Identification of Tyrosine 204 as the Photo-Cross-Linking Site in the DNA−*Eco*RI DNA Methyltransferase Complex by Electrospray Ionization Mass Spectrometry†

David L. Wong‡ and Norbert O. Reich*

Department of Chemistry and Program in Biochemistry and Molecular Biology, University of California, Santa Barbara, California 93106

Received May 23, 2000; Revised Manuscript Received October 17, 2000

ABSTRACT: We describe a highly sensitive strategy combining laser-induced photo-cross-linking and HPLC-based electrospray ionization mass spectrometry to identify amino acid residues involved in protein−DNA recognition. The photoactivatible cross-linking thymine isostere, 5-iodoracil, was incorporated at a single site within the sequence recognized by *Eco*RI DNA methyltransferase (GAATTC). UV irradiation of the DNA−protein complex at 313 nm results in a >60% cross-linking yield. SDS−polyacrylamide gel electrophoresis and mass spectrometry were used to analyze the covalent cross-linked complex. The total mass is consistent with covalent bond formation between one strand of DNA and the protein with 1:1 stoichiometry. Protease digestion of the cross-linked complex yields several peptide−DNA adducts that were purified by anion-exchange column chromatography. A combination of mass spectrometric analysis and amino acid sequencing revealed that tyrosine 204 was cross-linked to the DNA. Electrospray mass spectrometric analysis of the peptide−nucleoside adduct confirmed this assignment. Tyrosine 204 resides in a peptide motif previously thought to be involved in AdoMet binding and methyl transfer. Thus, amino acids within loop segments but outside of "DNA binding" motifs can be critical to DNA recognition. Our method provides an accurate characterization of picomole quantities of DNA−protein complexes.

Sequence-specific DNA methylation involves the enzyme-catalyzed transfer of a methyl group from *S*-adenosyl-L-methionine (AdoMet) to the DNA substrate. Methylation occurs at cytosine C^5 or N^4 and adenine N^6 positions. Prokaryotic DNA methylation is involved in numerous processes, including gene regulation, DNA replication, and mismatch repair, and in restriction/modification systems that provide protection against infectious bacteriophage. Host DNA modification by methyltransferases protects the DNA against endonuclease cleavage, while foreign DNA from a species or strain with a different methylation pattern is destroyed by these restriction enzymes. The well-studied *Eco*RI enzymes fall into the type II restriction/modification class. The monomeric *Eco*RI methyltransferase (M.*Eco*RI, molecular mass of 37 913 Da) transfers a methyl group from AdoMet to the second adenine of its recognition site GAATTC to form N^6-methyladenine. The dimeric endonuclease (R.*Eco*RI, molecular mass of 31 059 Da) cleaves the same target site G/AATTC (*1*). The X-ray crystal structures of a DNA dodecamer containing an *Eco*RI site and the DNA−R.*Eco*RI complex have been determined (*2*−*4*). However, no crystal structure has been determined for M.*Eco*RI. Crystal structures of enzymes involved in cytosine C^5 and N^4 and adenine N^6 methylation are available (*5*−*8*), although only the cytosine C^5 enzymes have been co-crystallized with DNA.

M.*Eco*RI is a bilobal enzyme with an N-terminal catalytic domain (∼26 kDa) containing the active site and AdoMet-binding regions and a smaller C-terminal domain (∼12 kDa) with the target recognition region thought to be responsible for sequence-specific DNA binding (*9*). The preferred kinetic mechanism involves initial binding of nontarget DNA followed by AdoMet binding. AdoMet or AdoMet analogues facilitate target DNA site binding (*10*). M.*Eco*RI, as well as other adenine DNA methyltransferases, bends its target site upon binding (*11*). The DNA bending step facilitates the expulsion of the target adenine nucleoside into a pocket within the enzyme ($k_{flipping} = 20\ s^{-1}$) (*12*). Methylation ($k_{methylation} \sim 200\ s^{-1}$) occurs rapidly after nucleoside flipping (*12, 13*), and is followed by slow product release ($k_{cat} = 0.13\ s^{-1}$) (*14*). However, it is not clear how M.*Eco*RI or other adenine DNA MTases recognize their cognate sequences.

Ultraviolet (UV) light-induced photochemical cross-linking of protein to nucleic acids has been demonstrated using analogues such as 4-thiouracil, 8-azidoadenine, 5-bromouracil, and 5-iodouracil (*15*−*18*). Residues, especially those with aromatic side chains, in proximity to the DNA can be identified through structural characterization of DNA−protein adducts resulting from photo-cross-linking and amino acid sequencing (*19*−*21*). However, most of these analytical methods require large quantities of protein and DNA due to the low cross-linking efficiencies and the application of standard peptide sequencing methods. The combination of a highly sensitive mass spectrometry analytical procedure with the photo-cross-linking method is an extremely powerful approach for the identification of amino acids involved in protein−DNA recognition. Due to the conflicting require-

† This work was supported by NSF Grant MCB-9603567 to N.O.R.

* To whom correspondence should be addressed. Telephone: (805) 893-8368. Fax: (805) 893-4120. E-mail: reich@chem.ucsb.edu.

‡ Current e-mail address: dlwong@itsa.ucsf.edu.

10.1021/bi001164v CCC: $19.00
Published on Web 11/21/2000

ments for the ionization of different macromolecules, for example, proteins and peptides that form positive ions or nucleic acids that form negative ions, heteroconjugated species such as peptide−oligonucleotide adducts pose unique difficulties for mass spectrometric analysis. Our previous studies using 5-iododeoxyuracil (5-IdU) for the replacement of thymine resulted in significant photo-cross-linking yields (>60%) and low levels of photoinduced protein degradation (*22*). These ESI-MS studies of protein−oligonucleotide complexes demonstrated that we could obtain relatively high sensitivity with limited amounts of sample under the appropriate mass detection mode (*22*). Other mass spectrometric analyses of peptide−oligonucleotide adducts using related methods have been reported (*23*, *24*).

Since little is known about how DNA adenine methyltransferases recognize their cognate sites, we sought to develop a novel approach for identifying amino acids critical to DNA recognition. In this paper, we extended our previous studies using the photoactivatible thymine isostere 5-IdU and describe a general method with picomole sensitivity. Moreover, our results suggest that this approach may also be useful with cell extracts.

EXPERIMENTAL PROCEDURES

Materials. Phosphoramidites and DNA synthesis reagents were obtained from Milligen/Biosearch. 5-IdU-CE and N^6-Me-dA-CE were purchased from Glen Research. [γ-^{32}P]ATP (6000 Ci/mmol) was from Amersham. Sinefungin and 2-mercaptoethanol were purchased from Sigma Chemicals.

