324, 00161 Rome, Italy*
[b] *Istituto di Tecnologie Biomediche, CNR, Via Morgagni 30/E, 00161 Rome, Italy*
[c] *Cell Metabolism and Pharmacokinetics Laboratory, Regina Elena Cancer Institute, Via delle Messi d'Oro 156, 00158 Rome, Italy*

Received 28 August 1997; revised version received 23 September 1997

Abstract Using two synthetic oligonucleotides, we have constructed a new gene containing three zinc finger motifs of the Cys_2-His_2 type. We named this artificial gene 'Mago'. The Mago nucleotide triplets encoding the amino acid positions, described to be crucial for DNA binding specificity, have been chosen on the basis of the proposed recognition 'code' that relates the zinc finger's primary structure to the DNA binding target. Here we demonstrate that Mago protein specifically binds the 'code' DNA target, with a dissociation constant (K_d) comparable to the K_d of the well known Zif268 protein with its binding site. Moreover, we show that the deduced Mago 'code' and the 'experimental' selected DNA binding sites are almost identical, differing only in two nucleotides at the side positions.

Key words: Zinc finger-DNA interaction; Protein design; Recognition code; Casting selection; Dissociation constant

1. Introduction

The ability to design DNA binding proteins able to recognize specific target sequences can be a powerful tool in research, biotechnology, and medicine. Various strategies have been reported for generating DNA binding proteins with new or altered target specificity [1–6]. The zinc finger of the class Cys_2-His_2 appears to be optimal for this purpose because of its relatively simple and versatile structure. It is a small motif of 28–30 amino acids, folded into a compact globular module, that comprises an α-helix containing two invariant histidine residues coordinated through a zinc atom to two cysteines of a single β-turn [7,8]. The zinc finger domain, originally discovered in the transcription factor TFIIIA [9], has been subsequently found in a huge variety of regulatory proteins containing different numbers of tandem arrays of zinc finger domains, coupled with distinct functional domains [7,10–13]. The X-ray crystal structures of the three zinc finger domains of the transcription factor Zif268 bound to a DNA target site reveal that the individual finger domains bind DNA with specific contacts through the amino-terminal part of the α-helix [14]. Zinc finger protein recognition of DNA involves an antiparallel arrangement of protein in the major groove of DNA, the amino-terminal region is involved in 3′ contacts with the target sequences, whereas the carboxyl-terminal region is involved in 5′ contacts. Each finger domain appears to contact three adjacent base pairs of DNA mainly through three key residues, located in positions −1, +3 and +6 of the α-helix. In addition the amino acid at position +2 appears to play an auxiliary role enhancing the specificity of the amino acid at other positions. To design new zinc finger proteins, each finger domain can be altered in these specific amino acid positions and multiple zinc fingers can be 'mixed and matched' to obtain novel DNA binding specificity [1,8,15–17]. In particular, a 'code' that relates the amino acids of a single zinc finger to its associated DNA target has been proposed for a variety of zinc finger domains. This 'code', initially predicted to be 'alphabetic', with an invariant corresponding amino acid/base pair, is now described to be 'syllabic', depending on structural characteristics of every single zinc finger module and the context in which it is inserted [15–21]. Using and expanding this restricted 'code' several research groups have recently engineered and selected peptides containing zinc finger domains able to bind and recognize specific DNA sequences [1,15–23]. Zinc finger proteins related to TFIIIA and Zif268 appear to provide the most versatile framework for design. Taking advantage of these data, we have designed and engineered a new three zinc finger peptide, named 'Mago', able to bind the 'code' predicted DNA sequence: 5′-ATG TGG GTT-3′.

