Volume 323, number 3, 252–256 FEBS 12523 June 1993

Folding topology and DNA binding of the N-terminal fragment of Ada protein

Hitoshi Sakashita[a], Takahiko Sakuma[b], Tadayasu Ohkubo[a], Masatsune Kainosho[b], Kunihiko Sakumi[c], Mutsuo Sekiguchi[c] and Kosuke Morikawa[a]

[a]*Protein Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565, Japan,* [b]*Tokyo Metropolitan University, 1-1, Minami-ohsawa, Hachioji-shi, Tokyo, 192-03, Japan* and [c]*Medical Institute of Bio-Regulation, Kyushu University, Fukuoka 812, Japan*

Received 23 March 1993; revised version received 8 April 1993

Three amino terminal fragments of *Escherichia coli* Ada protein (39 kDa) with different molecular masses (14 kDa, 16 kDa and 20 kDa) were prepared in large quantities from an *E. coli* strain harboring plasmids constructed for the overproduction of the truncated proteins. The three fragments can be methylated to an extent similar to that of the intact molecule. The methylated 16 kDa fragment specifically binds to the ada box on a DNA duplex. NMR analyses revealed that the 14 kDa fragment comprises two α-helices and a β-sheet with parallel and anti-parallel mixed strands. A comparison of the ^{15}N-^{1}H HMQC spectra of the fragments has led to the conclusion that this tertiary structure within the 14 kDa fragment is retained in the larger 16 kDa and 20 kDa fragments.

Ada protein; DNA binding activity; Gel mobility shift assay; Nuclear magnetic resonance; Secondary structure

1. INTRODUCTION

Exposure of *Escherichia coli* cells to a low concentration of alkylating agents enhances their resistance to both mutagenic and cytotoxic effects. This adaptive response induces the production of various repair enzymes which remove alkylated nucleotides from damaged DNA duplexes [1,2]. The Ada protein plays a central role in this adaptive response with two major functions. First, this protein transfers the methyl groups from methyl-phosphotriesters and O^6-methylguanines in alkylated DNA duplexes to the residue Cys^{69} and Cys^{321} of its own molecule, respectively [3,4]. Secondly, this protein can act as a transcriptional regulator not only for its own *ada* gene but also for other genes belonging to the *ada* regulon, after the acceptance of a methyl group to the amino terminal half of the protein. In fact, the methylation at Cys^{69} converts the Ada protein from a weak to a strong activator of transcription [5,6]. In *E. coli* cells, the intact 39 kDa molecule is cleaved by an unknown cellular protease to produce 20 kDa (N-ada 20k) and 19 kDa fragments. The methylated 20 kDa N-terminal fragment can bind to the ada box and act as a repressor, whereas the 19 kDa C-terminal fragment cannot [7]. Recently, Myers et al. [8] have reported that the N-ada 20k fragment possesses a site that strongly binds a single zinc ion. This metal binding site is located within the first 92 amino acid residues from the N-terminus (N-ada 10k), and contains the motif $Cys\text{-}X_3\text{-}Cys\text{-}X_{26}\text{-}Cys\text{-}X_2\text{-}Cys$.

We have found that the N-ada 20k fragment can be further truncated to either 16 kDa or 14 kDa fragments, by releasing C-terminal portions via mild proteolysis. These two shorter fragments have been found to almost completely retain the DNA methyltransferase activity. In order to investigate the functional and structural properties of these fragments, we have constructed a protein overproduction system in *E. coli*, which allows large scale preparation of the N-terminal 20 kDa, 16 kDa and 14 kDa fragments. In this report, we describe the functional properties of the three fragments, such as the methyltransferase activity and the specific recognition with the cognate DNA sequences, and the secondary structures and their folding topology of the 14 kDa fragment, determined by NMR analyses.

Correspondence address. H. Sakashita and T. Ohkubo, Protein Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565, Japan. Fax: (81) (6) 872 8210.

