Volume 278, number 1, 115–119 FEBS 09318 January 1991

ADONIS 001457939100034B

The histone H1-*lacZ'* fusion protein produced in *Escherichia coli* binds to the 5'-TTGGCAnnnTGCCAA-3' motif on DNA

Riitta-Maaria Mannermaa and Jouko Oikarinen

Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, Kajaanintie 52A, SF-90220 Oulu, Finland

Received 22 October 1990; revised version received 12 November 1990

The coding region of the chicken histone H1.03 gene was cloned to a bacterial expression vector, and the 291-amino acid H1-β-galactosidase fusion protein was isolated after induction with IPTG. The fusion protein recognizes the 5'-TTGGCAnnnTGCCAA-3' motif on DNA. The H1 globular domain was initially shown to be responsible for the sequence-specific binding by functional deletion analysis. This function may be indispensable for the role of H1 as a determinant of nucleosome positioning and as a eukaryotic repressor.

Histone H1; Nucleosome; Chromatin; DNA binding; Repression

1. INTRODUCTION

Histone H1 has been thought to function as a general repressor in eukaryotes, being able to aggregate chromatin in an inactive state [1–4]. At least 4–5 subtypes of H1 exist in chromatin in each species, binding to hinge regions of nucleosomes [5], formed by two turns of DNA twined around core histones [6,7].

It has been observed that the 3D structure of the globular domain of histone H5 [8], an erythrocyte-specific H1, displays similarity to the DNA binding domain in CRP, the α-carbon backbones of the two being superimposable. Based on these data and tertiary structure homology of the globular domain with DNA-binding homeodomains, it has been suggested that H1 may be a sequence-specific DNA binding protein [9]. It has recently been demonstrated that H1 binds to a CTF/NF-I recognition sequence in the mouse α_2(I) collagen promoter [10] and to specific sites in the rat albumin promoter [11]. H1 can also be isolated by DNA affinity chromatography using the consensus for CTF/NF-I binding sites, 5'-TTGGCAnnnTGCCAA-3' [12], as the ligand (unpublished results). H1 may therefore be one of the proteins of regulatory importance binding to that motif [13].

The present work was undertaken to prove definitely that histone H1 is a sequence-specific DNA binding protein by studying this function in an H1-*lacZ'* fusion protein produced in *E. coli*.

Correspondence address: J. Oikarinen, Department of Medical Biochemistry, University of Oulu, Kajaanintie 52A, SF-90220 Oulu, Finland

Abbreviations: IPTG, isopropyl-β-D-thiogalactopyranoside; CTF/NF-I, CCAAT transcription factor/nuclear factor-I; PCR, polymerase chain reaction; DMSO, dimethylsulfoxide; EMSA, electrophoretic mobility shift assay; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; PCA, perchloric acid

2. EXPERIMENTAL

2.1. *Cloning and production of histone H1.03 in bacteria*

Chicken genomic DNA was isolated as described elsewhere [14]. The coding region of the H1.03 gene was amplified by PCR [15] with 1 μg of the genomic DNA as the template in the presence of 10% DMSO, using a Perkin-Elmer Cetus DNA Thermal Cycler. Oligonucleotides of the sequences 5'-ACGCCAAGCTTCATGGCT GAGACCGCTCCTGTCGCT-3' and 5'-CGAGCTCGGTACCTT TTTCTTAGGGGCCGCCTTCTT-3', corresponding to a part of the sense strand at the 5' end and to a part of the antisense strand at the 3' end respectively, were used as the primers. These contained 5' overhangs for the *Hin*dIII and *Kpn*I sites respectively. The amplified fragment of 696 bp was purified by PAGE, cut with the two enzymes and ligated to plasmid pUC19 between the *Hin*dIII and *Kpn*I sites. *E. coli* JM109 was transformed by the construct. A deletion mutant lacking the sequence coding for the C-terminal tail of H1 was generated using an oligonucleotide of the sequence 5'-GAGCTCGGTACCTTCTTAGGAGCCTTCTCCTTCAC-3' as the 3' primer and plasmid PCR51 as the template. Plasmid preparations were isolated, and subjected to DNA sequencing by the dideoxy method [16].