2. Materials and methods

2.1. Construction of the Mago gene

Using as a model the three zinc finger peptide backbone proposed by Choo et al. [1] we synthesized two overlapping oligonucleotides named '5p' (sense) and '3p' (antisense), respectively 176 and 175 nucleotides long. The synthesis was performed with a standard scale of 40 nmol. 5 μg of each oligonucleotide was annealed, by the 20 nt long overlapping region, in a total volume of 50 μl in the following buffer: 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, and 7.5 mM DTT. The reaction was heated to 75°C and slowly cooled to 37°C, then 4 μl of 10 mM dNTP and 2 μl (10 U) of Klenow (BioLabs) were added and incubated for 20 min at 37°C. 1 μl of the resulting mixture was used as template for the following amplification reaction: (94°C-30 s, 65°C-30 s, 72°C-1 min)×32 cycles, +10 min at 72°C, using as primers two 24 nt long oligonucleotides (5′-CCCAAGCTTGGATCCATGTATC-CA-3′ and 5′-CCCAAGCTTAGATCTACTCGAGTT-3′), located at the ends of the 322 bp long DNA fragment (Fig. 1). 10 μl of the amplification reaction was checked on a 1.5% agarose gel; the remaining part was phenol extracted, ethanol precipitated, resuspended in 47 μl of H_2O and digested in the appropriate conditions with *Bam*HI and *Bgl*III restriction enzymes (BioLabs). Finally, 100 ng of a treated DNA fragment was ligated in the pGEX-4T-3 expression vector (Pharmacia Biotech).

We arranged, by PCR amplification, the three zinc finger coding region of the Zif268 gene [24] in the pGEX-4T-3 expression vector. To this end we realized two oligonucleotides located respectively from nt '1294 to 1320' and from nt '1428 to 1454' on the cDNA sequences of Zif268. The two oligos were designed with two extra restriction sites at the ends, compatible with the cloning strategies.

2.2. Electrophoretic mobility shift assay

The electrophoretic mobility shift assay (EMSA) was performed by

*Corresponding author. Fax: (39) (6) 4985.2505.
E-mail: Passananti@ifo.it

PII S0014-5793(97)01257-X

incubating 0.5–1 ng of oligonucleotide probe and purified GST proteins (0.5–2 μg), for 30 min at room temperature in the presence of the binding buffer: 20 mM HEPES (pH 7.9), 100 mM KCl, 1 mM DTT, 2 mM $MgCl_2$, 0.1% NP40, 10 μM $ZnCl_2$, 10% glycerol, 50 μg/ml BSA, 100 ng of poly(dI-dC). After incubation the mixture was loaded onto a 5% polyacrylamide gel (30:1 acrylamide/bisacrylamide) and run in 0.5×TBE running buffer at room temperature. The DNA sequences of the oligonucleotide probes were: 5′-ATTAACCATGTGGGTTGGCTGCTCC-3′ (Mago 'code' deduced DNA binding site) and 5′-ATTAACCGCGTGGGCGGGCTGCTCC-3′ (Zif268 DNA binding site). Unlabeled competitor oligonucleotides were added, in the reaction mixture, at 100-, 50- and 25-fold molar excess of the amount of labeled DNA. The mutant oligonucleotide competitors were completely mutagenized in the core region (underlined) corresponding to the DNA binding sites of both Mago and Zif268. All the oligonucleotides used as probes have TTT 5′ overhanging from one strand, and were labeled using a terminal Klenow reaction [23].

2.3. Filter binding

To derive the apparent dissociation constant (K_d) we used the filter binding analysis as described by Stockley [25]. Basically we determined the fraction of labeled DNA oligonucleotides bound at a series of peptide concentrations (GST-Mago and GST-Zif268). Binding reactions containing oligonucleotide probes (20 000 cpm, corresponding to <1 ng), peptide (from a dilution series between 10^{-10} M and 10^{-6} M) and binding buffer (B) (20 mM HEPES (pH 7.9), 100 mM KCl, 0.5 mM DTT, 50 μg/ml BSA, 10 μM $ZnCl_2$) were incubated in a final volume of 50 μl for 30 min at room temperature. Then, each sample, diluted to 100 μl with buffer B, was applied on a Millipore 1225 filter (presoaked for several hours in filter binding buffer: 20 mM HEPES (pH 7.9), 100 mM KCl, 0.2 mM EDTA), and placed on a filtration apparatus connected to a vacuum pump. After a quick wash with 5 ml of buffer B, each filter was removed, dried and placed in a vial with 3 ml of scintillation liquid. Radioactivity associated with each sample was determined by counting on an open channel. The value of each sample was corrected by subtracting the count of a background sample (no protein). A freshly thawed sample of the peptide was used for each set of experiments. Each experiment was performed in duplicate.