Abbreviations: N-ada 14k, Ada fragment containing residues 1–129; N-ada 16k, Ada fragment containing residues 1–146; N-ada 20k, Ada fragment containing residues 1–178; HMQC, heteronuclear multiquantum coherence; PCR, polymerase chain reaction; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; NOE, nuclear Overhauser effect.

2. MATERIALS AND METHODS

2.1. *Construction of the Ada fragment overproducing strains*

The respective N-terminal fragments of the Ada protein were overproduced by constructing fragments of the *ada* gene using a PCR-based approach. A 5′-primer that encodes the *SmaI* restriction site and the initiation codon was combined with a 3′-primer that encodes a stop codon and the *PstI* restriction site following either Lys^{129} (14 kDa), Lys^{146} (16 kDa) or Lys^{178} (20 kDa). Each PCR product from pYN3059 [4] was ligated into pEXP7 [9] and transformed into *E. coli* JM109.

Published by Elsevier Science Publishers B V

2.2. *Sample preparation*

The respective N-terminal fragments of the Ada protein were expressed by *E. coli* harboring each high expression plasmid. These fragments were purified from the harvested *E. coli* cells by phosphocellulose chromatography. The SDS-PAGE analysis indicated single bands for all three fragments and suggested purities above 95%. Using this expression system, 10–20 mg of each protein fragment could be obtained from one liter of the culture. It should be noted that the sequence of the *E. coli* K-12 Ada polypeptides used in the present study is different at Leu[7], Thr[151] and Phe[159] from the sequence of the *E. coli* B Ada protein [4,10]. The uniformly ^{13}C- and ^{15}N-labeled protein fragments were prepared from *E. coli* cells grown in M9 minimal medium with [^{13}C]glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources. Specific residues of the protein were labeled with [^{15}N]alanine, [^{15}N]arginine, [^{15}N]lysine, [^{13}C]threonine or [^{13}C]leucine. Specimens for NMR measurements were dialyzed and dissolved in either D_2O or 90% H_2O/10% D_2O containing acetate buffer, pH 5.5, and protein concentrations were set at 1–2 mM.

2.3. *Gel mobility shift assays*

The deoxyoligonucleotides were synthesized with an Applied Biosystems DNA synthesizer Model 380B. For gel mobility shift assays, a DNA sequence containing the ada box: 5′-GCGAAAAAAATTAAAGCGCAAGATTGTTGGTTTTTGC-3′ was used, where the sequence of only one strand of the duplex is presented. DNA binding assays were carried out in a 10 μl solution containing 50 mM HEPES-KOH (pH 7.8), 50 mM KCl, 10 mM $MgCl_2$, 5% glycerol, 1 pmol of 5′-end ^{32}P-labeled double stranded oligonucleotide, and either 0 or 100 ng unlabeled calf thymus DNA (Sigma). Binding reactions were performed at 22°C for 60 min. Either methylated or non-methylated proteins (1 pmol) were used for the assay. The methylated fragments were prepared according to the procedure described by Nakabeppu and Sekiguchi [4]. The samples were loaded on a 5% polyacrylamide gel. The electrophoresis was operated at 120 V for 1.5 h, and the labeled DNA was visualized by autoradiography.

2.4. *NMR spectroscopy*

2D and 3D NMR spectra were measured with a Bruker AMX-600 spectrometer. Proton detected ^{15}N-^1H 3D NOESY-HMQC and TOCSY-HMQC spectra were recorded with a NOE mixing time of 100 ms and a MLEV-17 mixing time of 30 ms, respectively. A series of triple resonance experiments: HNCO, HNCA, HN(CO)CA, HN(CA)HA and HCACO, were performed for a uniformly ^{13}C- and ^{15}N-labeled sample with a triple resonance probe at 600 MHz for protons [11,12]. For all NMR experiments, the temperature was set at 30°C. The data were processed with UXNMR on an X32 computer. ^1H chemical shifts were referenced to 2,2′-dimethyl-2-silapentane-5′-sulfonate (DSS) and were measured with respect to an internal standard of acetone. The ^{15}N chemical shifts were measured relative to an external standard of NH_4Cl and are referred to liquid NH_3. Slowly exchanging amide protons were identified from the ^{15}N-^1H HMQC spectra.