Synthesis of the fusion protein was induced by 10 mM IPTG [17], a 10-ml culture giving a yield of about 10 μg. In order to isolate the fusion protein, the bacteria were lysed under the conditions described for plasmid isolation [14]. The lysates were loaded onto an anti-β-galactosidase affinity column under the conditions described for DNA affinity chromatography [10], and the bound material was eluted by raising the pH to 10.8. The eluted material was fractionated and the pH was adjusted to 7.5 as soon as possible.

2.2. *Assays for DNA binding*

The binding of proteins to DNA was assayed by EMSA [18,19], slightly modified [20]. Preparative EMSA was carried out as described [21]. Briefly, the isolated protein was incubated with an oligonucleotide labeled at its 5' ends. The retarded band was cut off, the proteins eluted and separated by glycerol-SDS-PAGE. The effect

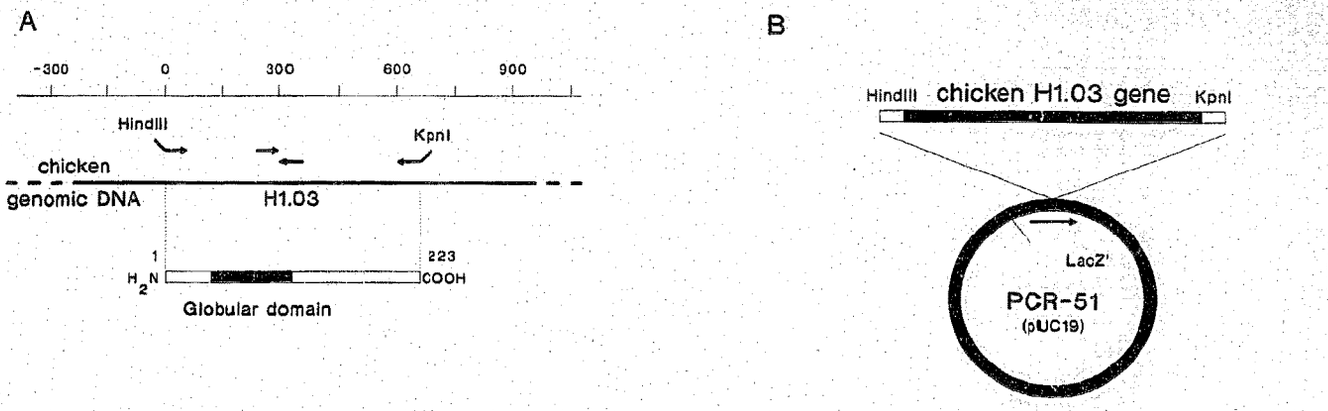

Fig. 1. The strategy used for amplification of the coding region of the histone H1.03 gene (Panel A) and cloning in the frame in conjunction with the *lacZ'* gene (Panel B). The locations of both the sense and antisense strand PCR primers that contained overhangs for the *Hin*dIII and *Kpn*I restriction sites and the internal sequencing primers are depicted in Panel A.

of H1 antibodies on binding of the fusion protein to DNA was demonstrated by EMSA, as described [10].

2.3. *Additional methods*

An H1 fraction capable of binding to the CTF/NF-I recognition sequence was isolated from rat liver nuclear extracts by heparin-agarose and two consecutive DNA affinity chromatographies [10].

Oligonucleotides were synthesized using a Beckman DNA Synthesizer. The primers in the PCR or dideoxy sequencing were used without further purification, while the oligonucleotides used for the synthesis of the double-stranded oligonucleotide containing the CTF/NF-I recognition sequence were annealed and purified by PAGE. Oligonucleotides were radioactively labeled at their 5' ends using polynucleotide kinase.

The fusion protein material for protein sequencing was separated by glycerol-SDS-PAGE and the proteins stained with Coomassie brilliant blue for a short period of time. The proteins were electroeluted and subjected to sequencing by Edman degradation using an Applied Biosystems 477A Protein Sequencer and 120A Analyzer. Glycerol-SDS-PAGE was performed as described previously [22]. The gels were subjected to staining with Coomassie brilliant blue or silver [23].