Oligonucleotide Analogue Synthesis. DNA substrates (14mers) were prepared on a Biosearch 3810 DNA synthesizer using β-cyanoethyl phosphoramidites. The upper strand of each duplex substrate (CT, CTI-1, or CTI-2), d(GGCG-**GAATTC**GCGG), contains the M.*Eco*RI recognition site, and 5-IdU was incorporated into the first or second thymine position. The lower strands (CBM, BMI-1, and BMI-2), d(CCGC**GAMTTC**CGCC), all contain N^6-methylated adenine (M) and, when indicated, 5-IdU, at one of the two thymine positions. The DNA was deprotected in NH_4OH at room temperature overnight, the solvent removed, and the sample resuspended in 0.1 M TEAA (pH 7.0). The DNA was purified on a Dynamax C18 reversed-phase PureDNA column (Rainin Instrument Co.) (*25*). Purified oligonucleotides were diluted in 10 mM Tris (pH 8.0) and 1 mM EDTA, and the concentrations were determined by absorbance at 260 nm. Complementary single strands were annealed, and the double-stranded form was confirmed by autoradiography of ^{32}P-labeled DNA with nondenaturing polyacrylamide gel electrophoresis.

M.EcoRI Preparation and Enzymatic Analysis. M.*Eco*RI from *Escherichia coli* strain MM294 harboring the overexpression plasmid pXRI was purified as described previously (*26*). The purified enzyme was dialyzed extensively in 100 mM NaCl, 10 mM potassium phosphate (pH 7.0), 1 mM EDTA, 7 mM β-mercaptoethanol, 1 mM NaN_3, and 20% glycerol at 4 °C. Concentrations of purified M.*Eco*RI were determined spectrophotometrically using an extinction coefficient $E^{1\%}_{280}$ of 10.8 (*1*).

Binding Assay and Electrophoresis. The apparent thermodynamic dissociation constants (K_d) for the various DNA substrates were determined using a standard gel mobility shift assay as previously described (*27*). Briefly, the double-stranded DNA substrates were radiolabeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. Excess ATP was removed using a Bio-gel P6 column (Bio-Rad), and the reaction cocktail was diluted with buffer containing 10 mM Tris (pH 8.0), 1.0 mM EDTA, and 100 mM NaCl. The binding assay (30 μL) contained 0.1 nM DNA, 100 mM Tris (pH 8.0), 10 mM EDTA, 200 μg/mL BSA, 10 mM DTT, 20 μM sinefungin, and various M.*Eco*RI concentrations (0−5.0 nM). Dissociation constants were determined by fitting data obtained by densitometry to a standard hyperbolic binding isotherm.

Photochemical Cross-Linking Reaction. Small-scale reaction mixtures (30 μL) contained 4 nM ^{32}P-labeled DNA and 20 nM M.*Eco*RI in the binding assay buffer as described above except BSA was eliminated. Larger-scale reaction mixtures (up to 0.5 mL) contained 10 μM unlabeled DNA and 10 μM M.*Eco*RI in the identical buffer. All reactions were conducted in 1.5 mL micro-test tubes and the mixtures incubated on ice during the UV irradiation. The energy output of the tunable YAG 8010 Pump ND 6000 frequency-doubled dye (DCM in methanol) laser (Continuum Co.) at the various wavelengths was kept at 20 mJ/pulse (pulse width of 6 ns and laser beam diameter of 5 mm). All photo-cross-linking reactions were performed for 1−2 min. After UV irradiation, covalently cross-linked complexes were separated by 12% SDS−polyacrylamide gel electrophoresis (SDS−PAGE), and the photo-cross-linking yields were quantified by densitometry as described in ref *22*.

Dialysis and Proteolysis of the Covalent Complex. The cross-linked samples were dialyzed against 50 mM NH_4HCO_3 (pH 7.8) at 4 °C for 8 h (molecular weight cutoff of 14000−16000, 1 cm in diameter, Spectrum Co.). The samples were then digested by trypsin (Sigma) (5% of protein, w/w) or L-1-(tosylamido)-2-phenylethyl chloromethyl ketone-treated chymotrypsin (Boehringer Mannheim) (5% of protein) in 50 mM NH_4HCO_3 (pH 7.8) buffer at 37 °C for 12 h. The reaction was stopped by treatment for 10 min at 90 °C. Digests were analyzed by SDS−PAGE.

Purification of the Cross-Linked Adduct. The digested adducts were separated from free peptides by high-performance anion-exchange chromatography using a Vydac column (301VHP552, DEAE, 5 μM, 5 mm × 25 mm). A gradient of 0 to 60% buffer B [25 mM TEAA (pH 8.0) and 0.5 M NaCl] was performed in 20 min using a flow rate of 0.5 mL/min [buffer A, 25 mM TEAA (pH 8.0)]. The digested cross-linked adduct was detected by simultaneous monitoring at 215 and 260 nm (one spectrum per second) using a Waters 996 photodiode array detector. Collected fractions were further desalted with Sep-Pak Plus C_{18} cartridges (Waters Co.), followed by ethanol precipitation. Sample pellets were washed twice with 70% ice-cold ethanol and dried in a Savant Speed-vac.

ESI-MS on the Cross-Linked DNA−Peptide Adduct. All ESI-MS analyses were performed on a VG Platform II mass spectrometer (Fisons Instrument). The N_2 nebulizing gas was maintained at 100 psi, with a flow of the drying gas of 5.0 L/min. The purified cross-linked adduct was brought up in a 10 mM NH_4OAc/methanol (1:1) solution at a concentration of 5 pmol/μL. The samples were infused utilizing a syringe infusion pump 22 (Harvard Instruments) at a flow rate of 10 μL/min. Experiments were performed in the negative

ionization mode, and the ESI source temperature was kept at 120 °C. In most cases, 15 scans were averaged in the "MCA" mode over the mass to charge range (*m*/*z*) of 300−1700.

Nuclease Digestion of the Cross-Linked DNA−Peptide Adduct. Samples containing the cross-linked DNA−peptide adduct and uncomplexed DNA were treated with nuclease P1 (Pharmacia Biotech) and calf intestinal alkaline phosphatase (Pharmacia Biotech) to produce mononucleosides and peptide−mononucleoside adducts. Briefly, approximately 50 pmol of cross-linked adduct along with the uncomplexed DNA mixture (purified DEAE fractions) was treated with 1 unit of nuclease P1, 30 mM NaOAc (pH 5.3), and 2 mM $ZnCl_2$. Digestion was performed at 37 °C for 4 h. The solution was adjusted to 10 mM Tris-HCl, 50 mM KCl, and 1 mM $MgCl_2$ (pH 8.0) for the alkaline phosphatase digestion. Approximately 1 unit of alkaline phosphatase was added at 37 °C to digest the above mononucleotide samples overnight.