3. RESULTS

3.1. *DNA binding of truncated Ada fragments*

In order to gain insight into the functional and structural properties of the three truncated N-terminal 20 kDa, 16 kDa and 14 kDa fragments, we have constructed a high level *E. coli* expression system which is under the control of the *tac* promoter. These overproducing strains allowed us to prepare large amounts of fragments, which were sufficient for NMR studies. DNA binding activities for the methylated and non-methylated fragments were measured in the absence or presence of a non-specific duplex DNA (Fig. 1A and 1B). In the absence of the non-specific DNA, retardation of the ada box containing labeled oligonucleotide was observed, only in the presence of either the N-ada

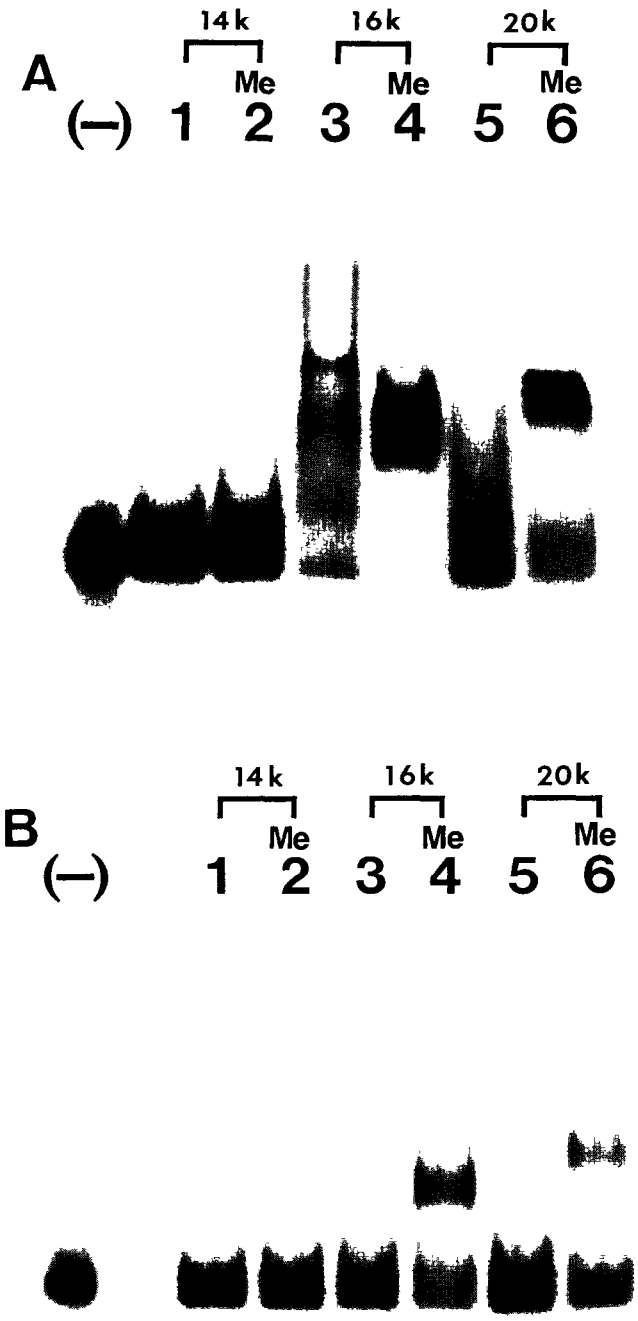

Fig. 1. Photographs of autoradiograms illustrating DNA binding by three truncated Ada fragments, as detected by gel mobility assays in the absence (A) and presence (B) of non-specific DNA. ^{32}P-labeled ada box containing duplex oligomer was mixed with three truncated fragments: lanes 1 and 2, N-ada 14k; lanes 3 and 4, N-ada 16k; lanes 5 and 6, N-ada 20k. Data for the methylated forms of the fragments (lanes 2, 4 and 6) and the non-methylated form fragments (lanes 1, 3 and 5) are shown. The lanes designated (–) indicate the minus-protein controls.

