Identification of the Template-Binding Cleft of T7 RNA Polymerase as the Site for Promoter Binding by Photochemical Cross-linking with Psoralen†

Srinivas S. Sastry*

Laboratory of Molecular Genetics, The Rockefeller University, Box 174, 1230 York Avenue, New York, New York 10021

Received April 24, 1996; Revised Manuscript Received July 25, 1996⊗

ABSTRACT: We describe a novel method of photo-cross-linking DNA-binding proteins to DNA employing psoralen as a tether. We apply this method for the interaction of T7 RNA polymerase to its promoter. The crystallographic model of T7 RNA polymerase shows a cleft formed by the palm, thumb, and fingers domains. It was proposed that template DNA binds in the cleft. Here we directly and positively identify, in solution, the cleft as the seat of template binding. We photo-cross-linked a 23 bp promoter DNA to T7 RNA polymerase. We then determined the masses of cross-linked tryptic peptides by mass spectrometry and analyzed their amino acid composition. The cross-linked peptides were projected on the crystal structure of T7 RNA polymerase. The peptides nicely decorated the back, front, and side wall of the cleft. In a previous work [Sastry et al. (1993) *Biochemistry 32,* 5526−5538] we used site-specific psoralen furan-side monoadducts for cross-linking DNAs to DNA-binding proteins. We cross-linked a single-stranded 12-mer oligonucleotide to T7 RNA polymerase. We isolated and purified a DNA cross-linked tryptic peptide. We then used mass spectrometry and amino acid composition analysis to identify the location of this peptide on the T7 RNA polymerase primary sequence. In the present work we have mapped this peptide on the 3-D structure of T7 RNA polymerase. This peptide maps in the fingers domain of the polymerase. On the basis of a comparison of the map positions of peptides that cross-linked to either promoter DNA or single-stranded oligo-DNA, we propose that different functional domains may be involved in binding of double-stranded promoter DNA and nonspecific single-stranded DNA. Whereas the cleft of the polymerase is the seat of double-stranded promoter binding, the fingers domain may be used by the polymerase to grab single-stranded DNA (or RNA) in a nonspecific manner. Alternatively, the single-stranded oligo binding site may be an RNA product-binding site during transcription. The photochemical techniques we have developed [Sastry et al. (1993) *Biochemistry 32,* 5526−5538; this work] can be applied to other DNA-protein complexes to map DNA−binding domains.

T7 RNA polymerase (T7 RNAP;[1] 98.8 kDa) belongs to a class of single-subunit RNAPs that includes T3 and SP6 phage RNAPs (Chamberlin & Ryan, 1982; McAllister, 1993). It can be purified to homogeneity in milligram quantities using relatively straightforward procedures (Goldberg & Dunn, 1988). A wealth of mutagenesis data has revealed the amino acid residues important for catalysis and DNA binding [e.g., see Bonner et al. (1992, 1994a,b), Gross et al. (1992), Mookthiar et al. (1991), Osumi-Davis et al. (1994), and Patra et al. (1992)]. So far, T7 RNAP is the only RNAP whose X-ray crystal structure has been solved (Sousa et al., 1993). In vitro, T7 RNAP transcribes DNA without additional protein factors. These features make this "simple" polymerase a very attractive model RNAP for structural studies using physicochemical methods. T7 RNAP−promoter complexes have been characterized by footprinting and low-resolution NMR (Rastinejad & Lu, 1993; Muller et al., 1989). Elongation complexes have been characterized by site-specifically placing a psoralen road block in the path of the polymerase (Sastry & Hearst, 1991a,b). The 3-D structure of T7 RNAP shows high α-helicity with a deep cleft (∼60 Å length × 15−25 Å width × 25−40 Å depth) that has been proposed to accommodate template DNA (Sousa et al., 1993). T7 RNAP has a striking structural similarity to *Escherichia coli* DNA polymerase Klenow fragment and HIV-1 RT (Joyce & Steitz, 1994, 1995; Sousa et al., 1993).

In a previous work we used psoralen furan-side monoadducts as photoreagents for cross-linking DNA-binding proteins to DNA (Sastry et al., 1993a,b). Here, a more direct cross-linking approach is presented. Psoralens are a class of three-ring heterocyclic compounds consisting of a furan linearly fused to a coumarin. Psoralens photochemically alkylate DNA by reacting principally with pyrimidines in nucleic acids (Cimino et al., 1985). Thymidine is most reactive. The three main adducts between thymidine and psoralen and photochemical reaction pathways are shown in reference (Sastry et al., 1993a). Photoreactions take place at the 3,4 or 4′,5′ double bonds of psoralen and the 5,6 double bond in pyrimidines (see Figure 6 for the structure of a psoralen). The planar psoralen first intercalates between base

† S.S.S. is a Louis B. Mayer Foundation Fellow. Financial resources from the Mayer Foundation are gratefully appreciated. The Hewlett-Packard Co. is thanked for an instrumentation grant.

* Corresponding author. Telephone 212-327-8987. FAX 212-327-8651. E-mail: sastrys@rockvax.rockefeller.edu.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1996.

[1] Abbreviations: amu, atomic mass units; DMSO, dimethyl sulfoxide; HPLC, high-performance liquid chromatography; IR, infrared; LC/MS, liquid chromatography/mass spectrometry; Maf, psoralen furan-side monoadduct; T7 RNAP, bacteriophage T7 RNA polymerase; XL, T7 RNA polymerase−DNA cross-link; UVA, long-wave ultraviolet light; UM, unmodified DNA; 8-MOP, 8-methoxypsoralen; HMT, 4′-hydroxymethyl-4,5′,8-trimethylpsoralen; RT, reverse transcriptase; TCA, trichloroacetic acid; TPCK, L-1-(tosylamino)-2-phenylethyl chloromethyl ketone; TFA, trifluoroacetic acid; TBE, tris−borate−EDTA buffer.

S0006-2960(96)00982-8 CCC: $12.00

pairs of a ds DNA (or RNA), and when exposed to long-wave UV (320−400 nm; UVA), the intercalated psoralen reacts by [2+2] cycloaddition first forming either furan-side or pyrone-side monoadducts. Absorption of a second photon converts the furan-side monoadduct to a cross-link with the adjacent pyrimidine on the complementary DNA or RNA strand. The 3-D structures of the psoralen−DNA furan-side monoadduct and cross-link were determined by NMR (Spielmann et al., 1995). Psoralens also react efficiently with proteins by as yet unknown photochemical mechanisms (Lerman et al., 1980; Megaw et al., 1980; Midden, 1988; Schiavon & Veronese, 1986; Veronese et al., 1981, 1982; Yoshikawa et al., 1979) [see ref Sastry et al. (1993b) for a discussion of the chemistry of cross-linking]. Noncovalent affinity of psoralen to proteins varies quite dramatically depending on the protein and the psoralen derivative. It may be higher for proteins than with DNA. For example, free 8-MOP or angelicin (an angular isomer of psoralen) bound BSA or DNA polymerase noncovalently with higher affinity [$\sim 2 \times 10^4$ M^{-1} (Granger et al., 1982; Rodighiero et al., 1988)] than the reported affinity with DNA [$\sim 2.5 \times 10^3$ M^{-1} (Isaacs et al., 1977)]. Noncovalent binding affinities of psoralen derivatives to protein could not be directly correlated to photobinding (Fredericksen & Hearst, 1979; Midden, 1988). According to one study, 8-MOP formed one photoadduct per 103 amino acids, whereas with DNA the corresponding value was one adduct per 1000 bp (Schmitt et al., 1995).

In an effort to understand the structure of T7 RNAP−DNA complexes, for which as yet no high-resolution structure is available, we used psoralen as a photochemical probe. Here we identified the cleft in the 3-D structure of T7 RNA polymerase as the seat of promoter interactions. Our results nicely fit in with the crystallographic model (Sousa et al., 1993) and confirm that, in solution, the promoter DNA indeed binds in the cleft. We found that T7 RNAP (and other DNA-binding proteins) can be cross-linked to DNA with added psoralen. Either 8-MOP or HMT can be used for cross-linking. The yields of direct cross-linking reactions are comparable or sometimes higher than cross-linking using furan-side monoadducts. Our photo-cross-linking methods can be potentially used to identify DNA-binding domains in DNA-binding proteins.

MATERIALS AND METHODS

DNAs and Proteins. 8-MOP was purchased from ICN Biochemicals Inc. (Irvine, CA). Stock solutions (5 mg/mL) of 8-MOP were made in DMSO. Complementary strands of 23 bp of a T7 RNAP promoter sequence were commercially synthesized by automated solid-phase procedures and further purified by HPLC as previously described (Sastry et al., 1992; Spielmann et al., 1992). The concentrations of DNAs were calculated from their respective molar extinction coefficients at 260 nm ($\sim 10^4$ M^{-1} cm^{-1} per nt). T7 RNAP was prepared locally according to published procedures (Goldberg & Dunn, 1988). The concentration of the purified enzyme was determined using a molar extinction coefficient of $\epsilon_{280} = 1.4 \pm 0.1 \times 10^5$.

Gel Mobility-Shift Assay for DNA Binding. DNAs (0.5 pmol) were labeled at their 5′ ends with ^{32}P and added to transcription buffer at room temperature. Transcription buffer consisted of 50 mM Tris-HCl, pH 8.0, + 10 mM $MgCl_2$ + 1 mM DTT + 1 mM spermidine + 5% (v/v) glycerol. When needed, 8-MOP was added to 36 μg/mL (final concentration), and the DNA was incubated at 37 °C for 10 min before polymerase addition. T7 RNAP was added at various concentrations to each reaction separately (see legend for Figure 2). The final volume of the reaction was 50 μL. The binding reactions were incubated at 37 °C for 10 min, and then glycerol was added to 5% (final) without loading dyes. The reactions were run on an 8% acrylamide−TBE nondenaturing gel (14 cm × 16 cm). Before the samples were applied, the gel was prerun and the buffer was replaced with fresh buffer. The samples were electrophoresed at 5 V/cm until the bromophenol blue dye (in separate lanes) was ~2 cm from the bottom of the gel. The gels were soaked for 20 min in 5% acetic acid + 5% MeOH + 3% glycerol, dried, and autoradiographed. Fractional binding was determined as follows: $[BL_b/(BL_b + BL_f)]$ deducted by $[C/(C + C_f)]$, where BL_b is the integrated band area of the storage phosphor signal representing the bound DNA, BL_f is the band area representing the free DNA in the same lane, C is the band area in the control (no T7 RNAP) and C_f is the band area corresponding to the free DNA in the same lane. All band areas were obtained by deducting nonspecific background storage phosphor counts in a portion of the gel where no bands were present. All quantitations were done with the aid of the ImageQuant program using a phosphorimager.