HPLC/ESI-MS on the Cross-Linked Nucleoside−Peptide Adduct. A Michrom Ultrafast HPLC system (Michrom Bioresources, Sunnyvale, CA), equipped with a micro-flow cell UV absorbance detector set at 215 nm, was used for all subsequent chromatography steps. The reverse-phase C_{18} HPLC column (1.0 mm × 150 mm, 5 μm, 300 Å, Michrom Bioresources) was used to separate the nucleoside−peptide adduct from the free nucleosides. A typical sample contained 1−10 pmol of nucleoside−peptide adducts injected onto the HPLC/MS system with 2% acetonitrile and 0.1% trifluoroacetic acid (solvent A) and 90% acetonitrile and 0.1% trifluoroacetic acid (solvent B). The column was eluted with a linear gradient of 0 to 70% solvent B over the course of 70 min at a flow rate of 50 μL/min.

To perform the on-line HPLC/ESI-MS analysis, the outlet of the UV cell was connected to the electrospray ion source with PEEK tubing (0.005 in. inside diameter). It was essential to use 0.005 in. tubing to maintain high resolution and decrease the "dead volume" between the column and the ion source. All MS analyses were carried out by ion spray MS using a Fisons VG Platform II quadrupole mass spectrometer. The ion spray voltage was set at 3.0 kV, and the N_2 nebulizer gas pressure was maintained at 100 psi. Experiments were performed in the positive ionization mode, and the source temperature was kept at 120 °C. The mass-to-charge ratio (*m*/*z*) range from 220 to 1420 Da was scanned once every second. Data were collected in centroid mode and analyzed with the MassLynx (Micromass Inc.) software.

Amino Acid Sequencing. Amino acid sequence analysis of the purified cross-linked adduct (100 pmol) was conducted at Amgen, Inc. (Thousand Oaks, CA), on an Applied Biosystems protein sequencer using standard procedures.

Photo-Cross-Linking of Crude Cell Extracts. Two different crude cell extract samples were prepared for photo-cross-linking. A cell extract containing M.*Eco*RI was prepared from *E. coli* strain MM294 harboring the overexpression plasmid pXRI. A control crude cell extract sample was obtained by using plasmid pSE380 (Invitrogen), which lacks the gene for M.*Eco*RI. Cells were grown at 37 °C in LB medium containing 100 μg/mL ampicillin to an optical density of 0.6 at 600 nm, and induced with isopropyl 1-thio-β-D-galactopyranoside to a final concentration of 1.0 mM. Cells were then grown for an additional 4 h at 37 °C and harvested by centrifugation. Cell pellets were washed with EB.2 buffer [0.2 M NaCl, 10 mM phosphate buffer, 1 mM EDTA, 7 mM βME, and 1 mM NaN_3 (pH 7.0)], resuspended in EB buffer, and lysed by sonification (4 × 20 s, μ-tip, energy output level = 2, 50% duty cycle, Branson 450 sonifier). After centrifugation, about 0.5 μg of each cell extract (1 or 10 μL of supernatants) was incubated with 4 nM ^{32}P-labeled duplex DNA (CTI-1/CBM) (small-scale) or 10 μM unlabeled DNA (large-scale) at room temperature for 20 min. Photo-cross-linking reactions were performed at 308 nm for 1 min as described above, and the results were analyzed by 12% PAGE or 12% SDS−PAGE.

Table 1: ESI-MS Data of Various Oligonucleotide Substrates Used in This Study[a]

sample	calcd mass (Da)	measured mass (Da)
CT [d(GGCGGAATTCGCGG)]	4345.1	4344.7 ± 0.43
CTI-1 [d(GGCGGAAITCGCGG)]	4456.1	4455.3 ± 0.08
CTI-2 [d(GGCGGAATICGCGG)]	4456.1	4454.6 ± 0.05
CB [d(CCGCGAATTCCGCC)]	4184.1	4183.6 ± 0.54
CBM [d(CCGCGAMTTCCGCC)]	4198.1	4197.4 ± 0.29
BMI-1 [d(CCGCGAMITCCGCC)]	4310.7	4308.9 ± 0.05
BMI-2 [d(CCGCGAMTICCGCC)]	4310.7	4308.8 ± 0.60

[a] The 14mer oligonucleotides contain the M.*Eco*RI recognition site (underlined), in which I is 5-iodouridine and M is N^6-methylated adenine.

RESULTS

Formation of the Cross-Linked M.EcoRI−DNA Complex. Prior to the photo-cross-linking experiment, it was necessary to confirm the incorporation of the modified bases in the oligonucleotides and to examine the affinities of binding of the M.*Eco*RI to these DNA substrates. Our previous study showed that M.*Eco*RI binds its methylated DNA products poorly (*31*). Therefore, we used hemimethylated DNA duplexes in all photo-cross-linking studies to force the M.*Eco*RI to bind in only one orientation. The DNA substrates used in the study were (M.*Eco*RI recognition site is in bold and underlined; I, 5-iodouracil; M, N^6-methylated adenine) (a) CT/CBM [d(GGCG**GAATTC**GCGG)/d(CCGC**GAMTTC**CGCC)], (b) CTI-1/CBM [d(GGCG**GAAITC**GCGG)/d(CCGC**GAMTTC**CGCC)], (c) CTI-2/CBM [d(GGCG**GAATIC**GCGG)/d(CCGC**GAMTTC**CGCC)], (d) CT/BMI-1 [d(GGCG**GAATTC**GCGG)/d(CCGC**GAMITC**CGCC)], and (e) CT/BMI-2 [d(GGCG**GAATTC**GCGG)/d(CCGC**GAMTIC**CGCC)].

Negative ionization electrospray mass spectrometric analysis was utilized to obtain the accurate mass of these oligonucleotides. As shown in Table 1, the mass data agreed well with the incorporation of a methyl group and/or an iodine atom into the oligonucleotides. The ESI mass spectra exhibit excellent mass resolution and accuracy. Only a few picomoles of DNA was needed to generate conclusive spectra under the optimal mass spectrometric conditions.