16k or the N-ada 20k fragments, regardless of whether or not these fragments were methylated (Fig. 1A, lanes 3–6). However, neither the methylated nor the non-methylated N-ada 14k fragment caused retardation of the oligonucleotide (Fig. 1A, lanes 1 and 2). These results indicate that both the N-ada 16k and the N-ada 20k fragments can non-specifically bind to the DNA oligomer duplex, whereas the N-ada 14k fragment has lost binding activity, regardless of its methylation state. In the presence of the non-specific DNA, the bands of the ada box containing oligomer duplex were shifted on the gel, only when the methylated N-ada 16k and N-ada 20k fragments were used (Fig. 1B, lanes 4 and 6). On the other hand, the oligonucleotide mixed with either the methylated N-ada 14k fragment or all the non-methylated fragments showed no significant shift on the gel (Fig. 1B). These results show that the methylated N-ada 16k fragment, and the N-ada 20k fragment, retain the ability of specific recognition with the ada box containing DNA, although the N-ada 14k fragment lacks both specific and non-specific DNA binding activities.

3.2. *NMR measurements*

The sequence-specific resonance assignment of the N-ada 14k fragment spectrum was made by surveying through-bond and through-space connectivities, according to the procedure of sequential resonance assignments introduced by Wuthrich et al. [13]. This procedure relies upon various through-bond correlation experiments, such as 3D ^{15}N-^{1}H TOCSY-HMQC spectra, to identify amino acid spin systems. Subsequently, through-space correlation experiments, such as 3D ^{15}N-^{1}H NOESY-HMQC spectra, are required to establish sequence-specific assignments using the amino acid sequence data. Due to the many overlaps of the CαH and NH proton signals, it was difficult to identify sequential NOE connectivities, and hence the conventional sequential resonance assignments were complex. To solve this problem, the N-ada 14k fragment was uniformly labeled with ^{13}C and ^{15}N, and triple resonance experiments were used in parallel. The resolution was improved by the use of the ^{13}C- and ^{15}N-labeled protein. However, signal overlaps that obstructed assignments still remained in the 3D spectra. Nevertheless, sequential assignments were successful for the backbone signals of 98 out of 129 residues, and this information allowed the identification of secondary structures in the main parts. In the present analysis, sequential assignments were unable to be done for the 30 residues between His^{83} and Ser^{112}. These residues have little chemical shift dispersion, and several residues indicate weak NH proton signals because of signal broadening. Therefore, the secondary structure in this region remains to be determined.

Secondary structure elements were deduced from the pattern of sequential and medium-range NOE connectivities. The N-ada 14k fragment consists of four β-strands and two α-helices, as shown in Fig. 2. The helical structure is characterized by sequential strong dNN connectivities and a number of medium-range dαN(i,i + 3) connectivities. Based on these data, two α-helices were identified between residue 9 and 15 and between residue 58 and 65. A β-sheet is characterized by strong dαN connectivities and interstrand connectivities