Transcription Assay. For G-ladder synthesis, 0.2 μM 23 bp template was mixed with 50 μM GTP containing 15 pmol of [α-^{32}P]GTP (specific radioactivity 3000 Ci/mmol; New England Nuclear) in transcription buffer (see above). T7 RNAP was then added to 0.2 μM, and the reaction mixture (50 μL) was incubated at 37 °C. Two and a half microliter aliquots of the reaction mixture were withdrawn at different time points and mixed with 5 μL of 8 M urea−TB−10 mM EDTA and heated in a boiling water bath for 5 min and run on a 24% acrylamide−TBE−8 M urea gel. The transcripts were visualized by autoradiography, and the percent ^{32}P incorporation was measured directly by counting the ^{32}P in gel strips containing the G-ladder in a scintillation counter.

Small-Scale Photo-Cross-linking and Quantitation. The nontemplate strand of the promoter was 5′ end labeled with the aid of [γ-^{32}P]ATP (6000 Ci/mmol) and T4 polynucleotide kinase (Maniatis et al., 1982). The kinase was inactivated by heating at 95 °C for 10 min. Unincorporated [γ-^{32}P]ATP was removed by passing the solution through a 1 mL Sephadex G25 spin column. The complementary strands (~1 μM) of the promoter were annealed by first mixing equimolar amounts of the two strands plus 8-MOP (5 μg/mL) in transcription buffer. The mixture was then heated at 70 °C for 5 min and cooled slowly to room temperature over a period of 3 h. Approximately 10 pmol of double-stranded 23-mer promoter DNA (with the ^{32}P-labeled nontemplate strand) containing 8-MOP was incubated with 20 pmol of T7 RNAP in 50 μL of transcription buffer (see above) at room temperature (~25 °C) for 30 min. Irradiations were carried out with a 200 W Hg−Xe arc lamp (Oriel Corp., Stratford, CT). The lamp housing was fitted with a water filter to eliminate IR and with a mercury line band-pass filter centered at 365 nm (~25% of light transmitted; 10 nm bandwidth; Oriel Corp. Model 56531). A 3.8 cm fused silica lens was used to focus the UV light beam on to the slit of a 100 μL quartz cuvette (1 cm path length)

containing the reaction mixture. The cuvette was placed in a custom-made cuvette holder that was thermostated at 37 °C using a pumping Lauda bath containing a mixture of 50% ethylene glycol and 50% water. The UVA light passing through the cuvette was measured using an International light meter (Model IL1700) fitted with a calibrated UVA probe (Model SFD038). Five microliter aliquots of the photoreactions were removed at specified intervals of time (see Figure 1A) and mixed with an equal volume of SDS (1%) buffer and heated at 95 °C for 5 min. The samples were run on a 10% polyacrylamide−SDS gel. For small-scale cross-linking experiments, the 365 nm band-pass filter was used instead of the dichroic mirror (see below) because the intensity of UV light passing through the sample was low enough to be accurately measured with our UVA probe and the rate of cross-linking was slow enough to be quantifiable.

To estimate the cross-link yield in terms of the amounts of T7 RNAP, we translated the sum of the total storage phosphor counts of all cross-linked bands to the amount of ^{32}P using a standard curve. We constructed the standard curve using graded amounts of [γ-^{32}P]ATP of known specific radioactivity. Many rows of [γ-^{32}P]ATP were spotted on a 3MM filter paper, dried, and exposed to a phosphor screen. The same samples were also counted for ^{32}P in a scintillation counter. The ^{32}P Cerenkov counts of these samples were plotted against the corresponding storage phosphor counts. The relationship between ^{32}P Cerenkov counts (or the actual amount in picomoles of [^{32}P]ATP) and the corresponding storage phosphor counts was perfectly linear over a large dynamic range of radioactivity. Using the standard curve and the observed storage phosphor counts, we obtained the ^{32}P counts for each cross-linked and free DNA band (e.g., Figure 1A). From the specific radioactivity of the ^{32}P label we calculated the amount of DNA in the cross-linked bands. At concentrations of T7 RNAP equivalent to the estimated macroscopic binding constant (see Figure 2B) about half of the input DNA is bound to the polymerase. The amount of polymerase used here was about equivalent to the estimated macroscopic binding constant. Assuming that one polymerase molecule cross-linked per promoter, we calculated the amount of cross-linked T7 RNAP. The percent T7 RNAP cross-link yield was obtained by dividing the picomoles of cross-linked polymerase with the picomoles of the total amount of polymerase in the reaction mixture and multiplying the result by 100.

Large-Scale Photo-Cross-linking. Promoter DNA was annealed by heat-cooling as described above. The DNA was first incubated with 8-MOP (36 μg/mL) for 10 min at room temperature in the dark. The polymerase was then added, and the tubes were incubated at 37 °C for 30 min before irradiation. Irradiations were carried out with the Hg−Xe arc lamp described above except that a dichroic mirror with greater than 90% reflectance in the 350−450 nm range (Oriel Corp. Model 66218) replaced the 365 nm band-pass filter. We prepared two batches (one mL each) of a reaction mixture containing 7 nmol of T7 RNAP mixed with 7 nmol of 23 bp promoter DNA. Typically, we included ~100 pmol of ^{32}P-labeled DNA top strand to serve as a radiotracer during purification of cross-links. Cross-linking was done at 37 °C for 5 min in a thermostated quartz cuvette (1 mL; 1 cm path length) in three ~0.3 mL aliquots for each batch of 1 mL. After irradiation, the reaction mixture was extracted with a mixture of phenol−chloroform−isoamyl alcohol (Maniatis et al., 1982). The interphase containing the protein and conjugates was precipitated with EtOH. Precipitation with 10% TCA was also effective. The cross-linked protein was resuspended in 3.5 mL of 2 M urea + 100 mM Tris-HCl (pH 8.0). A freshly prepared solution of TPCK-treated trypsin (Worthington Biochemical Corp., Freehold, NJ) (Sastry et al., 1993b) was added to T7 RNAP in three aliquots at 1 h intervals to achieve a final T7 RNAP:trypsin mass ratio of 6:1. The T7 RNAP cross-links were digested for about 4 h at 37 °C. The digest was then lyophilized down to approximately 1 mL and injected into an HPLC anion-exchange column (Vydac oligonucleotide column, catalogue no. 3040L54). The HPLC method was as follows. Buffer A: 0.01 M sodium phosphate (pH 6.7) + 20% CH_3CN. Buffer B: 0.4 M sodium phosphate (pH 6.7) + 20% CH_3-CN. The gradient was 100% buffer A in 0−10 min and 0−100% buffer B in 11−60 min. One milliliter fractions were collected at a flow rate of 1 mL/min, aliquots of the fractions were run on 10% acrylamide−Tris−Tricine SDS gels, and the cross-links were identified by autoradiography (Figure 5). Cross-linked peptides were pooled and desalted using Waters Sep-Pak C18 reversed-phase cartridges (procedure and columns were from Millipore Corp., Milford, MA). The cross-links were photoreversed with 254 nm UV light in a Rayonet photochemical reactor (Southern New England UV Co., Branford, CT) to detach the DNA from the peptides (Sastry et al., 1993b). The yield of peptides was approximately 25−50 μg as estimated by the bicinchoninic acid protein microassay method (Pierce Chemical Co., Rockford, IL). Portions of the cross-linked peptide samples were sent to protein structure facilities located at the University of California at Davis, to the W. M. Keck Foundation facility at Yale University, and to Prof. James Manning's laboratory at Rockefeller University, for automated Edman sequencing or amino acid composition analysis. The analyses were carried out under a monetary contractual agreement. The data were analyzed in our laboratory. Other portions of the samples were subjected to LC/MS. The LC phase of the analysis was carried out on a 10 cm C18 analytical reversed-phase column using a TFA−acetonitrile gradient. The elution conditions were as follows: buffer A, 0.1% TFA + 10% CH_3CN; buffer B, 0.1% TFA + 70% CH_3CN; linear gradient from 0 to 70% buffer B in 20 min.

Fluorescence Spectroscopy. A 0.16 mM solution of 8-MOP in 10 mM Tris-HCl, pH 7.5, + 1 mM EDTA was split in half. Half was irradiated with UVA under conditions identical to those of in the large-scale cross-linking reaction (see above). The other half was kept in the dark. The 8-MOP samples or the isolated peptide−cross-link samples (500 μL) were loaded in a quartz cuvette, stirred in a Hitachi F-2000 spectrofluorometer, and equilibrated at room temperature for at least 5 min. The excitation and emission scans were collected with a bandwidth of 5 nm at a scan speed of 15 nm/min. The excitation and emission wavelengths are given in Figure 6. The fluorescence data were acquired using a PC that was interfaced with the spectrofluorometer. The data were signal-averaged and digitized.