Substitution with 5-IdU at single positions within the recognition site was used to probe the M.*Eco*RI−DNA interface. Since the van der Waals radius of iodine is only 8% larger than the methyl group of thymine, the replacement of thymine with 5-IdU should leave the DNA−protein complex relatively unchanged (*18*, *23*). M.*Eco*RI binds to a double-stranded substrate in which the upper strand is substituted with 5-IdU in place of the first thymine (CTI-

Table 2: ESI-MS Data of the Cross-Linked Adducts[a]

sample	calcd mass (Da)	measured mass (Da)
M.*Eco*RI (325 amino acids)	37913.4	37916.9 ± 30.52
DNA−M.*Eco*RI complex	42241.5	42242.9 ± 21.28
CTI-I* [d(GGCGGAAUTCGCGG)]	4329.1	4328.2 ± 0.53
CTI-I* and T29 (2123 Da)[b]	6450.03	6450.0 ± 0.53
CTI-I* and T29-GVSGF (1676 Da)[b]	6003.1	6003.8 ± 1.62
CTI-I* and Y30-32 (2341 Da)	6669.1	6668.3 ± 2.41
T29 (2123 Da) and dU (228 Da)	2351.03	2350.9 ± 0.61

[a] Abbreviations: CTI-I*, CTI with the loss of an iodine atom; T29, trypsin-digested peptide (GVSGFIVPEHYEL**Y**GTEAR); T29-GVSGF, photocleaved peptide T29; Y30-32, incomplete chymotrypsin-digested peptide (LGVHLGRGVSGFIVPEHYEL**Y**). The cross-linked site Y^{204} is in bold. [b] Sequences of the cross-linked peptides (T29 and T29-GVSGF) were also confirmed by Edman amino acid sequencing.

1/CBM) with similar affinity (K_d = 0.55 ± 0.02 nM) for the thymine-substituted duplex (CT/CBM) (0.43 ± 0.08 nM) (*22*). The substitutions of 5-IdU for the three other thymine positions were also tested with the electrophoretic mobility shift assay to determine whether the 5-IdU substitution perturbed the DNA−M.*Eco*RI complex. Our results indicated that none of these 5-IdU-substituted probes significantly altered the stability of the complex (data not shown). We therefore used these 5-IdU-containing substrates for the spatial probing studies. CTI-1/CBM gave the highest cross-linking efficiency (∼50%), and the efficiencies of the other three were 5% (CTI-2/CBM), 3% (CT/BMI-1), and 2% (CT/BMI-2). Purified M.*Eco*RI and the DNA−protein mixture without UV irradiation did not result in any complex band in the denaturing PAGE. The cross-linked DNA−protein complex was also identified in our previous ESI-MS study (*22*). The total mass (42 243 Da) of the complex was consistent with irradiation-dependent covalent bond formation between one strand of DNA (CTI with the loss of an iodine atom) and the protein (Table 2). No cross-linked complexes could be detected by ESI-MS in the control samples. Large-scale photo-cross-linking reactions (10 μM) were also performed to generate more cross-linked samples for the ESI-MS study. As can be seen in Figure 1, the cross-linking yields for the 5-IdU-substituted DNA analogue (CTI-1/CBM) and the non-IdU-containing hemimethylated DNA substrate (CT/CBM) are ∼50 and <3%, respectively (lanes 3 and 5).

Isolation and Identification of the Peptide−Oligonucleotide Cross-Linked Adduct. To identify the peptide involved in the protein−DNA cross-link, the sample was subjected to extensive proteolysis at 37 °C. To verify whether the cross-linked adduct can undergo such lengthy proteolysis, the cross-linked complex was incubated at 37 °C for 24 h and at 90 °C for 20 min. The sample was also dialyzed against the buffer that was suitable for proteolysis. Then, protease was added to the sample and incubated at 37 °C for ∼12 h. The result of the complete proteolysis was demonstrated by SDS−PAGE (Figure 1, lanes 6 and 7). Trypsin and chymotrypsin were independently used to generate two sets of peptides.

Following the digestion with trypsin or chymotrypsin, the cross-linked peptide−DNA adducts were separated from non-cross-linked free peptides by anion-exchange chromatography. The eluted samples were monitored at 215 and 260 nm simultaneously (Figure 2). The non-cross-linked peptides

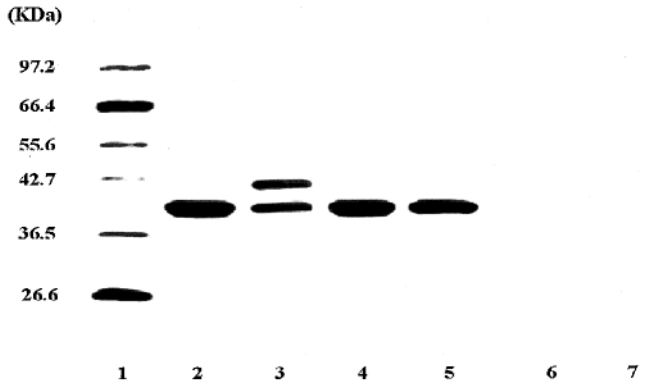

FIGURE 1: Coomassie-stained 12% SDS−PAGE of photo-cross-linking samples. Large-scale cross-linking reactions were performed using 10 μM M.*Eco*RI and 10 μM DNA (CTI-1/CBM or CT/CBM). Densitometric analysis indicates an ∼50% cross-linking yield: lane 1, protein markers; lanes 2 and 3, M.*Eco*RI with DNA substrate (CTI-1/CBM) (without and with irradiation); lanes 4 and 5, M.*Eco*RI with DNA substrate (CT/BM) (without and with irradiation); lane 6, complete tryptic digestion on the cross-linked complex as in lane 3; and lane 7, complete tryptic digestion on the cross-linked complex as in lane 5.