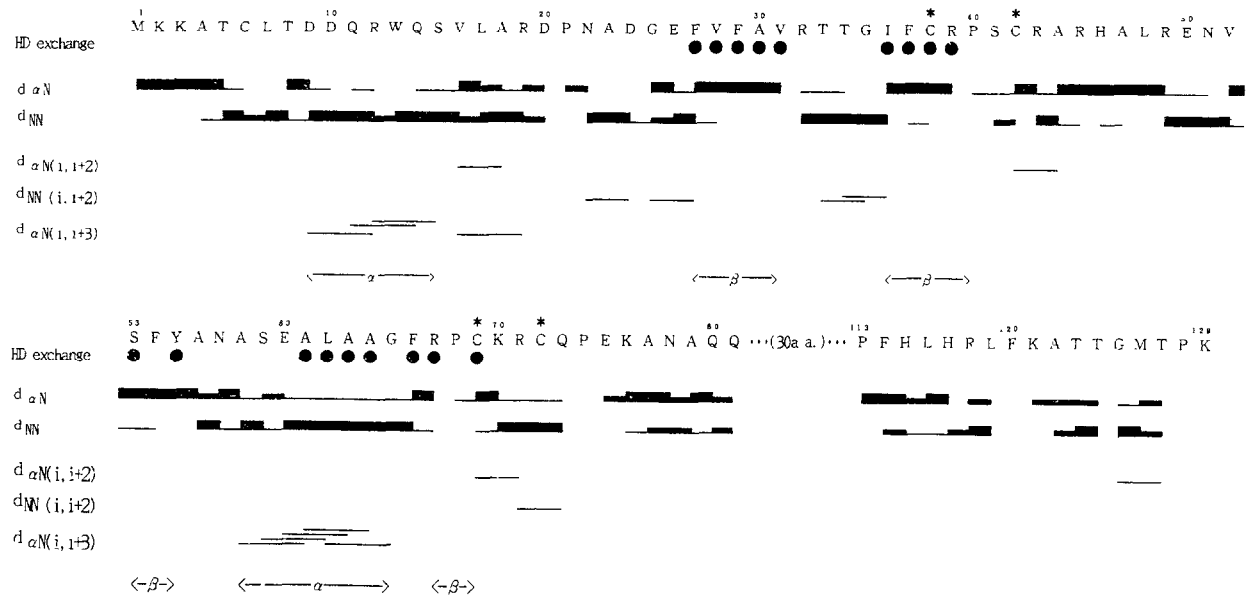

Fig. 2. Summary of the short and medium-range NOE's observed in the 3D ^{15}N-^{1}H NOESY-HMQC spectrum of the N-ada 14k fragment. The thickness of the lines indicates the relative intensity of the NOE cross-peaks. Backbone amide protons with slow H-D exchange rates are indicated by solid circles. Secondary structure assignments deduced from these data are indicated at the bottom of the figure. Asterisks indicate the four cysteines that form the proposed Zn^{2+} binding site.

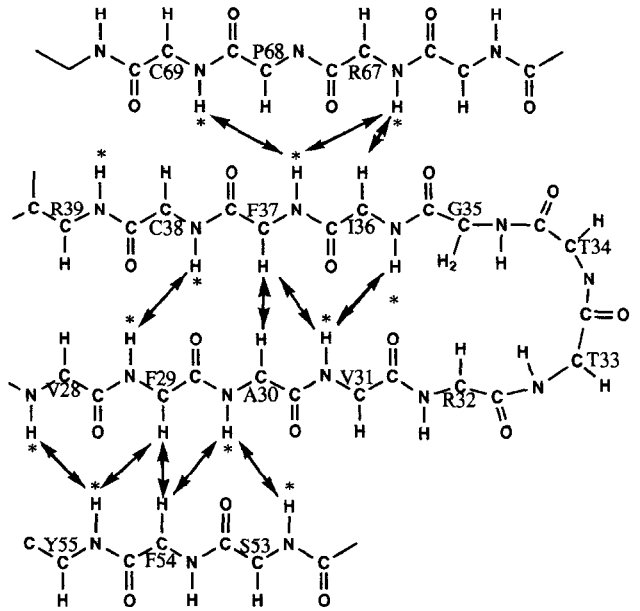

Fig. 3. Schematic diagram of the arrangement of the four-stranded β-sheet of the N-ada 14k fragment. Arrows indicate the observed interstrand NOE's. Amide protons with slow H-D exchange rates are indicated by asterisks.