RESULTS AND DISCUSSION

Cross-Linking of T7 RNAP to DNA with a Psoralen. Figure 1A shows a time−course for cross-linking T7 RNAP to a 23 bp DNA containing a slightly modified T7 promoter

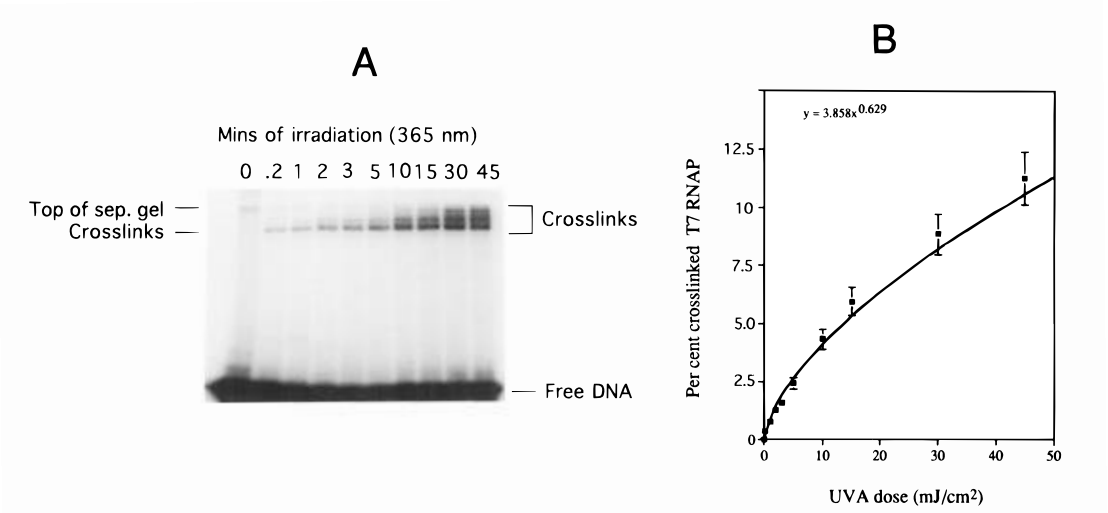

FIGURE 1: Cross-linking of T7 RNAP to a 23-mer DNA containing T7 promoter sequence. Panel A: Autoradiogram of a 10% acrylamide−SDS gel. The cross-links migrate approximately close to a ~100 kDa marker. Identical samples were irradiated for the indicated length of time (except 0, which was not irradiated). Panel B: Quantitation of photo-cross-linking. Autoradiograms, similar to those in panel A were used for the quantitation (see Materials and Methods). The total dose of UV light that each sample received was varied by changing the length of irradiation time. Each point is an average of three separate experiments.

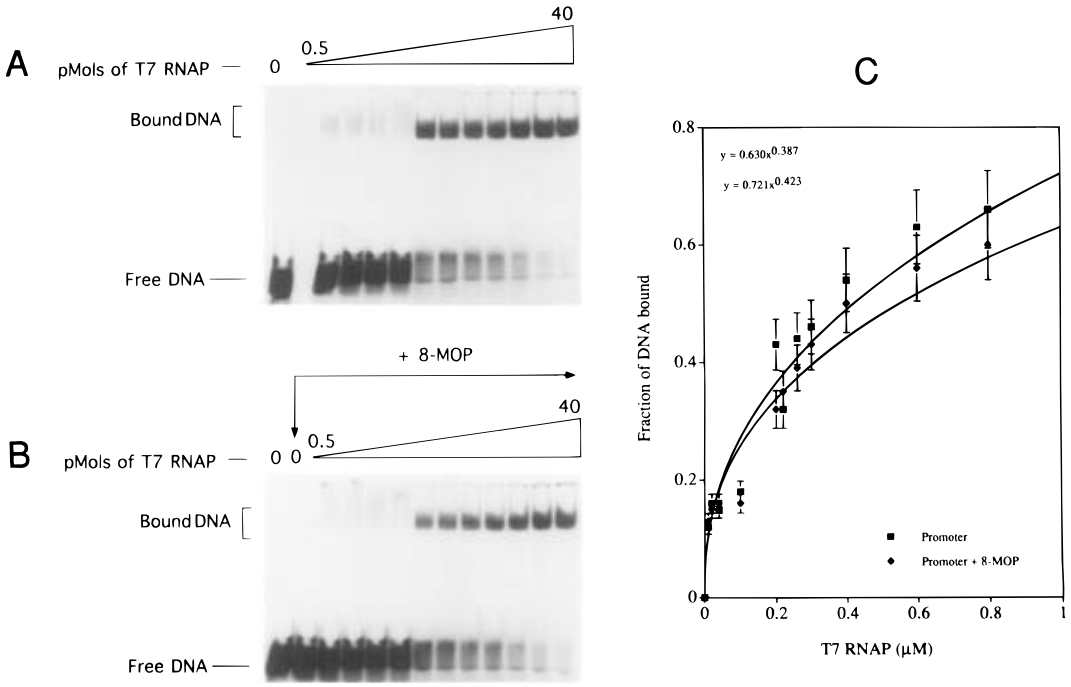

FIGURE 2: Gel-shift assay for promoter−T7 RNAP interactions in the presence and absence of 8-MOP. Panel A: Binding of T7 RNAP to promoter without 8-MOP. Panel B: Binding of T7 RNAP to promoter with 8-MOP. Panel C: Calculated binding isotherms. Squares represent promoter without 8-MOP, and diamonds represent promoter plus 8-MOP. Each point in the graph represents an average of three independent determinations.

sequence (5′-TAATACGACTCACTATAGGGAAG-3′). The cross-linking was carried out using a Hg−Xe arc source fitted with a 365 nm band-pass filter. The cross-links migrated approximately near a ~100 K marker, consistent with the size of T7 RNAP (e.g., see Figure 4). The amount of cross-linking increases with irradiation time. Short irradiation times (up to 5 min) yielded mostly a single band representing T7 RNAP−DNA cross-links. Upon prolonged UVA exposure (>10 min) more than one band of cross-link was seen. We believe that multiple bands represent alternate conformers of cross-links within the gel. They may arise owing to cross-linking at multiple sites and/or partial or complete denaturing states of the cross-links. This interpretation is consistent with our present results (see below). The total amount of cross-linked T7 RNAP increases as a function of UVA dose, indicating that cross-linking is completely light-dependent (Figure 2B). No cross-links were seen without psoralen. Because only the psoralen is photoreactive at 365 nm (DNA and protein do not absorb UV at this wavelength), it is clear that covalent bonds are formed between psoralen, DNA, and T7 RNAP [see also Sastry et al. (1993b)]. Experiments with nonpromoter DNA (at the same T7 RNAP:DNA ratio as with promoter DNA) showed about 100-fold lower yields of cross-links compared with promoter (not shown). The lower cross-link yields with nonpromoter DNA are consistent with a decreased (10^2−10^3-fold) binding affinity of T7 RNAP to nonpromoter DNA. In a previous study, we showed that non-DNA-binding proteins, such as BSA, were not cross-linked to the DNA via psoralen.

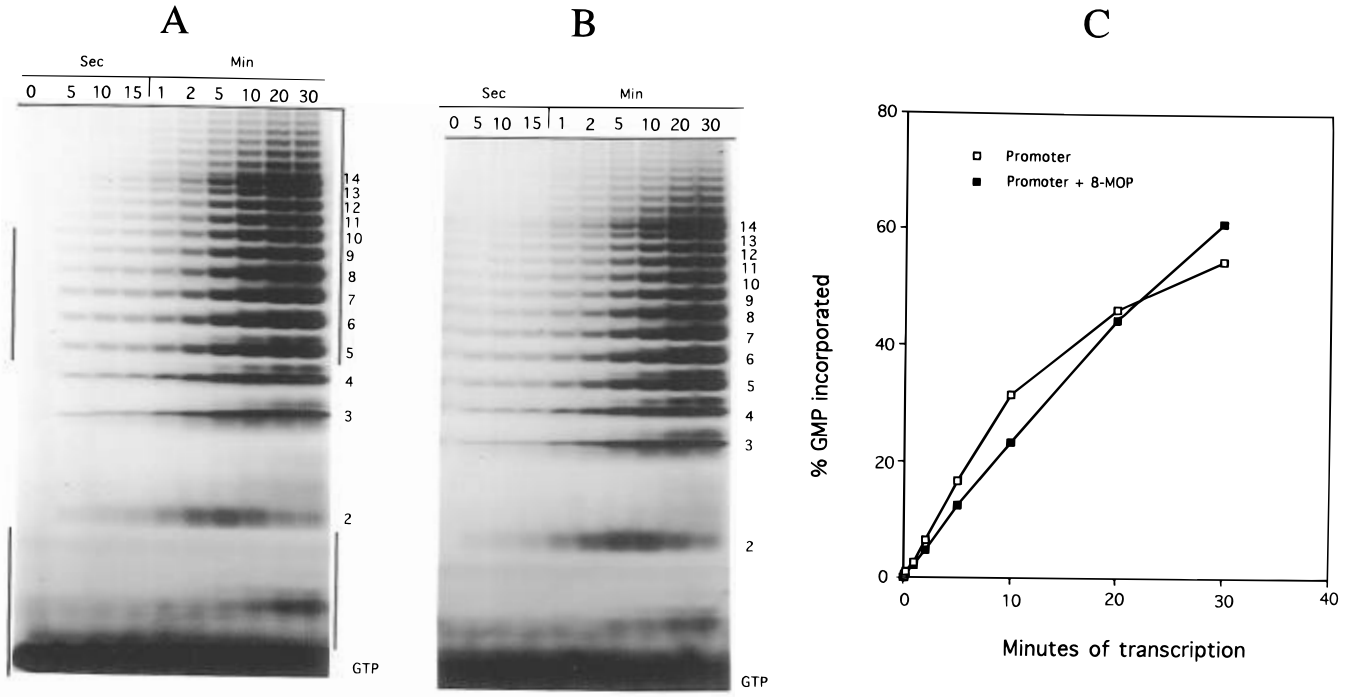

FIGURE 3: Transcription initiation is normal in the presence of 8-MOP. Panels A and B show the time−course of poly(G) ladder synthesis without and with 8-MOP, respectively. Panels A and B are autoradiograms of 24% acrylamide gels. The numbers at the top of the gels indicate time points at which samples were removed from a transcription reaction and run on gels. Zero is the start of transcription. The numbers on the right indicate the number of G residues incorporated into the ladder. Panel C shows quantitation of typical incorporation rates.

Our estimated psoralen-mediated photo-cross-link yields with T7 RNAP (Figure 2B) are comparable to those mediated by 254 nm UV or aryl azides or halogenated nucleotides. The reported yields obtained with some other cross-linking methods are in the order of 4−25% cross-linked protein (Williams & Konigsberg, 1991; Budowsky & Abdurashidova, 1989).