2.4. DNA sequencing

Every construct generated and presented in this paper has been checked by nucleotide sequence analysis, performed using the Sequenase kit according to the manufacturer's instructions (United States Biochemical).

2.5. Expression of recombinant fusion protein

Mago and Zif268 DNA fragments were inserted into the *Bam*HI site of pGEX-4T-3 expression vector (Pharmacia Biotech). Each recombinant molecule was expressed in BL21 (DE3) host bacteria by IPTG induction and purified using glutathione-agarose beads [26]. The isolation of each recombinant protein was checked by Coomassie blue staining of SDS polyacrylamide gels.

2.6. CASTing DNA target selection

The CASTing DNA target selection was performed following the methods reviewed by Wright and Funk [27]. In particular, a 44 base oligonucleotide was synthesized to contain a central region of 12 random bases flanked by 16 base regions with the defined sequence: 5′-ACGCGTCGACGGATCC(A/C/G/T)$_{12}$AAGCTTGTCGACACG-C-3′. To yield a mixture of double-stranded DNA fragments enriched in each possible permutation, one partially overlapping oligonucleotide was annealed and extended by AmpliTaq polymerase (Perkin Elmer). This mixture was incubated with GST-Mago fusion protein bound to beads (5 μl) in binding buffer (20 mM HEPES (pH 7.9), 100 mM KCl, 1 mM DTT, 2 mM $MgCl_2$, 0.1% NP40, 10 μM $ZnCl_2$, 100 ng of poly(dI-dC), 10% glycerol, 50 μg/ml BSA) in a final volume reaction of 25 μl, for 30 min [28]. Beads were centrifuged and washed four times with binding buffer and then boiled for 5 min. The oligonucleotide mixture eluted from the beads by boiling was used for PCR amplification with the amplifying oligonucleotides corresponding to the invariable region [28]. After four rounds of selection-amplification, a fraction of the PCR reaction was cloned in the pGEM-T Easy Vector and each individual insert was end-labeled and further examined by EMSA (performed as described above). The inserts showing the most intense signal were then sequenced and aligned.

3. Results

3.1. Design of the Mago gene

Using as a model the three zinc finger peptide backbone proposed by Choo et al. [1], we engineered the artificial gene named 'Mago', directed to target the predicted DNA sequence: 5′-ATG TGG GTT-3′. As shown in Fig. 1, we synthesized a DNA fragment 322 bp long, encoding a peptide containing three zinc finger domains derived from TFIIIA protein [1], and the epitope tag from an influenza hemagglutinin (HA) subtype [29]. The design of the zinc finger region of the Mago gene was based on the available recognition 'code' that relates specific amino acid positions of the α-helix of a single finger domain to its cognate DNA target triplet. In particular, we chose the amino acid/nucleotide base contacts more frequently represented in the list of the syllabic 'code' signatures reported by Choo and Klug [17,30]. We decided to modify only the positions most critical in determining DNA binding specificity (positions −1, +3 and +6 of the α-helix) and to keep invariant the zinc finger peptide backbone proposed by Choo and Klug. Schematically, the resulting structure of the Mago protein was as follows: the third finger domain, at the carboxyl-terminal region, was designed to recognize the DNA triplet 5′-ATG-3′, through the amino acid residues arginine, valine, and asparagine, respectively in the crucial positions of the α-helix −1, +3 and +6; the second finger was designed to target the DNA triplet 5′-TTG-3′ through the amino acid residues arginine, histidine, and threonine (−1, +3 and +6) and finally the first finger was planned to recognize the triplet 5′-GTT-3′ through the amino acid residues asparagine, threonine and arginine (−1, +3 and +6) (Fig. 1).