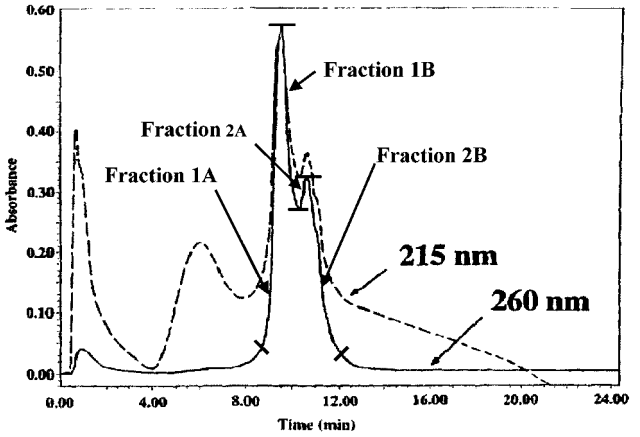

FIGURE 2: HPLC chromatograms of the DEAE column separation of the trypsin-digested cross-linked complex. The photo-cross-linked complex (100 pmol) was digested with trypsin and subjected to DEAE anion-exchange column separation. The cross-linked peptide−oligonucleotide adducts were eluted in 10.5 min (after the major DNA peak). HPLC chromatograms were monitored at 215 and 260 nm simultaneously.

eluted early in the gradient followed by the protease (trypsin), the free oligonucleotides, and the peptide−oligonucleotide adducts. The HPLC chromatogram shows that the cross-linked complexes (fractions 2A and 2B) have a higher A_{215}/A_{260} ratio than the free DNA (fractions 1A and 1B). Peak fractions were analyzed by 12% PAGE. As can be seen in Figure 3, bands of cross-linked peptide−DNA complexes migrate more slowly than the corresponding free DNAs (lanes 5 and 6). Fractions 2A and 2B, which contain the peptide−DNA adducts, were first desalted using a reverse-phase Sep-pak cartridge and then further analyzed by electrospray ionization mass spectrometry.

Cross-linked peptides were analyzed by both positive and negative ESI-MS methods. The total ion current measured in the positive mode was similar to the total ion current measured in the negative mode (data not shown). However, the ion signals corresponding to the free oligonucleotides were much weaker, and sodium adducts were quite abundant in the positive ion ESI mode. Four ion species were detected

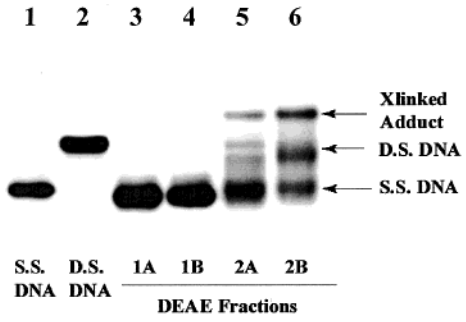

FIGURE 3: PAGE (12%) of the DEAE column-separated sample. DEAE column fractions (eluted from 8 to 12 min) were collected separately (0.5 mL/fraction). HPLC fractions were dried and radiolabeled with [γ-^{32}P]ATP. The upper bands seen in lanes 5 and 6 represent the cross-linked peptide−oligonucleotide adduct(s): lane 1, single-stranded DNA; lane 2, double-stranded DNA; and lanes 3−6, DEAE fractions 1A, 1B, 2A, and 2B, respectively.

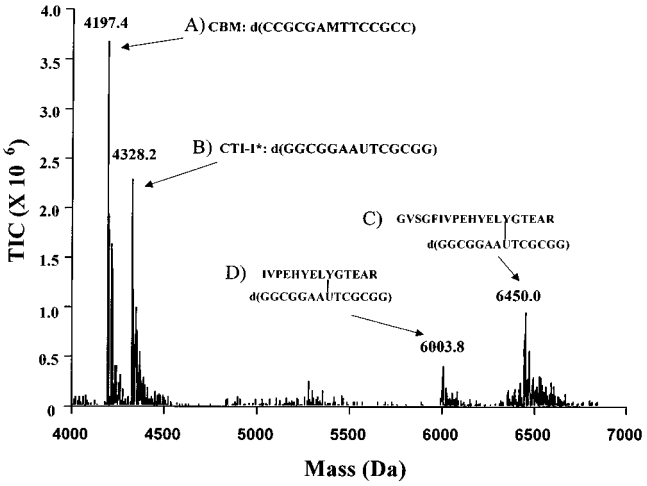

FIGURE 4: Electrospray ionization mass spectrum of DEAE fraction 2B. Fraction 2B was dissolved into a 10 mM NH_4OAc/methanol mixture (1:1, v/v) at about 5 pmol/μL. Samples were infused at a flow rate of 10 μL/min into the electrospray interface and detected in the negative ionization mode. Two single-stranded oligonucleotides and two cross-linked adducts can be seen.

from DEAE fraction 2B (Figure 4), which is consistent with the PAGE results. The deconvoluted mass spectrum exhibits two signals (4328.2 and 4197.4 Da) that can be assigned to the top strand (loss of the iodine atom) and bottom strand of the duplex oligonucleotide according to their theoretical masses. The peak with a mass of 6450 Da corresponds to the cross-linked tryptic digested peptide [T29 (GVSGFIVPEHYELYGTEAR)] and the top strand of DNA. The presence of a minor peak (6003.8 Da) in the tryptic digest suggested that another cross-linked adduct which represented the photocleaved peptide of T29 [T29-GVSGF (IVPEHYELYGTEAR)] was present. Similar ESI-MS experiments with the chymotrypsin-digested cross-linked adduct were also performed. A mass value of an ion species (6668.3 Da) was detected and assumed to represent the incomplete digested peptide [Y30−Y32 (LGVHLGRGVSGFIVPEHYELY)] cross-linked to the top DNA strand. The molecular weight of this sequence was in excellent agreement with the predicted mass (Table 2). The peptide fragments derived from both trypsin and chymotrypsin digestions are located in the same region of M.*Eco*RI.

LC/MS Analysis of the Peptide−Nucleoside Adduct. Nuclease P1 and alkaline phosphatase digestion of the peptide−oligonucleotide adduct was used to more precisely map the cross-linking site. This double digestion should generate a deoxyuridine−peptide species and six free mononucleosides (G, C, A, T, dU, and N^6-methylated A). Figure 5 shows the LC/MS data from approximately 10 pmol of the peptide−nucleoside adduct (fraction 2B). Two major peaks shown in the mass spectrum correspond to the free nucleosides and the T29−dU cross-linked adduct. All six free mononucleosides were detected in the first HPLC/MS elution peak (first major peak). A series of ion species (*m*/*z* 470.5, 587.5, 783.7, and 1175.0) were also identified for the second major peak in the mass spectrum, corresponding to the +5 to the +2 charge state of the T29−dU cross-linked adduct (Figure 5, inset). The total molecular mass of the ion species (2350.9 Da) closely matched the predicted value of a 2123 Da peptide (T29) cross-linked to a dU moiety (228 Da). This result again matched well with the above ESI-MS data of the peptide−oligonucleotide adduct. However, our LC/MS data could not confirm whether the minor peak prior to the major cross-linked adduct peak was derived from the other cross-linked peptide (T29-GVSGF) adduct.