T7 RNAP Appears to Bind Promoter Normally in the Presence of 8-MOP. Because our cross-linking technique involves irradiation of a mixture of promoter DNA, T7 RNAP, and 8-MOP, it was important to find out if T7 RNAP formed complexes with promoter in the presence of the psoralen. To monitor T7 RNAP binding, we titrated a fixed concentration of radiolabeled promoter DNA with increasing amounts of T7 RNAP (Figure 2). 8-MOP was added to promoter DNA and incubated for 10 min before addition of T7 RNAP (Figure 2B). Gel-shift assays (Figure 2A,B) indicated that T7 RNAP binds to the promoter DNA in the presence or absence of 8-MOP (36 μg/mL) with a similar affinity (Figure 4C). A macroscopic binding constant was derived by fitting the data to a general power equation ($y = bx^m$; Figure 4C). The slope ($m = dy/dx$) of the linear-transformed equation was equivalent to the binding constant. The derived macroscopic binding constant in the presence or absence of psoralen was the same within experimental error ($K_a \sim 3-4 \times 10^7\ M^{-1}$). This experiment demonstrated that T7 RNAP was bound to promoter probably in a normal fashion in the presence of 8-MOP. Binding experiments also indicated that nonpromoter DNA was unable to compete with the specific bands formed by T7 RNAP and the 23 bp promoter DNA, a result consistent with a previous report (Muller et al., 1988).

Transcription Initiation Is Also Normal in the Presence of Psoralen. We used a poly(G) ladder synthesis assay to measure transcription initiation (Gross et al., 1992; Ling et al., 1989; Martin et al., 1988; Sousa et al., 1992). In the presence of GTP as the sole nucleotide, T7 RNAP synthesizes a poly(G) ladder that extends to ~14 nts and abruptly tapers off above 14 nts (Figure 3A,B). The poly(G) ladder is the result of the reiterative slippage synthesis at the three C's on the template strand (+1 to +3; (Martin et al., 1988)). Figure 3A,B shows that, even at the saturating concentrations of 8-MOP (36 μg/mL; 23 psoralens/23 bp DNA) used in our cross-linking reactions, transcription initiation is not inhibited. Quantitation of the rate of [^{32}P]GMP incorporation confirmed that there was no inhibition of transcription initiation in the presence of the psoralen (Figure 3C). These experiments demonstrate that, in the presence of 8-MOP *before* irradiation and photobinding, promoter-specific complexes are formed.

Isolation of Cross-Links. The 23 bp promoter ds DNA was cross-linked to T7 RNAP through added 8-MOP in a large-scale reaction (see Materials and Methods). We added saturating amounts of free 8-MOP (36 μg/mL; 23 psoralens/23 bp DNA) to promoter DNA. Because intercalators are generally thought to obey "the nearest neighbor exclusion principle" (Saenger, 1984), psoralens might be intercalated between every second DNA base pair. Assuming a random mode of intercalation, at equilibrium, there should be at least two sets of DNA molecules in the population that have psoralen intercalated between alternating base pairs (other combinations may also occur). When irradiated, psoralen photochemically reacts with DNA and T7 RNAP and forms DNA−T7 RNAP cross-links. Cross-links are formed by the photoreaction of DNA-intercalated/stacked psoralen and bound T7 RNAP (Figure 2B). Since psoralen is a bifunctional reagent during protein−DNA cross-linking, photocoupling occurs by the absorption of two photons, most probably through two separate excitation events. The first absorbed photon probably leads to monoadducts with DNA.

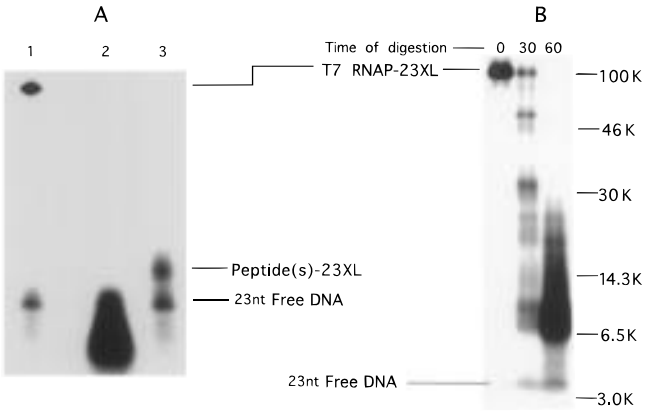

FIGURE 4: Precipitation and proteolysis of cross-linked T7 RNAP. Panel A: Autoradiogram of a 10% acrylamide−Tris−Tricine gel showing nonspecific cleavage of cross-linked T7 RNAP with proteinase K. Cross-linked T7 RNAP was precipitated with 10% TCA. The precipitated cross-links (10 μg of protein) were dissolved in 25 μL of TBE + 100 mM triethylammonium acetate (pH 7.0) + 100 mM Tris-HCl, pH 7.0, + 1% SDS. One microgram of proteinase K (in 50 mM Tris-HCl, pH 8, + 1 mM $CaCl_2$) was added, and the reaction was incubated at 37 °C for 30 min. Lane 1 contained a sample of TCA-precipitated cross-linked T7 RNAP before treatment with proteinase K. Lane 2 shows an aliquot of the supernatant containing free DNA following TCA precipitaion of the cross-links. Lane 3 contained TCA-precipitated, cross-linked T7 RNAP that was cleaved with proteinase K. Panel B: Time−course of trypsin digestion of cross-linked T7 RNAP. The T7 RNAP−DNA conjugates were precipitated twice with 10% TCA and resuspended in 3 M urea + TBE + 100 mM triethylammonium acetate, pH 7.0, + 100 mM Tris-HCl, pH 7.0, in a volume of 100 μL. TPCK-treated trypsin was freshly prepared in 10 mM HCl + 20 mM $CaCl_2$ and added to cross-links at a mass ratio of 1:6 trypsin to cross-links. Thirty microliter samples were removed at indicated intervals of time from the reaction mixture, heated in a boiling water bath for 10 min, and then subjected to 10% acrylamide−Tris−Tricine SDS−PAGE. This figure is intended to be a qualitative representation of the distribution of the products of protease digestion. Lane 60 was overloaded in order to get a better idea of the upper limit in the distribution of the digestion products.

Absorption of a second photon only by furan-side DNA monoadducts leads to cross-linking with polymerase. We have estimated the quantum yield (defined as moles of cross-link formed per mole of absorbed photons) for cross-linking single-stranded DNA furan-side monoadducts to DNA-binding protein. This was in the range of ∼0.002−0.01, depending on the wavelength of UVA.[2] These quantum yields are smaller than those reported for the conversion of furan-side monoadducts to interstrand DNA cross-links (0.02−0.04) (Shi & Hearst, 1987; Tessman et al., 1885). In our photoreactions, T7 RNAP that first reacted with free psoralen may not cross-link with DNA because (1) a specific stereochemical intercalation geometry is required for a [2+2] photo-cycloaddition to DNA which is less likely to occur with a psoralen that was first photochemically attached to T7 RNAP, (2) photomodified psoralen that is first attached to protein via the pyrone side (3,4 double bond or C2═O) will not absorb a long-wave UVA photon to undergo a [2+2] reaction with DNA, and (3) with 365 nm or longer wave UVA, photoreaction of free psoralen with T7 RNAP inactivates transcription and DNA binding by T7 RNAP.[3] Thus T7 RNAP that is first photoconjugated to 8-MOP is unlikely to bind and cross-link to DNA.

[2] Unpublished experiments.
[3] Unpublished experiments.

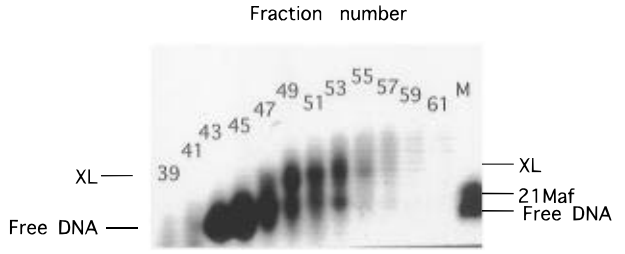

FIGURE 5: HPLC profile of trypsin-cleaved T7 RNAP−^{32}P promoter DNA cross-links. Ten microliters of fractions from anion-exchange HPLC was analyzed on 10% Tris−Tricine SDS−PAGE. Aliquots from every alternate fraction were loaded on the gel. Irradiated markers are 23-mer and 21 Maf (M).

The first step in our large-scale cross-link purification was precipitation of the cross-linked T7 RNAP with TCA (or EtOH plus NaOAc) to remove non-cross-linked DNA. Figure 4 (lane 1, Figure 4A, and lane 0, Figure 4B) shows an example of the precipitated, cross-linked T7 RNAP that is essentially free of non-cross-linked DNA. The free DNA is in the supernatant (lane 2, Figure 4A). After treatment with TCA the majority of free DNA migrated slightly faster, possibly due to some hydrolysis in the presence of TCA (lane 2, Figure 4A). T7 RNAP cross-linked to ^{32}P-labeled DNA is in one band migrating at ∼100 K (lane 0, Figure 4B). Digestion with proteinase K resulted in a smear of bands migrating slightly above the free DNA (lane 3, Figure 4A). The smear above the free DNA band presumably contained very short oligopeptides that remained cross-linked to DNA (lane 3, Figure 4A). With trypsin, a much wider distribution of peptides cross-linked to DNA was seen after 30 min (lane 30) or 1 h (lane 60). Tryptic digestion for 1 h produced mostly cross-linked peptides in the range of 6.5 to ∼15 kDa. Further digestion resulted in a distribution of smaller products. These experiments indicate that DNA in the cross-links is not dissociated during precipitation or proteolysis by both specific (trypsin) and nonspecific (proteinase K) proteases and that the DNA in peptide conjugates may be cross-linked to several different sites. A 4 h tryptic digest was subjected to anion-exchange HPLC using a quaternary amine column (Figure 5). This column selectively binds DNA very strongly. Control runs of the tryptic digest of T7 RNAP showed no retention of peptides on this column. Peptides without DNA eluted in the void volume, whereas DNA and DNA cross-linked peptides are eluted in later fractions (Figure 5). Fractions 49−55 were pooled, concentrated, and desalted on C18 reversed-phase cartridges. The UV absorption spectra of the isolated peptide mixture indicated the presence of DNA (not shown).