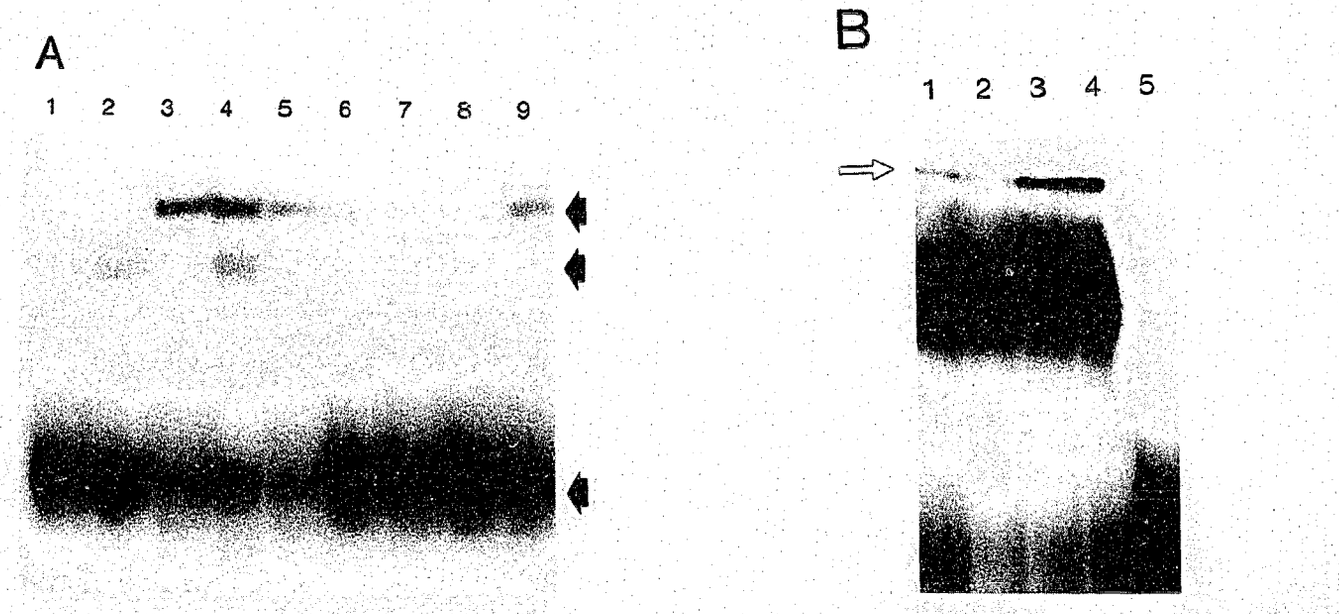

Fig. 2. EMSA using the H1-*lacZ'* fusion protein and the consensus oligonucleotide for the CTF/NF-I binding sequences. Panel A demonstrates the binding. Lane 1 contains only the labeled double-stranded oligonucleotide, and lanes 2 and 4 100 ng of H1 isolated from rat liver using DNA recognition site affinity chromatography, 100 ng of the H1.03-*lacZ'* fusion protein produced in *E. coli* and both of them, respectively, in addition to the oligonucleotide. Lanes 5 to 9 contain the fusion protein and lanes 6 to 8 a 10-, 50- and 100-fold excess of the oligonucleotide unlabeled, respectively. All the lanes contain 2.5 μg of double-stranded poly(dI-dC). The retarded complexes and the free labeled fragment are depicted by arrows. Panel B demonstrates the effect of H1 antibodies on the mobility of the fusion protein–DNA complex in EMSA. Lanes 1 and 2 contain H1-*lacZ'* incubated in the presence of an increasing concentration non-immune serum and lanes 3 and 4 that incubated in the presence of the H1 antiserum. Lane 5 is the labeled DNA fragment alone. The slow mobility band observable in the presence of antibodies to H1 is denoted by an arrow.

3. RESULTS AND DISCUSSION

The coding region of the chicken H1.03 gene [24] was cloned to the polylinker area of plasmid pUC19 between the *Hin*dIII and *Kpn*I sites (Fig. 1) using PCR. The primers were designed so that the genomic fragment obtained after digestion with the appropriate enzymes could be cloned in the correct orientation in conjunction with the *lacZ'* gene to generate an open reading frame of 873 nucleotides coding for an H1.03-*lacZ'* fusion protein of 291 amino acids. A number of positive clones was obtained by restriction site analysis, and one of these, PCR-51, was subjected to sequencing by the dideoxy method. Both strands were sequenced, and the construct was shown to contain the intact H1.03 coding region in conjunction with the *lacZ'* gene.