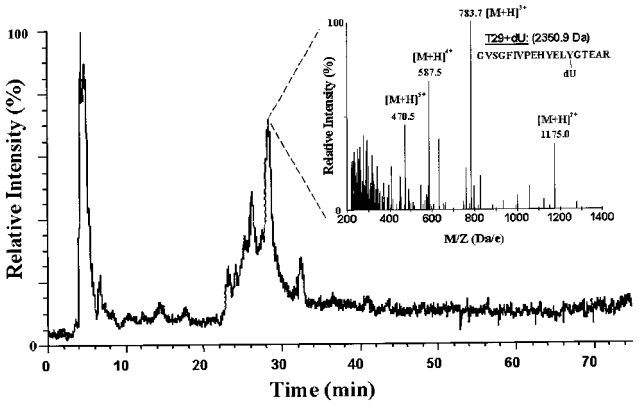

FIGURE 5: Electrospray ionization mass spectrum of the peptide−nucleoside adduct. In the inset is shown the background-subtracted mass spectrum of the ion signal peak at 28.33 min showing a series of charged species corresponding to a cross-linked peptide−nucleoside adduct.

Photo-Cross-Linking Occurs at Tyrosine 204 of M.EcoRI. To confirm the amino acid sequences of the identified cross-linked peptides, purified trypsin-digested cross-linked adducts (100 pmol) were subjected to automated peptide sequencing by Edman degradation for 15 cycles. The two peptide sequences were found to be GVSGFIVPEHYEL**X**G and IVPEHYEL**X**GTEAR, in which X represents a nonstandard amino acid. These peptide sequences are consistent with amino acid residues 191−205 and 196−209 of M.*Eco*RI, respectively. Thus, the ESI-MS data of tryptic and chymotrypsin-digested cross-linked adducts, together with the Edman degradation result, strongly indicate that photo-cross-linking of DNA occurred to Tyr 204 of M.*Eco*RI. Moreover, our mass spectrometry studies and amino acid sequencing data also indicate that photodegradation occurred, involving the peptide bond between Phe 195 and Ile 196 (T29). The results are consistent with our previous observation from 12% SDS−PAGE of the cross-linked complex (Figure 3A in ref *22*).

Photo-Cross-Linking of Crude Cell Extract Samples. Photo-cross-linking experiments of the small-scale cell extract were performed at 308 nm with UV irradiation for 1

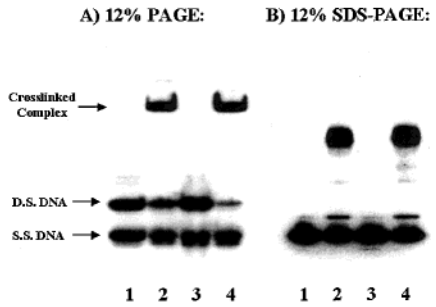

FIGURE 6: Photo-cross-linking of 5-IdU-substituted DNA to cell crude extracts (DNA analysis). (A) PAGE (12%) of the UV-irradiated (308 nm, 1 min) reaction mixtures (small-scale) containing 4 nM ^{32}P-labeled DNA (CTI-1/CBM) and various protein samples at approximately 20−100 nM (lanes 1−4, DNA duplex, purified M.*Eco*RI with DNA, pSE380 cell extract with DNA, and pXRI cell extract with duplex, respectively). (B) Denaturing SDS−PAGE (12%) of the same reaction mixtures as described for panel A.

min. To verify that the newly formed band is a specific covalent DNA−M.*Eco*RI complex, the following experiments were performed. First, an excess amount of λ-DNA was incubated with the duplex DNA (CTI-1/CBM) to protect it from nuclease cleavage. However, irradiation of this reaction mixture did not result in higher cross-linking yields or changes in nonspecific photo-cross-linking (data not shown). Thus, nucleases in the cell extract do not interfere with our procedures, and the λ-DNA did not compete with the DNA duplex for the M.*Eco*RI binding. Second, a cell extract without M.*Eco*RI was prepared by using plasmid pSE380, and irradiation of this cell extract in the presence of ^{32}P-labeled DNA duplex did not produce the band representing the DNA−M.*Eco*RI complex (Figure 6). Only one major cross-linked complex was seen in the M.*Eco*RI cell extract sample, and it has the same mobility as the band formed with pure M.*Eco*RI. The large-scale photo-cross-linking reaction was also performed with the M.*Eco*RI cell extracts. Figure 7 shows a Coomassie-stained denaturing gel of the M.*Eco*RI cell extract cross-linked to the DNA duplex. The cross-linked M.*Eco*RI−DNA complexes were clearly seen in the irradiated samples (lanes 5 and 6). No other covalent complexes are evident when comparing lane 4 (M.*Eco*RI cell extract, non-UV) with lane 6 (with UV). An ~50% cross-linking yield was achieved for this large-scale reaction.

DISCUSSION

Ultraviolet (UV) light-induced cross-linking of proteins to nucleic acids is a powerful technique for studying in vitro and in vivo DNA−protein interactions (*19*−*21*). Irradiation brings about the formation of a "zero-length" covalent bond between the protein and the nucleic acid to which it is bound. Several photoactivatible deoxynucleoside analogues, such as 5-bromouracil (5-BrU), 4-thiouridine, and 5-iodoracil (5-IdU), have been developed (*16*, *18*, *28*). However, the success of 5-BrU- or 4-thiouridine-based nucleic acid cross-linking to proteins is limited in part by low cross-linking yields (*21*, *28*, *29*). In addition, the identification of amino acids using these methods has historically required preparative-scale reactions using homogeneous nucleic acid and protein (up to milligram quantities), and purification of large quantities of cross-linked adduct (1−50 nmol) for sequence analysis

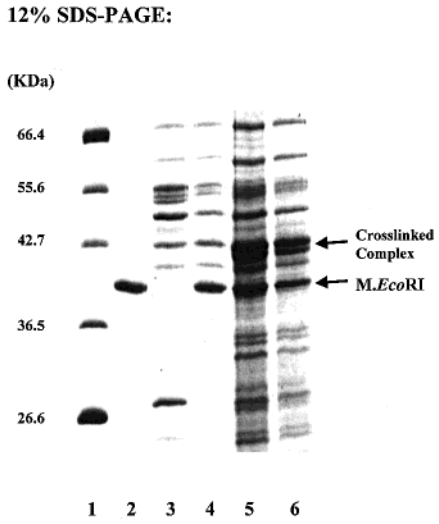

FIGURE 7: Photo-cross-linking of 5-IdU-substituted DNA to cell crude extracts (protein analysis). Coomassie-stained 12% SDS−PAGE of the reaction mixtures (large-scale) containing 10 μM 5-IdU-substituted DNA (CTI-1/CBM) and various protein samples (10 μM M.*Eco*RI or 5 μg of cell extract samples) [lanes 1−6, protein marker, pure M.*Eco*RI, pSE380 cell extract, 5 μg (1×) of pXRI cell extract, 25 μg (5×) of pXRI cell extract, and 1× pXRI cell extract, respectively]. Samples in lanes 5 and 6 were UV irradiated at 308 nm for 1 min.