among the CαH and the NH protons across the sheet. Successive strong dαN connectivities were observed for the peptide segments from Phe[27] to Val[31], from Ile[36] to Arg[39] and from Val[52] to Tyr[55]. Several interstrand NOEs found between the CαH protons, indicate that three segments of residues 27–31, 36–39 and 53–55 form an antiparallel β-sheet. Furthermore, some interstrand NOEs observed between the CαH or the NH protons in residues 36–39 and those in residues 67–69 show that nearly parallel β strands cross each other at this region. This arrangement of the β-strands was also confirmed by the locations of slowly exchanging NH protons. Thus, the analysis of these NOE data has led to the conclusion that the parallel and antiparallel mixed β-strands form a single β-sheet, as shown in Fig. 3.

4. DISCUSSION

Secondary structure prediction [14] of the Ada protein suggested that the 10 kDa N-terminal segment contains both α-helices and β-structures, while the segment between Leu[93] and Lys[146], corresponding to the 6 kDa segment at the C-terminus within the N-ada 16k fragment, comprises only α-helices and loops, but no β-structure. Notably, the 17 residues between Ala[130] and Lys[146], which constitute the 2 kDa extra carboxyl terminal segment of the N-ada 16k fragment, are abundant in basic residues, suggesting their involvement in DNA binding. Sequence analysis [15] suggested the presence of a helix-turn-helix motif in the N-ada 14k fragment. The secondary structure actually deduced from the present NMR analysis is not totally consistent with these predictions. The assignments of residues 83–112 were not successful because of signal broadening, and therefore the secondary structure remains unknown. At present, we cannot conclude whether this broadening is due to conformational flexibility or multi-conformations.

As described in the previous section, the results from the gel mobility shift assays revealed that the N-ada 20k fragment retains the same degree of specific DNA binding activity as found in the intact 39 kDa Ada molecule. Furthermore, the same assays for the other two fragments showed that the N-ada 16k fragment completely retains the binding activity with the DNA oligomer containing the ada box, whereas the N-ada 14k fragment loses this activity. These results imply that the basic segment, corresponding to the extra C-terminal 2 kDa region between residues 130 and 146 plays an essential role for binding to DNA either in specific or in non-specific manner to the ada box.

Although the N-ada 14k fragment lacks DNA binding activity, the ^{15}N-^{1}H HMQC spectra flow that most of the chemical shifts (in both the ^{15}N and amide protons) are the same as those of the N-ada 16k and N-ada 20k fragments. Therefore, we conclude that the same tertiary structure of the N-ada 14k fragment is retained in the larger N-ada 16k and N-ada 20k fragments, and that it is independent of the extra carboxyl terminal segments. The N-ada 14k fragment is presumed to contain both the Zn^{2+} binding site and the helix-turn-helix motif. The Zn^{2+} binding domain, which is thought to carry the DNA methyltransferase activity, mainly consists of β-structure with two short α-helices (Fig. 2). The peptide segment from Pro[113] to Lys[121] was predicted to be located in the second DNA recognition helix of the helix-turn-helix motif. The present NMR data indicate that, in contrast to this prediction, this region adopts a turn-like structure (Fig. 2). It might be possible that the second helix of the motif is disrupted due to the structural instability at the artificial terminus, that is introduced to produce the N-ada 14k fragment.

From a sequence comparison of the Ada proteins from *Salmonella typhimurium* [16] and *Bacillus subtilis* [17], Meyers et al. have proposed that four cysteines, Cys[38], Cys[42], Cys[69] and Cys[72], would form the Zn^{2+} binding site [8]. It is interesting to consider the locations with the secondary structure of the cysteine residues that are important for Zn^{2+} binding. Fig. 3 shows that Cys[69] and Cys[38] are located face to face with each other on the parallel β-sheet. Since Cys[72] and Cys[42] are only 3 and 4 residues away from Cys[69] and Cys[38], respectively, it is likely that these four cysteine residues are spatially close to each other and constitute a single Zn^{2+} binding site. Preliminary ^{113}Cd NMR measurement exhibits a single resonance at 627 ppm. Its chemical shift suggests that the cadmium ion is tetrahedrally coordinated [18]. The ^{1}H-^{113}Cd HMQC spectra also show that

all the residues coordinated to the ^{113}Cd atom are cysteines, and not histidines (data not shown). These results are consistent with the Zn^{2+} binding scheme proposed by Meyers et al. [8]. It is noteworthy that Cys^{69} can accept a methyl group from the methyl-phosphotriester in an alkylated DNA.