The protein coded by the fusion gene was produced in *E. coli*, and purified using chromatography on an anti-*lacZ'* antibody column. An apparent molecular mass 36 kDa protein was obtained, its mobility being in agreement with that of H1.03, which migrates in the 32 kDa position by globular standards despite consisting of only 223 amino acids. An additional molecular mass 14 kDa protein was observed, but this turned out to be lysozyme used to lyse the bacteria.

The 36 kDa band was recognized in Western blots by antibodies to H1 and β-galactosidase, and final proof of the identity of the proteins was obtained by protein sequencing, the proteins being separated by glycerol-SDS-PAGE, eluted from the gel and subjected to determination of the N-terminal sequence by Edman degradation. The molecular mass 14 kDa protein gave the lysozyme sequence and the other that of the fusion protein.

The H1-*lacZ'* fusion protein binds to the 5'-TTGGCAnnnTGCCAA-3' motif on DNA, as demonstrated by EMSA (Fig. 2A). A signal of slower mobility than in the case of H1 isolated from rat liver was obtained, the mobilities of the protein-DNA complexes being in good agreement with the relative molecular masses of the two proteins. The binding is specific, since it was counteracted by the addition of an excess of the unlabeled binding motif, and clear competition was obtained with increasing amounts of the competing oligonucleotide. On the other hand, actually no binding was observed to a synthetic oligonucleotide containing only one half-site of the symmetric CTF/NF-I binding motif (Fig. 3).

Two additional facts suggest that the relative mobility 36 kDa H1-*lacZ'* fusion protein causes the retardation. Firstly, only that protein could be detected by silver staining in a glycerol-SDS-polyacrylamide gel after electroelution and subsequent electrophoresis of proteins from the retarded band (data not shown), and secondly, incubation in the presence of H1 antiserum resulted in the appearance of a low mobility band in EMSA, as in the case of H1 isolated from rat liver (Fig. 2B).

A deletion mutant of H1.03 lacking amino acids 147 to 223 was produced in *E. coli* using the same strategy as for the production of the wild-type protein. The mutant protein binds in EMSA to the oligonucleotide containing the CTF/NF-I recognition motif with a two- to three-fold lower affinity compared with the wild-type protein (Fig. 4). The binding is specific, however, since