Moreover, we constructed a second gene encoding the three zinc finger domains of the well known Zif268 protein [24] in order to have a natural model system to compare with the functional properties of the synthetic Mago protein.

3.2. DNA binding activity

The bacterial produced and purified GST-Mago fusion protein was used in EMSA to test its ability to bind in a specific manner the code deduced DNA target sequence (Fig. 2A). A clear shift of the GST-Mago protein was observed in the presence of the labeled oligo containing the 'code' deduced

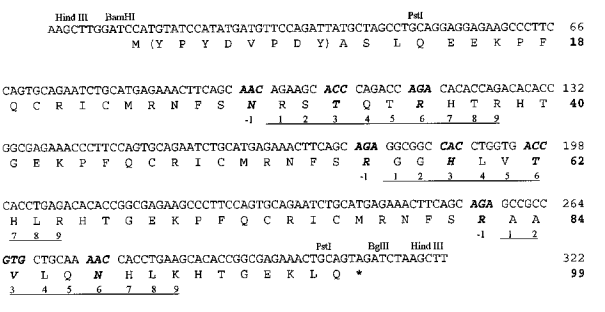

Fig. 1. Nucleotide and amino acid sequences of the synthetic three zinc finger gene, named Mago. The amino acid residues spanning the α-helix regions of each zinc finger are numbered and underlined. The nucleotides and the corresponding amino acid positions (−1, +3 and +6), described to be crucial for DNA binding specificity, are indicated in bold. The nine amino acids corresponding to the epitope tag HA are indicated between parentheses. Restriction enzyme sites used for the construction and cloning strategies of the gene are indicated.

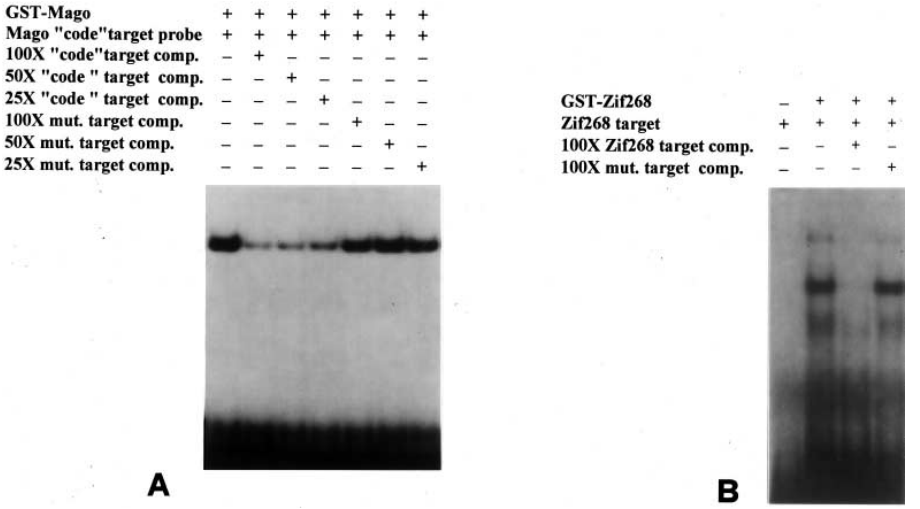

Fig. 2. GST-Mago and GST-Zif268 fusion protein binding activity analyzed by EMSA. A: Binding of GST-Mago to the recognition 'code' DNA target probe (5′-ATG TGG GTT-3′). Competition assays were performed by adding to the reaction mixture a 100-, 50- and 25-fold molar excess of unlabeled 'code' and mutant targets, as indicated above the lanes. B: Binding of GST-Zif 268 protein to its DNA target probe (5′-GCG TGG GCG-3′). Competitions were performed with a 100-fold molar excess of unlabeled DNA target and mutant target, as indicated above the lanes.

target 5′-ATG TGG GTT-3′, indicating a strong DNA/protein interaction. Formation of the complex DNA/GST-Mago was inhibited by competition with the unlabeled 'code' deduced target itself (up to 1:25 dilution), while no competition was observed in the presence of an oligo completely mutagenized in the nine base pair target sequence.