In summary, the present DNA binding experiments indicate that the 2 kDa basic segment between Ala^{130} and Lys^{146} is essential for both specific and non-specific DNA binding activities. The NMR measurements show that the Zn^{2+} binding domain mainly consists of four β-strands. In the primary structure of the Ada protein, the hypothetical helix-turn-helix motif and the 2 kDa basic segment, which are presumed to participate in DNA binding, are located outside the N-ada 10k segment (residues 1–92) containing the Zn^{2+} binding site. Therefore, they are considerably distant in the sequence from the Zn^{2+} binding domain, which includes the methylation site at Cys^{69}. Nevertheless, this methylation is the obvious trigger of the transcriptional enhancement in the *ada* regulon. Therefore, the determination of the tertiary structure of the N-ada 16k fragment should greatly contribute to the elucidation of this methylation-dependent switching mechanism of the Ada protein.

Acknowledgements. We are grateful to M. Ariyoshi for her assistance with the proteolysis experiments. This study has been partly supported by a grant from the Human Frontier Science Program (HFSP).

REFERENCES

[1] Lindahl, T., Sedgwick, B., Sekiguchi, M. and Nakabbepu, Y. (1988) Annu. Rev. Biochem. 57, 133–157.
[2] Volkert, M.R. (1988) Environ. Mol. Mutagen. 11, 241–255.
[3] Teo, I., Sedgwick, B., Demple, B., Li, B. and Lindahl, T. (1984) EMBO J. 3, 2151–2157.
[4] Nakabeppu, Y., Kondo, H., Kawabata, S., Iwanaga, S. and Sekiguchi, M. (1984) J. Biol. Chem. 259, 13723–13729.
[5] Nakabeppu, Y. and Sekiguchi, M. (1986) Proc. Natl. Acad. Sci. USA 83, 6297–6301.
[6] Teo, I., Sedgwick, B., Kilpatrick, M.W., McCarthy, T.V. and Lindahl, T. (1986) Cell 45, 315 324.
[7] Akimaru, H., Sakumi, K., Yoshiaki, T., Anai, M. and Sekiguchi, M. (1990) J. Mol. Biol. 216, 261–273.
[8] Myers, L.C., Terranova, M.P., Nash, H.M., Markus, M A. and Verdine, G.L. (1992) Biochemistry 31, 4541–4547.
[9] Ono, M., Matsuzaka, H. and Ohta, T. (1990) J. Biochem. 107, 21–26.
[10] Demple, B., Sedgwick, B., Robins, P., Totty, N., Waterfield, M.D. and Lindahl, T (1985) Proc. Natl. Acad. Sci. USA 82, 2688–2692.
[11] Clore, G.M. and Gronenborn, A.M. (1991) Progress in NMR Spectroscopy 23, 43–92.
[12] Ikura, M., Kay, L.E. and Bax, A. (1990) Biochemistry 29, 4659–4667.
[13] Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York.
[14] Nishikawa, K. and Noguchi, T (1991) Methods Enzymol. 202, 31–44.
[15] Dodd, I.B. and Egan, J.B. (1990) Nucleic Acids Res. 18, 5019–5026.
[16] Hakura, A., Morimoto, K., Sofuni, T. and Nohmi, T. (1991) J. Bacteriol. 173, 3663–3672.
[17] Morohoshi, F., Hayashi, K. and Munakata, N. (1990) Nucleic Acids Res. 18, 5473–5780.
[18] Michael, F. and Summers (1988) Coord. Chem. Revs. 86, 43–134.