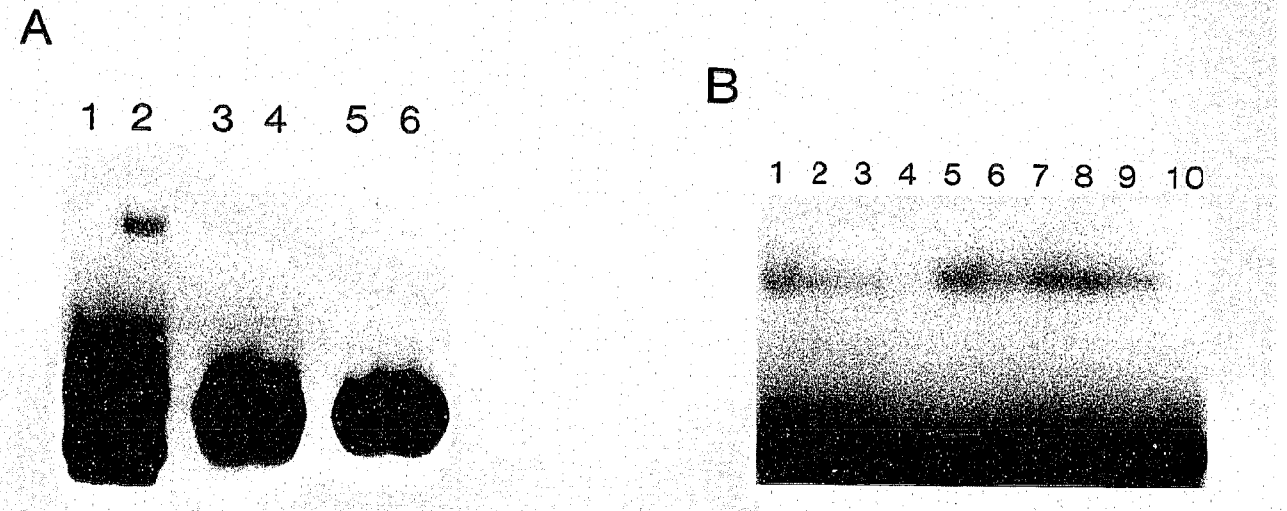

Fig. 3. Specificity of binding of the H1.03-*lacZ'* fusion protein. Panel A demonstrates binding of H1-*lacZ'* to synthetic labeled oligonucleotides containing the CTF/NF-I consensus sequence [12] (lane 2), a putative site from the mouse α_2(I) collagen promoter with one intact half-site [20] (lane 4) and an AP1 site from the rat transin gene [32] (lane 6). Lanes 1, 3 and 5 do not contain protein. Panel B demonstrates competition for binding of H1-*lacZ'* to the CTF/NF-I consensus motif from synthetic unlabeled oligonucleotides. Lanes 2 to 4 contain 1, 10 and 100 ng of the α_2(I) collagen oligonucleotide unlabeled, and lanes 6 to 8 that of the transin oligonucleotide, respectively. Lanes 1, 5 and 9 do not contain any competitor. Lane 10 does not contain protein.

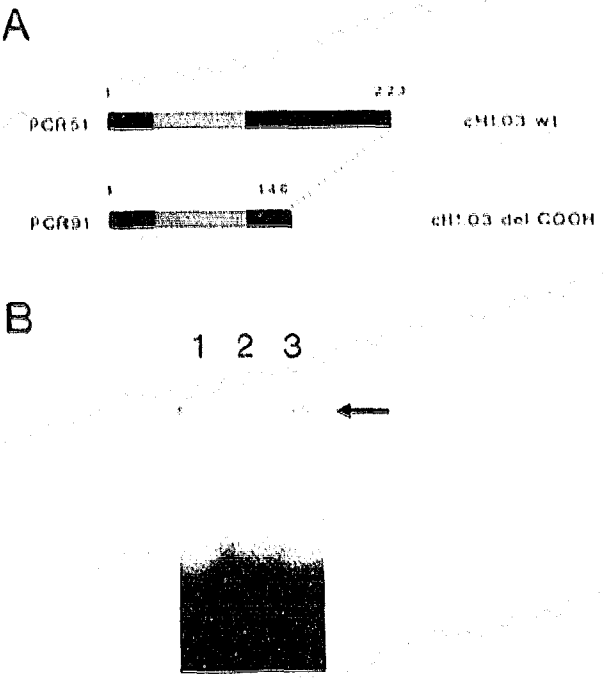

Fig. 4. Panel A depicts the strategy used for the production of an H1.03 deletion mutant that is lacking the C-terminal tail. Panel B demonstrates that the mutant protein binds to DNA in a sequence-specific manner. Binding of the fusion protein to the 5'-labeled oligonucleotide was assayed using EMSA. Lanes 1 and 3 contain 100 ng of the mutant protein, and lane 2 a 100-fold excess of the CTF/NF-I oligonucleotide unlabeled, as a competitor. The retarded complex is depicted by an arrow.

increasing amounts of the oligonucleotide unlabeled readily counteract it and no competition is observable with non-specific oligonucleotides (see Fig. 3).

The EMSA experiments demonstrate that the H1-*lacZ'* fusion protein is capable of sequence-specific binding to DNA, thus confirming the results obtained previously using H1 purified from rat liver by DNA affinity chromatography [10,20] and by PCA extraction [11]. The globular domain of H1 may be responsible for this interaction, as demonstrated using the deletion mutant, although it should be noted that the mutant protein contains approximately 30 extra amino acid residues both in the N- and C-terminal ends of the globular domain.

The major problem in demonstrating sequence-specific interaction of the H1-*lacZ'* fusion protein with DNA results from the high level of non-sequence-specific interaction. Consistent with this, many of the histones have recently been demonstrated to contain a novel DNA binding unit, an SPKK motif, capable of interaction with the exceptionally narrow DNA minor groove in A+T-rich regions [25,26]. The C-terminal tail of H1 contains several copies of this motif, but all these were removed in the deletion mutant. We suggested in our previous report that the N-terminal end of α3 of the H1 globular domain may contact with bases on the bottom of the DNA major groove in a similar manner to a number of prokaryotic regulatory proteins and homeodomain proteins [9]. The present results and several lines of previous experimental evidence lend thus support to this hypothesis [27,28].