Concomitantly, the GST-Zif268 DNA binding properties were analyzed and compared in EMSA performed in the same conditions described above for GST-Mago protein (Fig. 2B). We used as a probe an oligo containing the Zif268 DNA target sequence 5′-GCG TGG GCG-3′ [31], competed with itself unlabelled and with an oligo completely mutagenized in the nine base pair target sequence. These data together demonstrate that the binding of GST-Mago protein to its 'code' deduced target sequence is efficient and specific, similar to the 'natural' Zif268 protein.

3.3. Filter binding

The high affinity of the GST-Mago peptide for the 'code' deduced target revealed by EMSA was more precisely determined by measuring the relative dissociation constant (K_d) [25]. To derive the apparent K_d, we performed filter binding experiments with both GST-Mago and GST-Zif268 purified proteins, using as a probe their respective targets. Under our experimental conditions (see Section 2) we obtained for the GST-Mago/'code' target complex and the GST-Zif268/DNA complex comparable K_ds of 5 nM and 1 nM respectively, indicating a similar affinity of the two proteins toward their own targets.

3.4. CASTing DNA target selection

We used the GST-Mago protein to pick up the preferred DNA binding sites in a pool of random oligonucleotides. To this end, we performed cyclic amplification and selection of targets (CAST) experiments [27]. Basically, the pool of oligonucleotides obtained after four rounds of selection-amplification was cloned in the pGEM-T Easy Vector and the critical DNA region was amplified and labeled for further screening by EMSA (Fig. 3A). Using this assay, we analyzed 100 individual clones and 10 of them, chosen on the basis of the shift signal intensity, were sequenced. In particular, the DNA se-

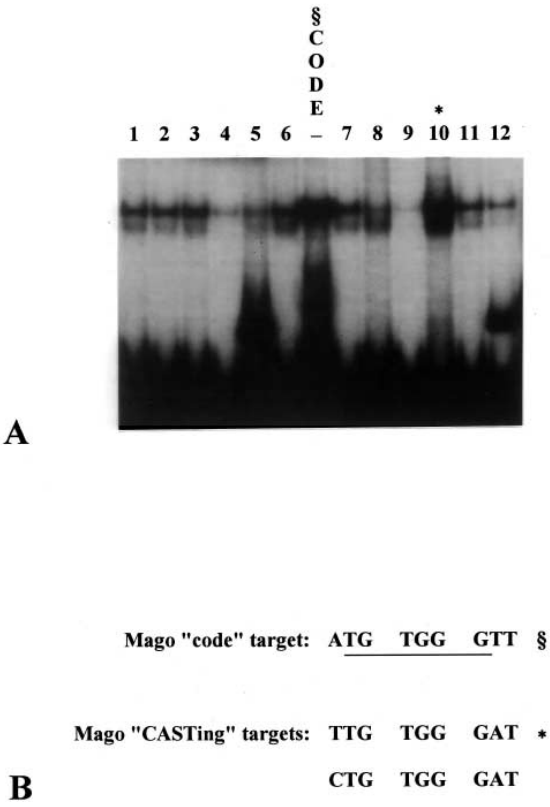

Fig. 3. Typical mobility shift assay of the CASTing selected GST-Mago DNA targets. A: Twelve end-labeled DNA fragments, obtained after four CASTing selection-amplification cycles, tested and compared with the 'code' DNA target probe, for their ability to bind the GST-Mago protein. B: Sequence alignment of Mago 'code' DNA target sequence and two CASTing selected DNA targets corresponding to the most intense gel shift signals. Clone 10 (shown in A) is indicated with an asterisk. The invariable core region of the Mago DNA targets is underlined.