Spectral Characteristics of Cross-Linked Peptides. To prove the presence of psoralen in isolated cross-links and to gain important insights about the cross-linking sites in the psoralen, we obtained UV absorbance and fluorescence spectra of isolated cross-linked peptides and compared their spectral characteristics with 8-MOP and irradiated 8-MOP. Figure 6 shows representative excitation and emission spectra. The respective UV absorption spectra (not shown) are roughly superimposable on the excitation spectra. The excitation spectrum of 8-MOP shows two absorption bands centered ∼250 and ∼300 nm (Figure 6A). Upon excitation at 330 nm, 8-MOP showed a strong emission band centered at ∼500 nm (Figure 6A). These spectra are in agreement with previous reports (Lerman et al., 1980; Yoshikawa et

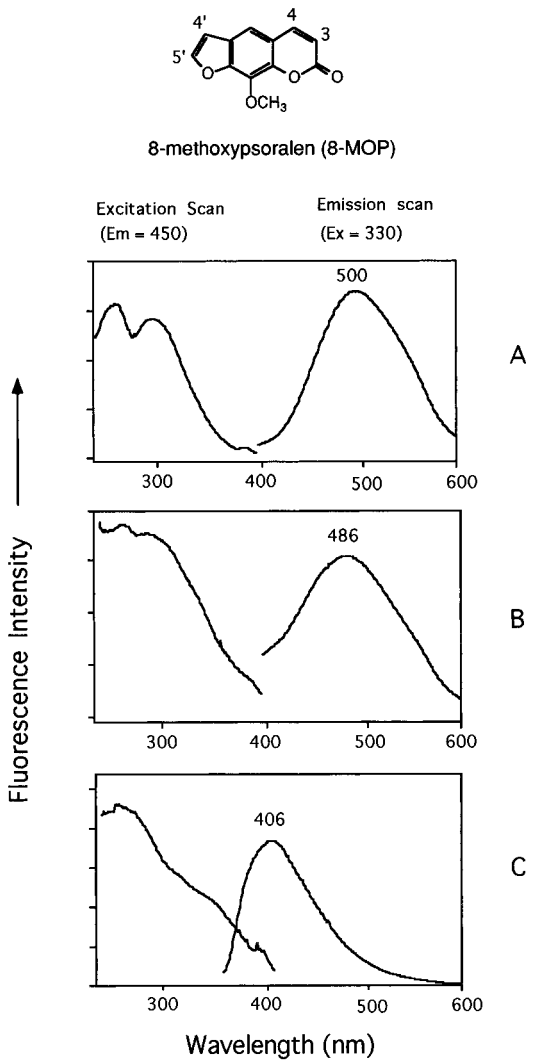

FIGURE 6: Spectral characteristics of cross-linked peptides following isolation and purification. All spectra were taken in 10 mM Tris-HCl, pH 7.5, + 1 mM EDTA. The chemical structure of 8-MOP is shown at the top. Panel A: Unirradiated 8-MOP. Panel B: Irradiated 8-MOP. Panel C: T7 RNAP−DNA cross-linked tryptic peptides.

al., 1979). Irradiation of 8-MOP with long-wave UV resulted in a pronounced decrease in the main absorption bands, particularly at ~300 nm that corresponds to the lowest energy excited state (Figure 6B). The emission maximum is blue shifted by ~14 nm relative to unirradiated 8-MOP (Figure 6A), indicating saturation of the 4′,5′ double bond of the psoralen (Blias et al., 1984; Moyson et al., 1993). Here, dimerization of 8-MOP via cycloaddition may have occurred, as previously noted (Krauch & Farid, 1967). Only the 4′,5′ adducts of psoralen are strongly fluorescent (Musajo et al., 1966). In cross-linked peptides with DNA, we observed that the modified psoralen was strongly fluorescent (Figure 6C). However, the emission peak was strongly blue shifted by ~94 nm compared with unirradiated 8-MOP or by 80 nm compared with irradiated 8-MOP. The absorption peak at ~300 nm (lowest energy transition) is strongly suppressed, and there is a small shoulder at ~350 nm (Figure 6C). These excitation and emission characteristics indicate that the furan side of the psoralen is saturated by bonding to DNA (as expected) and that the 3,4 double bond does not participate in cross-linking with peptide (otherwise the product would not be fluorescent). The presence of absorption bands in the 290−310 nm range in the excitation spectrum of purified cross-linked peptides (Figure 6C) indicated that the coumarinic chromophore was largely intact. Between 250−400 nm psoralens have two transitions, an n,π^* localized in the C2═O and a π,π^* localized in the ring system (Ben-Hur & Song, 1984; Mantulin & Song, 1973; Song & Tapley, 1979). Since the coumarin (3,4 double bond) system appears to be intact in the isolated cross-links, we suggest that in our photoreactions C═O may be the site of excitation for reactivity with protein. The suggested overall cross-linking pathway is

$$\text{psoralen} \xrightarrow{h\nu} 4',5' \text{ adduct with DNA} \xrightarrow{h\nu} \text{adduct with protein probably via C═O}$$

In this scheme only the furan-side monoadduct is reactive with protein, consistent with our previous results (Sastry et al., 1993). The emission peak of our purified cross-links (Figure 6C) is not identical with a purely furan-side DNA monoadduct (Moyson et al., 1993) perhaps because of electron (charge) transfer or other interactions due to attachment with peptide. The C═O in the psoralen is the site of direct excitation (either singlet or triplet n,π^* transitions) by near-UV (Ben-Hur & Song, 1984; Mantulin & Song, 1973; Song & Tapley, 1979). The reaction via C2═O with protein may involve H-abstraction or electron transfer (Turro, 1991), quite unlike the cyclobutane adducts with DNA.

Identification of the Cross-Linked Peptides. In order to facilitate the measurement of the exact masses of peptides with psoralen, the cross-links were photoreversed (254 nm UV-mediated cycloelimination) to detach the DNA from the peptides (Sastry et al., 1993b). Portions of the cross-linked peptide samples were subjected amino acid composition analysis and Edman sequencing. We could not obtain primary sequence information from Edman sequencing. Other peptide samples containing conjugated psoralen(s) were subjected to liquid chromatography/mass spectrometry (LC/MS). Figures 7 and 8 show electrospray ionization mass spectra of the isolated peptides. The hyper mass deconvolution method (Figures 7 and 8) allows the determination of the mass of peptides that generated multiply charged species. Signals (e.g., peak 489 in Figure 7) that did not generate multiply charged ion species were not considered. No attempts were made to match (see below) these singly charged species because their intensity was weak and mass determination was less reliable. Thus all the masses given in the tables of Figures 7 and 8 were the result of multiple determinations. These peptides represent the *maximum* number of peptides we have been able to identify with certainty (see Table 1).

To identify the peptides and the number of 8-MOP units per peptide, we assumed a conservation of mass of the 8-MOP and peptides after the photoreaction. Although in this work we have not presented chemical evidence supporting our assumption (besides spectroscopic data, Figure 6C), if H-abstraction or electron transfer mechanisms are involved as suggested above based on the spectral characteristics of the products (see Figure 6), one may expect photoproducts that are essentially conserved in mass, based on published mechanisms (Dorman & Prestwich, 1994; Turro, 1991). Since we were unable to obtain unambiguous amino acid sequence data for our peptide cross-links despite

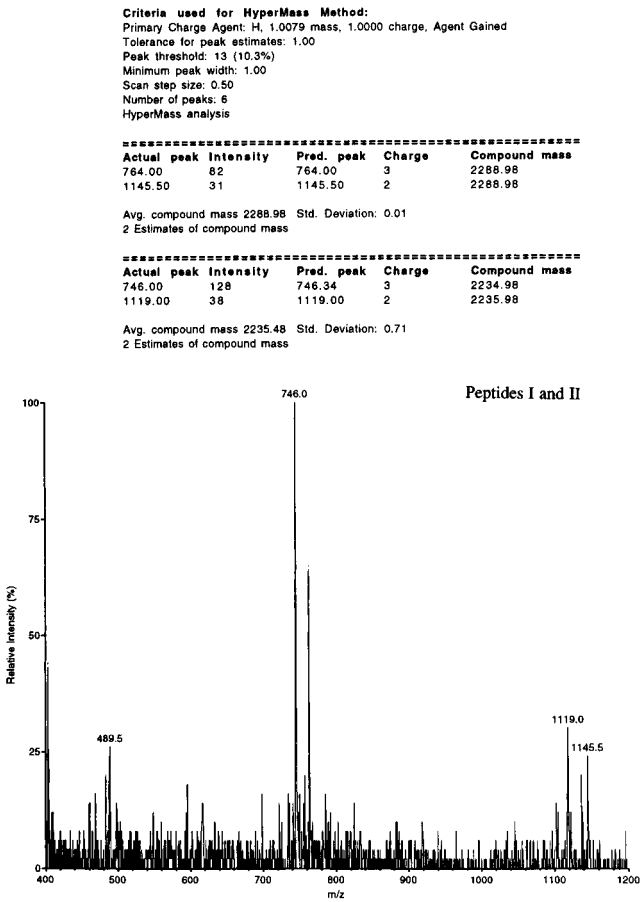

FIGURE 7: Electrospray ionization mass spectra of isolated T7 RNAP peptides attached to psoralen. The method of computing the peptide masses is shown at the top. The peptides were first passed through a reversed-phase HPLC column and subjected to mass spectrometry in the positive ion mode (Peptidogenic Inc., Livermore, CA). Peak 764 (3^+) is on the right-hand side of 746.