(*19*, *20*, *29*, *30*). We describe a highly sensitive photo-cross-linking mass spectrometry analytical procedure for identifying amino acid residues involved in protein−DNA recognition using only picomole amounts of both protein and DNA as starting materials.

Our photo-cross-linking experiments using 5-IdU-substituted DNA to map the interface between DNA and the M.*Eco*RI revealed that the methyl group of the central thymine at the top DNA strand was in close contact with the protein. This result is consistent with our earlier studies probing the functional significance of the thymine methyl moieties within the M.*Eco*RI recognition site (GAA**TT**C) (*31*). Removing the methyl group of the central thymine at the top strand decreased K_m^{DNA} by 5-fold, indicating this methyl group has a negative contribution to specificity. The replacement of this thymine with 5-IdU slightly increased its K_d value, presumably due to the larger iodine atom (*22*). Moreover, the absence of the methyl groups at the other three thymine positions only caused slight changes in K_m^{DNA}, while their 5-IdU-substituted DNA substrates bound to the M.*Eco*RI with affinities similar to that of the unmodified DNA (data not shown). Our cross-linking results showed that CTI-1/CBM gave the highest cross-linking yield (50−60%), and the efficiencies of the other three substrates (CTI-2/CBM, CT/BMI-1, and CT/BMI-2) were 5, 3, and 2%, respectively. Thus, our original functional analysis and the photo-cross-linking mapping data strongly suggest that the first thymine of the target site (GAA**T**TC) is in proximity to the M.*Eco*RI.

Our previous ESI-MS analysis of the cross-linked protein−DNA complex showed that a single strand of DNA (CTI with the loss of an iodine atom) formed a covalent bond with the enzyme with 1:1 stoichiometry (*22*). In the study presented here, two proteolytic digestions were performed independently to generate different peptide sequences; the ESI-MS analysis identified an incompletely digested peptide−nucleoside adduct in the chymotrypsin digestion. Our

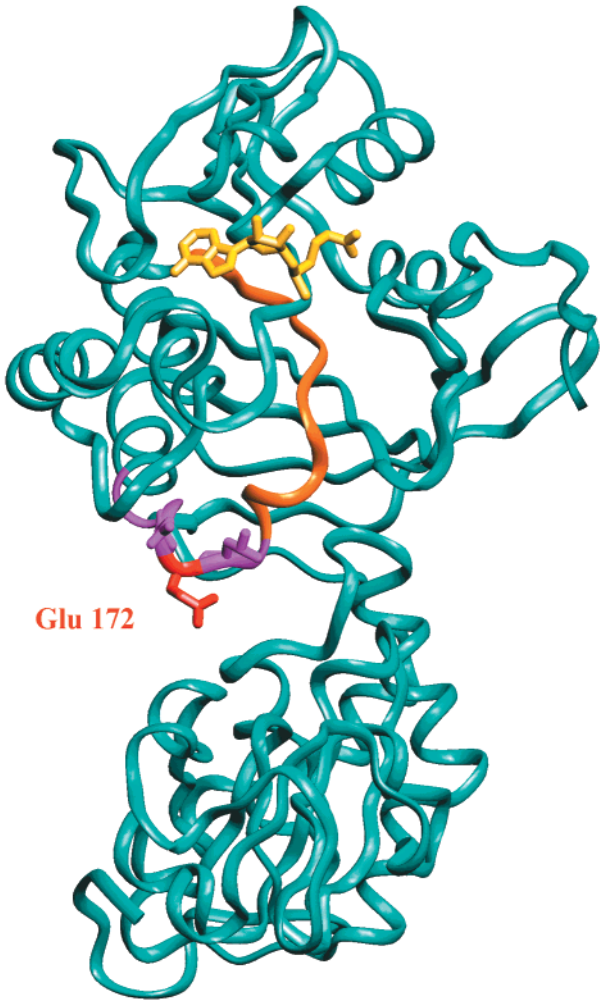

FIGURE 8: Crystal structure of M.*Taq*I in ribbon representation. Structure of M.*Taq*I (blue) in complex with the cofactor AdoMet (gold) showing that motif VI (orange), which was previously suggested to have a catalytic role (*7*), forms part of the surface of the pocket that may stabilize the extrahelical adenine target base. Amino acid E^{172} (red) located in the loop region (purple) is in the same position as Y^{204} in M.*Eco*RI.

three independent ESI-MS analyses of the DEAE anion-exchange fractions containing either trypsin- or chymotrypsin-digested cross-linked peptide−oligonucleotide adducts resulted in a common peptide sequence IVPEHYEL**Y**, in which the last tyrosine (Y^{204}) is most likely the site of cross-linking. The Edman peptide sequencing data of the trypsin-digested adducts confirmed this assignment. Previous studies suggested that tyrosine is frequently the site for photo-cross-linking (*19*, *29*, *30*). We conclude that amino acid Tyr 204 in M.*Eco*RI directly interacts with the methyl moiety of the first thymine via the major groove of the DNA.

The ESI-MS approach described here is extremely sensitive and has high mass accuracy. We have identified approximately 10−50 pmol of the cross-linked peptide−oligonucleotide adduct and 10 pmol of the peptide−mononucleoside adduct in this study. Furthermore, in contrast to an earlier report (*24*) using ESI-MS detection on similar peptide−oligonucleotide heteroconjugates, our negative ESI-MS data provide better ion signals. We also carried out MALDI mass analysis of the same heteroconjugate with several common matrixes for peptides and nucleic acids (*24*). However, only the free oligonucleotides could be detected.