quences of two of them corresponding to the most intense shift were: clone 10, 5′-TTG TGG GAT-3′ (Fig. 3A) and clone 30, 5′-CTG TGG GAT-3′ (data not shown). Alignment of these CASTing Mago DNA binding sites with the 'code' deduced target demonstrates that they are almost identical. In fact, as shown in Fig. 3B, the target sequences share a common core region 6 nt long and differ in only two nucleotides at the side positions.

4. Discussion

The possibility of engineering a DNA binding protein able to target a desired promoter and to exert a regulatory function when fused with appropriate domains is an attractive tool for a large number of applications [1–4]. Zinc finger proteins, because of their structure plasticity and modularity, have been chosen by several research groups as the framework to design new artificial transcription factors [15–23]. In particular, a 'code' that relates the primary structure of a single zinc finger to its associated DNA target has been proposed. This 'code', even if not complete and partially degenerate, can make it possible to generate optimal zinc finger domains for many desired target sequences. We used the list of the recognition 'code' signatures reported by Choo and Klug [17,30] to construct a synthetic gene, named Mago. Here we showed that Mago recognizes with high affinity and specificity the 'code' predicted DNA sequence 5′-ATG TGG GTT-3′. We used different approaches to study Mago protein properties. Mobility shift analysis of GST-Mago purified fusion protein revealed that Mago binds the 'code' DNA target in a specific manner. Moreover, the intensity of the Mago protein shift is comparable with the shift of the well characterized Zif268 protein combined with its natural DNA target. The affinity of the GST-Mago peptide for the recognition 'code' target has been determined by measuring the relative K_d. In our experimental conditions the K_d of Mago protein relative to the 'code' target was comparable to the K_d of Zif268 protein relative to its DNA target, indicating that the complex GST-Mago/'code' target was almost as stable as the GST-Zif268/DNA complex. Finally the CASTing target selection demonstrated that the deduced Mago 'code' and the 'experimental' selected DNA binding sites are almost identical, differing in only two nucleotides at the side positions. In agreement with this finding, the study of the interaction between transcription factors and their DNA targets has pointed out the presence, within many binding sites, of a conserved core region, flanked by nucleotide positions more susceptible to variation [31]. In the cell environment a transcription factor can bind multiple DNA targets with different affinity, exerting an additional level of transcriptional regulation [32]. The calibration of the binding affinity/specificity is one of the critical points in the design of synthetic zinc finger peptides [18,19,23]. Moreover, data accumulated in literature suggest that not only the amino acid positions described to be crucial for the recognition 'code' (−1, +3 and +6 of the α-helix), but also some other amino acid positions, for example position +2 of the α-helix and some context-dependent contacts, can play a role in zinc finger DNA binding properties [20,21,23]. All these speculations taken together suggest that proteins designed on the basis of the available 'code' need to be empirically tested for their DNA binding properties. Mago protein, constructed on the basis of the 'code', binds the selected DNA target with high affinity and specificity. The next step planned for Mago protein is to study its functionality in the appropriate eukaryotic cell system.

Acknowledgements: We are grateful to Prof. Paolo Amati for support and critical discussion. We thank Mrs. L. Tatangelo for valuable technical assistance. This work was supported by TELETHON (Progetto A63).