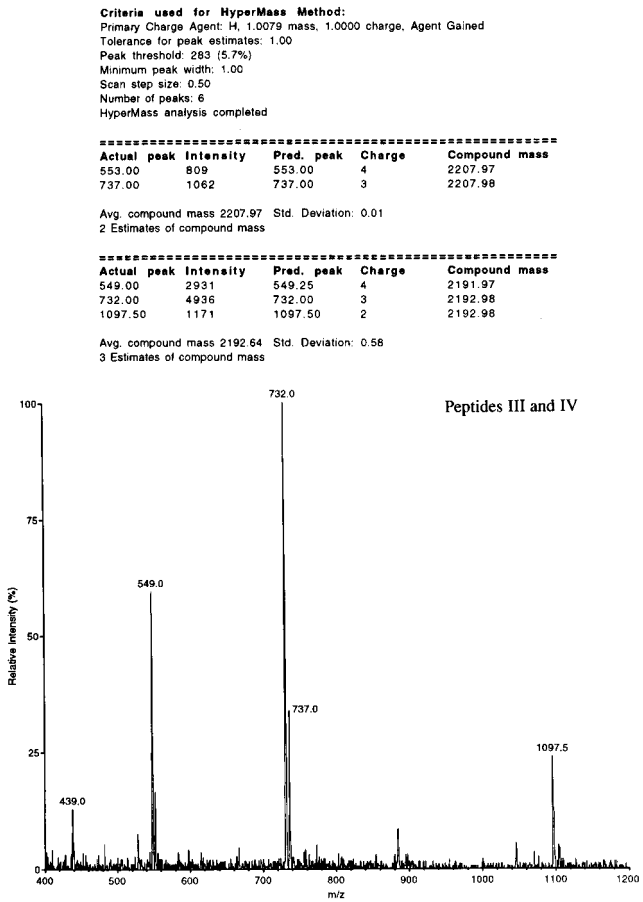

FIGURE 8: Electrospray ionization mass spectra of isolated T7 RNAP peptides attached to psoralen. The method of computing the peptide masses is shown at the top. The peptides were first passed through a reversed-phase HPLC column and subjected to mass spectrometry in the positive ion mode (Peptidogenic Inc., Livermore, CA). Peak 439 is 5^+, and peak 553 (4^+) is on the right side of 549 and belongs to peptide IV.

repeated attempts at Edman sequencing, we speculate that the photochemistry may involve peptide backbone modifications, in analogy to benzophenones (Dorman & Prestwich, 1994).

We generated a computer-assisted map of the partial and complete trypsin digest of T7 RNAP using software developed by the mass spectrometry laboratory at Rockefeller University and the PEP program of Intelligenetics Corp. (Mountainview, CA). T7 RNAP has 66 lys and 41 arg (Dunn & Studier, 1983; Moffatt et al., 1984). Cutting by trypsin on the C-terminal side of every Lys and Arg generated a partial−complete T7 RNAP digest consisting of 5886 peptides. Our mass spectroscopy data are accurate to 1 or 2 amu with a standard deviation of 0.01−0.7 amu. Therefore, in the computer program, we set a mass window for matching peptides at either 1 or 2 amu while searching for hits. With this computer program, given a mass, either there is a match or there is none. We then attempted to match each of the peptides after subtracting an integral number of psoralens. We searched the peptide bank for matches after subtracting 1 or 2 or 3, etc., of 8-MOP mass units and identified the peptides (Table 1). To obtain a match within 1 or 2 amu, no more or no less than the numbers of psoralens in Table 1 were subtracted from the observed peptide masses. There were partial peptides in the bank that came closest, within 5−138 amu, to the expected masses in Table 1. The positions of amino acids of these peptides are as follows. The calculated molecular weights are in parentheses: 85−95 (1379); 85−96 (1535); 85−98 (1720); 85−99 (1876); 144−153 (1163); 144−155 (1433); 154−163 (1256); 154−164 (1384); 161−172 (1514); 164−172 (1101); 164−173 (1258); 165−173 (1132); 165−172 (973); 165−179 (1893); 181−193 (1469); 194−206 (1428). None of these peptides matched the expected molecular masses given in Table 1. Other criteria were used to exclude all the above listed peptides from further consideration. (1) The difference between the calculated mass and the expected mass is exceedingly large, if we assume a conservation of mass after photochemistry (see below). The margin of difference should be within 2 amu on the basis of the accuracy of our mass determinations. (2) Severe tryptic digestion conditions were used. On the basis of the fact that very short peptides were isolated, we believe that tryptic digestion was mostly complete; therefore, partial peptides are less likely. Four other partial peptides existed in the data bank that were within a range of 3 and 35 amu of the observed masses given in Table 1 (i.e., excluding 8-MOP mass). These peptides spanned amino acids 121−143 (2265), 35−52 (2196), 32−50 (2232), and 81−98 (2192). These four peptides were considered not relevant because the UV absorption and fluorescence excitation and emission spectra of the isolated peptide mixture indicated the presence of psoralen (see

Table 1: T7 RNAP Peptides Cross-Linked to 8-MOP[a]

peptide no.	obsd mass	expected mass	amino acid sequence	no. of 8-MOP (8-MOP = 216.18)
I	2288 ± 1	1423	396-ISLEFMLEQANK-407	4
II	2235 ± 1	1803	426-VYAVSMFNPQGNDMTK-441	2
III	2192 ± 1	1111	462-IHGANCAGVDK-472	5
IV	2208 ± 1	1127	874-DILESDFAFA-883	5

aThe peptides were identified using a computer algorithm and a data bank of all possible trypsin fragments of T7 RNAP (see Results). The observed masses were from electrospray ionization mass spectra presented in Figures 7 and 8. The mass determinations have an accuracy of 1 or 2 amu. All observed masses are from at least two determinations. Signals that did not correspond with multiple ions were not considered in this table. The expected masses were obtained by subtracting the sum of the mass of the number of 8-MOP residues from the observed masses. See Results for a description of the procedure. The number on the left of each amino acid sequence corresponds to the N-terminal amino acid residue while the number on the right is a lysine corresponding to a trypsin cleavage site. Peptide **IV** corresponds to the C-terminal end of the T7 RNAP sequence.

above) and our purification procedure is selective for DNA-conjugated peptides. Therefore, the psoralen must be present on the peptides. By following the above procedure, peptides were correctly matched with their sequences (Table 1).

To further confirm the identity of the peptides in Table 1, we used amino acid composition analysis as a secondary piece of information. The following are the amino acids and their total numbers from Table 1. The corresponding numbers in the compositional analyses are given in parentheses. A = 6 (6); C = 1 (−); D = 4 (4); E = 3 (3); F = 4 (5); G = 3 (4); H = 1 (1); I = 3 (4); K = 3 (4); L = 3 (3); M = 3 (1); N = 4 (5); P = 1 (1); Q = 2 (2); S = 3 (3); T = 1 (1); V = 3 (4); Y = 1 (2). Amino acid composition data have an accuracy of up to 80−90%. Some amino acids are difficult to quantitate because they are unstable (e.g., C) or are oxidized (e.g., M) [e.g., see Blackburn (1978) for details]. Moreover, in our case, we do not know how psoralen modifies amino acids. Hence, we should emphasize here that amino acid composition analysis is only used to support our conclusions from mass spectrometry. Our results show that most of the amino acids predicted by mass spectrometry are represented in the amino acid composition analysis. For some amino acids, the numbers are higher than the expected numbers on the basis of mass spectrometry (Table 1). This may be because amino acid composition was carried out with portions of the samples prior to mass analysis. For mass analysis, the samples were subjected to an additional phase of HPLC purification. It is possible that some peptide(s) might have been lost during this additional step of purification. Alternatively, the differences may be due to the inherently low accuracy of amino acid composition analysis of mixtures of peptides.

There are differences in the number of attached psoralens per peptide. Up to 5 psoralens per peptide with a total of 16 psoralens are cross-linked (Table 1). Independently of the mass data, we measured the UV absorbance of peptide cross-links. From the absorbance at 312 nm we estimated the concentration of 8-MOP (ϵ_{312} = 900) in the peptide mixture to be $\sim 2.4 \times 10^{-5}$ M. The molar ratio of 8-MOP to total peptide concentration ($\sim 1.2 \times 10^{-6}$ M) that was estimated by the bicinchoninic acid protein assay method was roughly 20, in approximate agreement with the predicted total number of 8-MOP molecules. The variation in the number of psoralens per peptide may reflect individual proximity differences of amino acids with regard to the DNA-intercalated psoralens. We calculated that psoralen may photoreact with an amino acid that is within ~8 Å from a monoadducted pyrimidine base. From the NMR-derived structure of the furan-side monoadduct (Spielmann et al., 1995a,b) (proteins were not present in this work), it is clear that the psoralen was stacked coplanar with the base pairs and was attached to the thymine via a C4 cyclobutane ring. This geometry may confer a restricted degree of freedom to the psoralen in the DNA, although the DNA itself is flexible (Spielmann et al., 1995a,b). Since the photoreaction contained about 24 psoralens per 23 bp of DNA, it is conceivable that a large number of psoralens are acting like a "molecular glue" by photochemically cross-linking the protein−DNA contacting interfaces. On the other hand, some of the attached psoralens may not represent cross-links to DNA but may have photoreacted with the protein after the first cross-linking event. In this case, photochemical "fixation" of T7 RNAP to DNA by the first cross-link may promote photoreactions of additional psoralens that are in close proximity. These suggested mechanisms are consistent with the known promiscuous reaction of psoralens with amino acids in proteins (Midden, 1988).

In our experiments, it is unlikely that more than one molecule of polymerase was bound per promoter. The "footprint" of T7 RNAP on the promoter covers about 20 bp (Basu & Maitra, 1986; Ikeda & Richardson, 1986; Muller et al., 1989; Shi et al., 1988b). Our template DNA is 23 bp. We think there is not enough room for a second polymerase molecule to be stably bound to the same promoter DNA. We do not know the cross-linked amino acids nor do we know which strand of the template DNA was cross-linked to the peptides. It is conceivable that some cross-links are unstable during the analytic procedure and may have been missed. It appears that many survived the purification procedures. Our repeated failure to Edman sequence psoralen-cross-linked peptides may be due to significant alterations in the chemical properties of peptides (e.g., charge density and/or hydrophobicity). At present, mass spectrometry, aided by amino acid composition analysis, is the primary analytical technique that appears to be the most practical approach for identifying psoralen-cross-linked peptides. Efforts are continuing for the application of other advanced mass spectrometry methods for identifying cross-links and to understand the chemical basis of the photoreaction of psoralen to proteins and amino acids.