The finding that H1 recognizes a specific sequence on DNA raises interesting questions. Firstly, is the 5'-TTGGCAnnnTGCCAA-3' motif or a part of it a determinant of nucleosome positioning? In general, nucleosomes exist in certain positions along the chromatin, being unevenly spaced [29,30]. In the absence of H1 nucleosomes reconstituted in vitro are regularly spaced and capable of sliding along the DNA [31], while addition of H1 fixes them to certain sites with uneven spacing [30], this possibly resulting from binding of H1 to its specific recognition sequence (see [11]).

Secondly, the fact that the sequence recognized by H1 is also a recognition sequence for CTF/NF-I implies that the two factors may compete for the same sites on DNA, but it is not presently known whether the consensus sequences recognized by these two factors really are identical or only similar (see [12]).

Acknowledgements: This work was supported in part by grants from the Medical Research Council of the Academy of Finland and the Finnish Cancer Research Foundation. The expert technical assistance of Ritva Savilaakso is acknowledged.

REFERENCES

[1] Weintraub, H. (1985) Cell 42, 705–711.
[2] Wolffe, A.P. (1989) EMBO J. 8, 527–537.
[3] Sun, J.-M., Wiaderkiewicz, R. and Ruiz-Carrillo, A. (1989) Science 245, 68–71.
[4] Zlatanova, J. (1990) Trends Biochem. Sci. 15, 273–276.
[5] Staynov, D.Z. and Crane-Robinson, C. (1988) EMBO J. 7, 3685–3691.
[6] Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) Nature 311, 532–537.
[7] Morse, R.H. and Simpson, R.T. (1988) Cell 54, 285–287.
[8] Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K. and Zarbock, J. (1987) EMBO J. 6, 1833–1842.
[9] Mannermaa, R.-M. and Oikarinen, J. (1989) Biochem. Biophys. Res. Commun. 162, 427–434.
[10] Ristiniemi, J. and Oikarinen, J. (1989) J. Biol. Chem. 264, 2164–2174.
[11] Sevall, J.S. (1988) Biochemistry 27, 5038–5044.
[12] Nilsson, P., Hallberg, B., Thornell, A. and Grundström, T. (1989) Nucleic Acids Res. 17, 4061–4075.
[13] Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1985) Nature 334, 218–224.
[14] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, USA.
[15] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487–491.
[16] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[17] Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) Cell 51, 1079–1090.

[18] Garner, M.M. and Revzin, A. (1981) Nucleic Acids Res. 9, 3047–3060.
[19] Fried, M. and Crothers, D.M. (1981) Nucleic Acids Res. 9, 6505–6525.
[20] Oikarinen, J., Hatamochi, A. and De Crombrugghe, B. (1987) J. Biol. Chem. 262, 11064–11070.
[21] Rupp, R.A.W. and Sippel, A.E. (1987) Nucleic Acids Res. 15, 9707–9726.
[22] Schägger, H. and Von Jagow, G. (1987) Anal. Biochem. 166, 368–379.
[23] Morrissey, J.H. (1981) Anal. Biochem. 117, 307–310.
[24] Coles, L.S., Robins, A.J., Madley, L.K. and Wells, J.R.E. (1987) J. Biol. Chem. 262, 9656–9663.
[25] Suzuki, M. (1989) EMBO J. 8, 797–804.
[26] Izaurralde, E., Kas, E. and Laemmli, U.K. (1989) J. Mol. Biol. 210, 573–585.
[27] Crane-Robinson, C. and Privsyn, O.B. (1989) Protein Eng. 2, 577–582.
[28] Turnell, W.G., Satchwell, S.C. and Travers, A.A. (1988) FEBS Lett. 232, 263–268.
[29] Richard-Foy, H. and Hager, G.L. (1987) EMBO J. 6, 2321–2328.
[30] Rodriguez-Campos, A., Shimamura, A. and Worcel, A. (1989) J. Mol. Biol. 209, 135–150.
[31] Shimamura, A. and Worcel, A. (1989) J. Biol. Chem. 264, 14524–14530.
[32] Kerr, L.D., Miller, D.B. and Matrisian, L.M. (1990) Cell 61, 267–278.