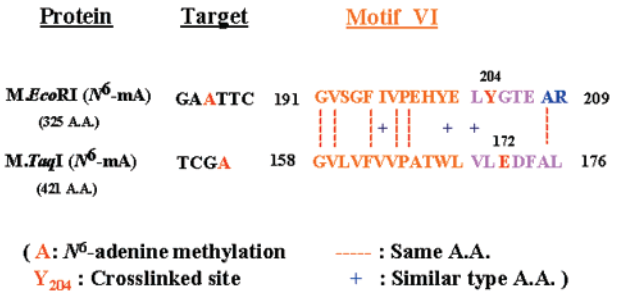

FIGURE 9: Sequence alignment (motif VI) of M.*Eco*RI with M.*Taq*I. The amino acid sequence alignment of the conserved motif VI (orange) from two N^6-adenine DNA Mtases (M.*Eco*RI and M.*Taq*I). Tyr^{204} of M.*Eco*RI and Glu^{172} of M.*Taq*I in the loop regions (purple) between motif VI and VII are positioned similarly.

Sequence comparisons of the two N^6-adenine DNA methyltransferases (group γ), M.*Eco*RI and M.*Taq*I, reveal that these proteins are quite similar (*9*). Both are bilobal and have nine conserved motifs in the same sequential order. The identified cross-linked peptide (T29) in M.*Eco*RI is located in motif VI that was not previously implicated in DNA recognition. However, the X-ray structure of M.*Taq*I shown in Figure 8 suggests that the loop structure (colored purple) between motif VI and motif VII directly contacts the DNA. The sequence alignment of motif VI (Figure 9) implies that Y^{204} in T29 of M.*Eco*RI and E^{172} in M.*Taq*I (colored red) are positioned similarly in relationship to the DNA. This suggests that DNA MTases recognize specific DNA sequences through peptide elements beyond those residing in "core" DNA recognition segments. Our results therefore suggest that protein engineering efforts aimed at altering the sequence specificity of these enzymes should consider such specificity determinants.

Our successful in vivo photo-cross-linking method provides a valuable analytical approach for identifying proteins and peptides involved in DNA binding. The high cross-linking yield (>50%) for the cell extract sample should generate large quantities of DNA−protein complexes and be amenable to in-gel digestion, which may be readily extended to the identification of the cross-linked amino acids using the combination of two-dimensional gel electrophoresis and mass spectrometric analysis techniques (*32*). Our goal is to provide a fast and sensitive method for identifying the amino acid(s) involved in the protein−DNA (RNA) interactions directly from the whole cell extracts.

ACKNOWLEDGMENT

We kindly thank Prof. Peter Ford and Dr. Steve Massick for their assistance in optimizing the laser instrument. We also thank Dr. Scott Patterson and Paul Courchesne at Amgen, Inc., for performing the Edman amino acid sequence analysis.

REFERENCES

1. Rubin, R. A., and Modrich, P. (1977) *J. Biol. Chem. 252*, 7265−7272.
2. Dickerson, R. E., and Drew, H. R. (1981) *J. Mol. Biol. 149*, 761−786.
3. Kim, Y. C., Grable, J. C., Love, R., Greene, P. J., and Rosenberg, J. M. (1990) *Science 249*, 1307−1309.
4. Rosenberg, J. M. (1991) *Curr. Opin. Struct. Biol. 1*, 104−113.

5. Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. (1997) *Nucleic Acids Res. 25*, 2702−2715.
6. Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994) *Cell 76*, 357−369.
7. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I., and Saenger, W. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10957−10961.
8. Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. (1994) *J. Mol. Biol. 238*, 626−629.
9. Malone, T., Blumenthal, R. M., and Cheng, X. (1995) *J. Mol. Biol. 253*, 618−632.
10. Reich, N. O., and Mashhoon, N. (1991) *Biochemistry 30*, 2933−2939.
11. Garcia, R. A., Bustamante, C. J., and Reich, N. O. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 7618−7622.
12. Allan, B. W., Reich, N. O., and Beechem, J. M. (1999) *Biochemistry 38*, 5308−5314.
13. Allan, B. W., Beechem, J. M., Lindstrom, W. M., and Reich, N. O. (1998) *J. Biol. Chem. 273*, 2368−2373.
14. Reich, N. O., and Mashhoon, N. (1993) *J. Biol. Chem. 268*, 9191−9193.
15. Gott, J. M., Willis, M. C., Koch, T. H., and Uhlenbeck, O. C. (1991) *Biochemistry 30*, 6290−6295.
16. Lin, S. Y., and Riggs, A. D. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 947−951.
17. Meffert, R., and Dose, K. (1988) *FEBS Lett. 239*, 190−194.
18. Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. R., and Koch, T. H. (1993) *Science 262*, 1255−1257.
19. Allen, T. D., Wick, K. L., and Matthews, K. S. (1991) *J. Biol. Chem. 266*, 6113−6119.
20. Blatter, E. E., Ebright, Y. W., and Ebright, R. H. (1992) *Nature 359*, 650−652.
21. Chen, A., Powell, L. M., Dryden, D. T., Murray, N. E., and Brown, T. (1995) *Nucleic Acids Res. 23*, 1177−1183.
22. Wong, D. L., Pavlovich, J. G., and Reich, N. O. (1998) *Nucleic Acids Res. 26*, 645−649.
23. Holz, B., Dank, N., Eickhoff, J. E., Lipps, G., Krauss, G., and Weinhold, E. (1999) *J. Biol. Chem. 274*, 15066−15072.
24. Jensen, O. N., Kulkarni, S., Aldrich, J. V., and Barofsky, D. F. (1996) *Nucleic Acids Res. 24*, 3866−3872.
25. Becker, C. R., Efcavitch, J. W., Heiner, C. R., and Kaiser, N. F. (1985) *J. Chromatogr. 326*, 9293−9305.
26. Greene, P. J., Heyneker, H. L., Bolivar, F., Rodriguez, R. L., Betlach, M. C., Covarrubias, A. A., Backman, K., Russel, D. J., Tait, R., and Boyer, H. W. (1978) *Nucleic Acids Res. 5*, 2373−2380.
27. Surby, M. A., and Reich, N. O. (1996) *Biochemistry 35*, 2209−2217.
28. McGregor, A., Rao, M. V., Duckworth, G., Stockley, P. G., and Connolly, B. A. (1996) *Nucleic Acids Res. 24*, 3173−3180.
29. Willis, M. C., Lecuyer, K. A., Meisenheimer, K. M., Uhlenbeck, O. C., and Koch, T. H. (1994) *Nucleic Acids Res. 22*, 4947−4952.
30. Liu, Y. P., Wang, Z., and Rana, T. M. (1996) *J. Biol. Chem. 271*, 10391−10396.
31. Reich, N. O., and Danzitz, M. J., Jr. (1992) *Biochemistry 31*, 1937−1945.
32. Hall, S. C., Smith, D. M., Masiarz, F. R., Soo, V. W., Tran, H. M., Epstein, L. B., and Burlingame, A. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1927−1931.

BI001164V