Mapping the Cross-Linked Peptides on the 3-D Crystal Structure. Figures 9 and 10 show stereo pairs of α-carbon backbone representation of the crystal structure of T7 RNAP (Sousa et al., 1993). The cross-linked peptides are shown in yellow and are numbered according to Table 1. Because we did not display the 3-D structure of T7 RNAP *before*

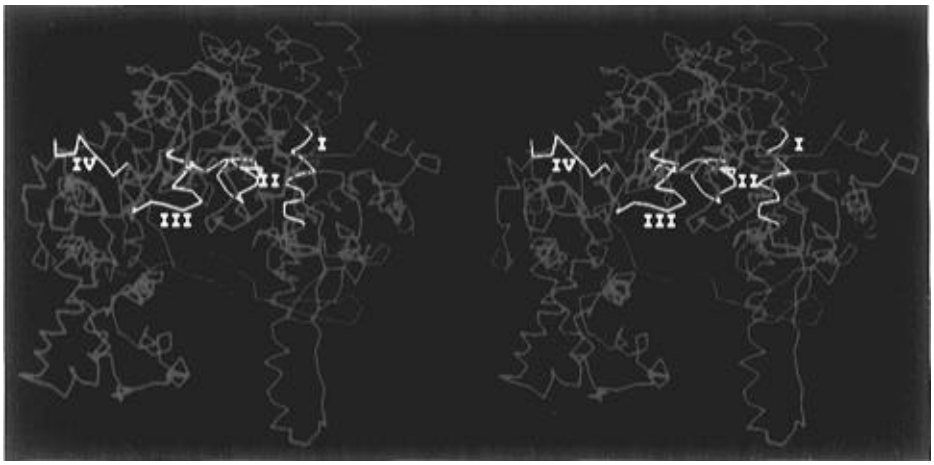

FIGURE 9: Stereoview of the crystal structure of T7 RNAP [after Sousa et al. (1993)]. The long axis of the cleft lies perpendicular to the plane of the paper. The coordinates for the T7 RNAP crystal structure were obtained from Brookhaven National Laboratories protein data bank (2RNP). Unknown amino acid coordinates were converted to alanines. The α-carbon backbone (in purple) model was computed using the Quanta program on a Silicon Graphics work station. The cross-linked peptides are shown in yellow. This figure shows the positions of the peptides from this work.

FIGURE 10: Stereoview of the crystal structure of T7 RNAP [after Sousa et al. (1993)]. The long axis of the cleft lies perpendicular to the plane of the paper. The coordinates for the T7 RNAP crystal structure were obtained from Brookhaven National Laboratories protein data bank (2RNP). Unknown amino acid coordinates were converted to alanines. The α-carbon backbone (in gray) model was computed using the Quanta program on a Silicon Graphics work station. The cross-linked peptide is shown in black. This figure shows the peptide mapped in a previous work (Sastry et al., 1993b).

identifying the peptides, there was no bias involved in identifying and mapping these peptides on the crystallographic model.

The polymerase molecule is organized around a cleft for template binding (front of the molecule). In analogy to the Klenow fragment of *E. coli* DNA pol I and HIV-1 RT, the T7 RNAP fold has a "thumb" (right-hand part of Figures 9 and 10), a "palm" (central portion of Figures 9 and 10), and "fingers" (left-hand side of Figures 9 and 10). Peptides **I**, **II**, and **III** are nicely located within the template-binding cleft, thus positively identifying the cleft as the place of template binding, confirming the crystallographic model in solution (Sousa et al., 1993). Peptide **I** is toward the upper end of the long O-helix and forms a part of the right wall of the cleft. Helix O is presumed to be a part of the thumb domain, and bending of the thumb domain may be involved in narrowing of the cleft around the DNA. Part of peptide **II** forms the ceiling of the back side of the cleft and is partly β-sheet 4 leading to helix P (Sousa et al., 1993). β-Sheet 4 is a part of the active site and has been proposed to contact the template strand in the crystallographic model. Peptide **III** forms the end of helix Q and the loop connecting helix R. As shown here, peptide **III** protrudes into the front of the cleft. Peptide **IV** is in the "leading edge" of the polymerase and is presumed to make downstream contacts with the template. Peptide **IV** forms a part of helix FF and the C-terminal end. Peptide **IV** lies in the junction of the palm and fingers but is physically not a part of the DNA-binding cleft, at least in the static crystallographic view of the polymerase. Template cross-linking to this peptide may suggest a conformational change in the protein, bringing this region close to the template during promoter binding. Alternatively, the long axis of the DNA may be deflected to the left of the T7 RNAP molecule as depicted here, suggesting a wrapping around or bending of the DNA as it emerges from the cleft. Peptide **IV** contains residues identified as important for processivity (Mookthiar et al., 1991). This part of the C-terminal region of the polymerase is thought to be critical in catalysis and, potentially, substrate discrimination.

It seems unlikely that the clustering of peptides seen in Figure 9 would occur by chance alone.

Comparison of the Promoter-Binding Cleft with a Non-specific Single-Strand DNA-Binding Region and the Functional Relevance of the Cleft. In a previously published work we used a single-stranded DNA oligomer containing a psoralen furan-side monoadduct as a tether to cross-link T7 RNAP (Sastry et al., 1993b). T7 RNAP has a strong affinity

to nonspecific single-stranded DNA (Sastry et al., 1993b; Sousa et al., 1992). We isolated and purified a T7 RNAP peptide that cross-linked to the furan-side oligomer by HPLC methods similar to those described in this paper. By a combination of mass spectrometry and amino acid composition analysis we mapped this single-stranded oligo binding peptide to a region between residues 558 and 608 on the T7 RNAP primary sequence (Sastry et al., 1993b). Because the crystallographic model was not available at that time, we could not locate this peptide on the T7 RNAP 3-D structure. We have now projected this single-stranded oligonucleotide-binding peptide region onto the 3-D model (Figure 10). This peptide maps in the fingers domain helix W (Figure 10) that is one of the helices forming the "framework" on which the cleft "rests" (Sousa et al., 1993). Comparison of Figures 9 and 10 reveals the binding regions for promoter and for single-stranded oligo. The double-stranded promoter bonding occurs in the cleft whereas the nonspecific single-stranded oligo is contacted by an α-helical region of the fingers domain. This result suggests that, while the fingers region may be used by T7 RNAP to make nonspecific contacts with single-stranded DNA, the single-stranded DNA itself is not moved into the cleft, perhaps because it is catalytically nonphysiological. The cleft may be suitably reserved for template promoter. In models derived from X-ray crystal structures of the Klenow fragment of *E. coli* DNA polymerase I and HIV-1 RT, the fingers structures were shown to interact with single-stranded DNA in the case of Klenow and with RNA template in the case of HIV-1 RT (Joyce & Steitz, 1995; Kohlstaedt et al., 1992; Sousa et al., 1993). Consistent with our cross-linking results (Sastry et al., 1993b) T7 RNAP has fingers more like the Klenow fragment rather than HIV-1 RT (Joyce & Steitz, 1995; Sousa et al., 1993).

Alternatively, the single-stranded oligo binding site we mapped in T7 RNAP may be an RNA-binding site in binary complexes [akin to *E. coli* RNAP−RNA complexes; see Altmann et al. (1994)] or an RNA product binding site in ternary complexes. In the T7 RNAP crystallographic model, the single-stranded template and the 3′-terminus of a three-base RNA were shown to lie close to helix Y and the turn C-terminus to helix Y (motif B) (Bonner et al., 1992; Delarue et al., 1990; Sousa et al., 1993). Part of helix W is in the vicinity of helix Y in the 3-D structure. It is conceivable that the RNA chain may be contacted partly by the peptide shown in black (helix W in Figure 10).

Although we did not isolate any peptides in the fingers domains using the promoter DNA, this does not preclude this region from interacting with the promoter at some early stage(s) during promoter complex formation. In Figure 9 we may be witnessing a T7 RNAP molecule in a final-stable complex. Results shown in Figures 9 and 10 imply that depending on the type of DNA, double-strand promoter or single-strand DNA (or RNA), T7 RNAP makes qualitatively different contacts using different domains. This may be related to functional differences between promoter and single-stranded DNA. Although the polymerase bonds the promoter, by some accounts (Sousa et al., 1992) with a weaker overall affinity than single-stranded oligonucleotides, as shown in our works, the promoter is located in the cleft and the single-stranded oligo in the fingers. The location of the promoter in the cleft may facilitate an efficient sliding mechanism that is perhaps necessary during T7 RNAP transcription, in a manner similar to that of other polymerases and transcription factors (Burley et al., 1993; Gamper & Hearst, 1982; Kuriyan & O'Donnell, 1993). Because most single-stranded DNAs are in general very poor substrates for transcription, the cleft may not play a critical role for this sort of activity, and thus they are located outside the cleft. Therefore, the cleft appears to be optimally designed for promoter interactions during transcription initiation.

We have not carried out large-scale cross-linking experiments using nonpromoter DNA because of extremely low cross-linking yields. It is possible that one or more of the same peptides may be cross-linked to nonpromoter DNA because T7 RNAP binds both promoter and nonpromoter DNAs with less than a 1000-fold affinity difference (Gunderson et al., 1987). Our cross-link mapping studies have not yet extended to other DNA-binding proteins to permit us to draw any general conclusions as to the discriminative capability of our technique regarding specific versus nonspecific interactions. In this respect, we do not claim that the peptides we have identified are exclusively for promoter DNA. Hence, our observations are limited to the identification of the T7 RNAP cleft as the seat of promoter binding in solution, confirmation of the crystallographic model using an independent photochemical method, and a discussion of the potential role of the cleft.