References

[1] Choo, Y., Sánchez-García, I. and Klug, A. (1994) Nature 372, 642–645.
[2] Pomerantz, J.L., Sharp, P.A. and Pabo, C.O. (1995) Science 267, 93–96.
[3] Wu, H., Yang, W.P. and Barbas III, C.F. (1995) Proc. Natl. Acad. Sci. USA 92, 344–348.
[4] Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T. and DeGrado, W.F. (1995) Science 270, 935–940.
[5] Sera, T. and Schultz, P.G. (1996) Proc. Natl. Acad. Sci. USA 93, 2920–2925.
[6] Rebar, E.J., Greisman, H.A. and Pabo, C.O. (1996) Methods Enzymol. 267, 129–149.
[7] Schmiedeskamp, M. and Klevit, R.E. (1994) Curr. Opin. Struct. Biol. 4, 28–35.
[8] Berg, J.M. and Shi, Y. (1996) Science 271, 1081–1085.
[9] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609–1614.
[10] El-Baradi, T. and Pieler, T. (1991) Mech. Dev. 35, 155–169.
[11] Rosati, M., Marino, M., Franzé, A., Tramontano, A. and Grimaldi, G. (1991) Nucleic Acids Res. 19, 5661–5667.
[12] Passananti, C., Corbi, N., Paggi, M.G., Russo, M.A., Perez, M., Cotelli, F., Stefanini, M. and Amati, P. (1995) Cell Growth Diff. 6, 1037–1044.
[13] Perez, M., Rompato, G., Corbi, N. and Passananti, C. (1996) FEBS Lett. 387, 117–121.
[14] Pavletich, N.P. and Pabo, C.O. (1991) Science 252, 809–817.
[15] Desjarlais, J.R. and Berg, J.M. (1992) Proc. Natl. Acad. Sci. USA 89, 7345–7349.
[16] Desjarlais, J.R. and Berg, J.M. (1994) Proc. Natl. Acad. Sci. USA 91, 11099–11103.
[17] Choo, Y. and Klug, A. (1994) Proc. Natl. Acad. Sci. USA 91, 11168–11172.
[18] Rebar, J.E. and Pabo, C.O. (1994) Science 263, 671–674.
[19] Jamieson, A.C., Kim, S.H. and Wells, J.A. (1994) Biochemistry 33, 5689–5695.
[20] Gogos, J.A., Jin, J., Wan, H., Kokkinidis, M. and Kafatos, F.C. (1996) Proc. Natl. Acad. Sci. USA 93, 2159–2164.
[21] Kim, J.S., Kim, J., Cepek, K.L., Sharp, P.A. and Pabo, C.O. (1997) Proc. Natl. Acad. Sci. USA 94, 3616–3620.
[22] Jamieson, A.C., Wang, H. and Kim, S.H. (1996) Proc. Natl. Acad. Sci. USA 93, 12834–12839.
[23] Greisman, A.H. and Pabo, C.O. (1997) Science 275, 657–661.
[24] Christy, B.C., Lau, L.F. and Nathans, D. (1988) Biochemistry 85, 7857–7861.
[25] Stockley, P.G. (1994) in: Methods in Molecular Biology (Kneale, G.G., Ed.), DNA-Protein Interactions, Vol. 30, pp. 251–262.
[26] Smith, D.B. and Johnson, K.S. (1988) Gene 67, 31–40.
[27] Wright, W.E. and Funk, W.F. (1993) Trends Biochem. Sci. 18, 77–80.
[28] Zweidler-Mckay, P.A., Grimes, H.L., Flubacher, M.M. and Tsichlis, P.N. (1996) Mol. Cell. Biol. 16, (8) 4024–4034.
[29] Wadzinski, B.E., Eisfelder, B.J., Peruski Jr., L.F., Mumby, M.C. and Johnson, G.L. (1992) J. Biol. Chem. 267, (24) 16883–16888.
[30] Choo, Y. and Klug, A. (1994) Proc. Natl. Acad. Sci. USA 91, 11163–11167.
[31] Christy, B. and Nathans, D. (1989) Proc. Natl. Acad. Sci. USA 86, 8737–8741.
[32] Berg, J.M. (1992) Proc. Natl. Acad. Sci. USA 89, 11109–11110.