The photochemical techniques (Sastry et al., 1993b; this work) we have developed are quite novel. Unlike other photoprobes such as aryl azides, halogenated nucleotides which usually cross-link at best one or two residues in proteins, psoralen photoreacts promiscuously with many amino acids in proteins, perhaps by multiple mechanisms [reviewed in Midden (1988)]. This is also apparent from our data (Table 1) showing several psoralens attached to each peptide. Our photo-cross-linking techniques potentially allow one to identify global nucleic acid-binding domains (rather than pinpoint one or two contacting amino acids) in both single-stranded [using furan-side monoadducts as described in Sastry et al. (1993b)] and/or double-stranded nucleic acid-binding proteins (using free psoralen as in this work). The attractive features are the high cross-link yields and the fact that no prior chemical modification of DNA is required before cross-linking to psoralen. Our techniques are useful for confirming, in solution, crystallographic models (as shown here for T7 RNAP) or for gathering structural information in complexes where no high-resolution structures are available (as is the case with T7 RNAP). In addition, on the basis of the cross-linking results, specific amino acids can be targeted for site-specific mutagenesis experiments. To our knowledge this work is the first instance of a DNA-intercalated stacked photochemical probe used for cross-linking proteins to DNA.

ACKNOWLEDGMENT

I thank Antonio Parraga and David Wilson for assistance with crystal structure displays. I thank Drs. Brian Chait, Urooj Mirza, and Steven Cohen for helpful discussions. I also thank Prof. Joshua Lederberg for his interest in the project.

REFERENCES

Altmann, C. R., Solow-Cordero, D. E., & Chamberlin, M. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3784−3788.

Basu, S., & Maitra, U. (1986) *J. Mol. Biol. 190*, 425−437.

Ben-Hur, E., & Song, P.-S. (1984) *Adv. Radiat. Biol. 11*, 131−170.
Blackburn, S. (1978) in *Amino acid determination: methods and techniques* (Blackburn, S., Ed.) pp 8−35, Marcel Dekker, New York.
Blias, J., Vigny, P., Moron, J., & Bisagni, E. (1984) *Photochem. Photobiol. 39*, 145−156.
Bonner, G., Patra, D., Lafer, E. M., & Sousa, R. (1992) *EMBO J. 11*, 3767−3775.
Bonner, G., Lafer, E. M., & Sousa, R. (1994a) *J. Biol. Chem. 269*, 25120−25128.
Bonner, G., Lafer, E. M., & Sousa, R. (1994b) *J. Biol. Chem. 269*, 25129−25136.
Budowsky, E. I., & Abdurashidova, G. G. (1989) *Prog. Nucleic Acid Res. Mol. Biol. 37*, 1−65.
Burley, S. K., Clark, K. L., Ferre-D'Amare, A., Kim, J. L., & Nikolov, D. B. (1993) *Cold Spring Harbor Symp. Quant. Biol. 58*, 123−132.
Chamberlin, M. J., & Ryan, T. (1982) *Enzymes 15*, 87−108.
Cimino, G. D., Gamper, H., Isaacs, S. T., & Hearst, J. E. (1985) *Annu. Rev. Biochem. 54*, 1151−1193.
Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. (1990) *J. Protein Eng. 3*, 461−467.
Dorman, G., & Prestwich, G. D. (1994) *Biochemistry 33*, 3661−5673.
Dunn, J. J., & Studier, F. W. (1983) *J. Mol. Biol. 166*, 477−535; (1984) *J. Mol. Biol. 175*, 111−112 (erratum).
Fredericksen, S., & Hearst, J. E. (1979) *Biochim. Biophys. Acta 563*, 343−355.
Gamper, H. B., & Hearst, J. E. (1982) *Cell 29*, 81−90.
Goldberg, J., & Dunn, J. J. (1988) *J. Bacteriol. 170*, 1245−1253.
Granger, M., Toulme, F., & Helene, C. (1982) *Photochem. Photobiol. 36*, 175−180.
Gross, L., Chen, W. J., & McAllister, W. T. (1992) *J. Mol. Biol. 228*, 488−505.
Gunderson, S. I., Chapman, K. A., & Burgess, R. R. (1987) *Biochemistry 26*, 1539−1546.
Ikeda, R., & Richardson, C. C. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3614−3618.
Isaacs, S. T., Shen, C.-K. J., Hearst, J. E., & Rapoport, H. (1977) *Biochemistry 16*, 1058−1064.
Joyce, C. M., & Steitz, T. A. (1994) *Annu. Rev. Biochem. 63*, 777−822.
Joyce, C. M., & Steitz, T. A. (1995) *J. Bacteriol. 177*, 6321−6329.
Kohlstaedt, L. A., Wang, J., Friedman, J., Rice, P. A., & Steitz, T. A. (1992) *Science 258*, 1783−1790.
Krauch, C. H., & Farid, S. (1967) *Chem. Ber. 100*, 1685−1695.
Kuriyan, J., & O'Donnell, M. (1993) *J. Mol. Biol. 234*, 915−925.
Lerman, S., Megaw, J., & Willis, I. (1980) *Photochem. Photobiol. 31*, 235−242.
Ling, M. L., Risman, S. S., Klemant, J. F., McGraw, N., & McAllister, W. T. (1989) *Nucleic Acids Res. 17*, 1605−1681.
Maniatis, T., Frotsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp 122, 458, Cold Spring Harbor Laboratory Press, Plainview, NY.
Mantulin, W. W., & Song, P.-S. (1973) *J. Am. Chem. Soc. 95*, 5122−5129.
Martin, C. T., Muller, D. K., & Coleman, J. E. (1988) *Biochemistry 27*, 3966−3974.
McAllister, W. T. (1993) *Cell. Mol. Biol. 39*, 385−391.
Megaw, J., Lee, J., & Lerman, S. (1980) *Photochem. Photobiol. 32*, 265−269.
Midden, W. R. (1988) in *Psoralen DNA photobiology* (Gasparro, F. P., Ed.) Vol. 2, pp 16−49, CRC Press, Boca Raton, FL.
Moffatt, B. A., Dunn, J. J., & Studier, F. W. (1984) *J. Mol. Biol. 173*, 265−269.
Mookthiar, K. A., Peluso, P. S., Muller, D. K., Dunn, J. J., & Coleman, J. E. (1991) *Biochemistry 30*, 6305−6315.
Moyson, A., Voituriez, L., Cadet, J., & Vigny, P. (1993) *J. Photochem. Photobiol. B 17*, 263−271.
Muller, D. K., Martin, C. T., & Coleman, J. E. (1988) *Biochemistry 27*, 5763−5771.
Muller, D. K., Martin, C. T., & Coleman, J. E. (1989) *Biochemistry 28*, 3306−3313.
Musajo, L., Rodighiero, G., Brecccia, A., Dall'Aqua, F., & Malaseni, G. (1966) *Photochem. Photobiol. 5*, 739−745.
Osumi-Davis, P., Sreerama, N., Volkin, D. B., Middaugh, C. R., Woody, R. W., & Woody, A.-Y. M. (1994) *J. Mol. Biol. 237*, 5−19.
Patra, D., Lafer, E. M., & Sousa, R. (1992) *J. Mol. Biol. 224*, 307−318.
Rastinejad, F., & Lu, P. (1993) *J. Mol. Biol. 232*, 105−122.
Rodighiero, G., Dall'Acqua, F., & Averbeck, D. (1988) in *Psoralen DNA Photobiology* (Gasparro, F. P., Ed.) Vol. 1, pp 37−114, CRC Press Inc., Boca Raton, FL.
Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, pp 385−431, Springer-Verlag, Berlin.
Sastry, S. S., & Hearst, J. E. (1991a) *J. Mol. Biol. 221*, 1091−1110.
Sastry, S. S., & Hearst, J. E. (1991b) *J. Mol. Biol. 221*, 1111−1125.
Sastry, S. S., Spielmann, H. P., Dwyer, T. J., Wemmer, D. E., & Hearst, J. E. (1992) *J. Photochem. Photobiol. B 14*, 65−79.
Sastry, S., Spielmann, H. P., & Hearst, J. E. (1993a) *Adv. Enzymol. 66*, 85−148.
Sastry, S. S., Spielmann H. P., Hoang Q. S., Phillips, A. M., Sancar, A., & Hearst, J.E. (1993b) *Biochemistry 32*, 5526−5538.
Schiavon, O., & Veronese, F. M. (1986) *Photochem. Photobiol. 43*, 243−246.
Schmitt, I. M., Chimenti, S., & Gasparro, F. P. (1995) *J. Photochem. Photobiol. B 27*, 101−107.
Shi, Y.-B., & Hearst, J. E. (1987) *Biochemistry 26*, 3792−3798.
Shi, Y.-B., Gamper, H., & Hearst, J. E. (1988) *J. Biol. Chem. 263*, 527−534.
Song, P.-S., & Tapley, K. J. (1979) *Photochem. Photobiol. 29*, 1177−1197.
Sousa, R., Patra, D., & Lafer, E. M. (1992) *J. Mol. Biol. 224*, 319−334.
Sousa, R., Chung, Y. J., Rose, J. P., & Wang, B.-C. (1993) *Nature (London) 364*, 593−599.
Spielmann, P. H., Sastry, S. S., & Hearst, J. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4514−4518.
Spielmann, H. P., Dwyer, T. J., Hearst, J. E., & Wemmer, D. E. (1995a) *Biochemistry 34*, 12937−12953.
Spielmann, H. P., Dwyer, T. J., Sastry, S. S., Hearst, J. E., & Wemmer, D. E. (1995b) *Proc. Natl. Acad. Sci. U.S.A. 92*, 2345−2349.
Tessman, J. W., Isaacs, S. T., & Hearst, J. E. (1985) *Biochemistry 24*, 1669−1676.
Turro, N. J. (1991) *Modern molecular photochemistry*, University Science Books, Mill Valley, CA.
Veronese, F. M., Schiavon, O., Bevlacqua, B. F. & Rodighiero, G. (1981) *Photochem. Photobiol. 34*, 351−354.
Veronese, F. M., Schiavon, O., Bevlacqua, B. F., & Rodighiero, G. (1982) *Photochem. Photobiol. 36*, 25−30.
Williams, K. R., & Konigsberg, W. H. (1991) *Methods Enzymol. 208*, 516−539.
Yoshikawa, K., Mori, N., Sakakibara, S., Mizuno, N., & Song, S. (1979) *Photochem. Photobiol. 29*, 1127−1133.

BI960982D