Measurement of Interstrand Cross-Link Frequency and Distance between Interruptions in DNA Exposed to 4,5′,8-Trimethylpsoralen and Near-Ultraviolet Light†

Noboru Matsuo‡ and Peter M. Ross*

The Laboratory for Investigative Dermatology, The Rockefeller University, New York, New York 10021-6399

Received July 14, 1986; Revised Manuscript Received October 31, 1986

ABSTRACT: Bifunctional psoralens react photochemically with DNA to form single-strand adducts and interstrand, chemical cross-links. Cross-link formation is first order with [P], the concentration of added psoralen, when [P] $\ll K_d$, the psoralen–DNA dissociation constant. DNA molecules containing interstrand cross-links are reversibly bihelical and so are readily detected. It was not heretofore possible to determine cross-link frequency in polydisperse DNA from the mass F of DNA spared cross-linkage. We have derived a statistical relation to calculate cross-link frequency at fixed light exposure and variable [P]. We show here that S, the initial slope of the curve described by $-\ln F$ as a function of [P], is proportional to $\bar{M}_w$, the weight-average molecular weight of nick-free DNA. The cross-link frequency at any [P] can be determined from k, a constant measured for DNA of known $\bar{M}_w$ at low cross-linkage. This relation is valid for DNA of any molecular weight distribution. In experiments with uniform length DNA, $-\ln F$ (cross-link frequency) increased in simple proportion to [P]. Intact and restriction endonuclease *Hin*dIII digested phage λ DNA molecules have discrete lengths. S for each was proportional to $\bar{M}_w$ of the twin helix even though the molecular weight distribution of the restriction fragments was skewed. S was proportional to $\bar{M}_w$ and to the median molecular weight of sheared cellular DNA over a wide range. Also, we found that $1/S$ was linear with exposure of cellular DNA to γ radiation. S can therefore be used to calculate $\bar{L}$, the average distance between interruptions in the double helix.

Psoralen and its derivatives are furocoumarins that intercalate between bases in native DNA [see Song and Tapley (1979) for review]. If the psoralen–DNA complex is exposed to near-ultraviolet light (around 360 nm), covalent linkage can occur by psoralen cycloaddition to pyrimidine bases (Straub et al., 1981). Bifunctional psoralens can form interstrand chemical cross-links by sequential photoreaction with pyrimidine bases at adjacent levels on opposite sides of the antiparallel DNA chain (Cole, 1970; Johnston et al., 1977; Kanne et al., 1982).

For uniform length DNA from bacteriophage or virus, cross-link frequency can be deduced from the fraction of DNA mass remaining free from cross-links, assuming random cross-link distribution. This fraction can be calculated from measurements of molecular weight or reversible bihelicity (Cole, 1970; Ross & Howard-Flanders, 1977). These methods have not been applied effectively to DNA purified from cells because mechanical shearing introduces nonrandom breaks, thwarting simple statistical analysis. No conventional approximations described our data throughout the range of accurate measurement, so we were unable to determine cross-link frequency from measurements of cross-linkage, the fraction of DNA mass containing cross-links.

We set out to measure cross-link frequency in cellular DNA in a manner that would be unaffected by the molecular weight distribution of the sample under study. We first derived a statistical description of cross-link frequency in DNA of any molecular weight distribution. We then tested our model and compared the results with those from published procedures.

Table I: Predicted Dark Binding of TMP and 8-MOP to DNA in Limited Reaction Volume

[S] (mol/L)	[DNA][a] (μg/mL)	[PS]/[S] 8-MOP[b] (L/mol)	[PS]/[S] TMP[c] (L/mol)	[TMP][d] to 5% decreased slope (mol/L)
1×10^{-12}	2.3×10^{-6}	768	17 600[e]	4×10^{-6}
1×10^{-7}	0.23	766	17 800	$>1 \times 10^{-6}$
1×10^{-5}	23	763	15 100	6×10^{-6}
5×10^{-5}	115	738	9 390	1×10^{-5}
1×10^{-4}	230	714	3 900	2.5×10^{-5}
2×10^{-4}	460	666	950	$>5 \times 10^{-5}$

[a] Assuming [S] = 0.3[DNA]; see text. [b] $K_d = 1.3 \times 10^{-3}$ mol/L; [8-MOP] = 10^{-6} mol/L. [c] $K_d = 5.6 \times 10^{-5}$ mol/L; [TMP] = 10^{-7} mol/L. [d] Concentration of ligand above which more than 5% of binding sites are saturated. [e] [TMP] = 1.7×10^{-6} mol/L.

Theory

Binding of Psoralen to DNA. Psoralens intercalate between base pairs of duplex nucleic acid (Dall'Acqua et al., 1978, 1979a; Isaacs et al., 1977). The dissociation constant for this "dark", noncovalent binding of psoralen to DNA is defined as

$$K_d = [P_f][S_f]/[PS]$$

where $[P_f]$ is the concentration of free psoralen, $[S_f]$ is the concentration of unoccupied binding sites, and [PS] is the concentration of bound psoralen. If [S] and [P] are the total concentration of binding sites and added psoralen, then $[S] = [S_f] + [PS]$ and $[P] = [P_f] + [PS]$. Solving for [PS]

$$[PS] = \tfrac{1}{2}\{[P] + [S] + K_d - \sqrt{[P]^2 - 2([S] - K_d)[P] + ([S] + K_d)^2}\} \quad (1)$$

Interstrand cross-links form in proportion to [PS] (Dall'Acqua

† This work was supported by Research Grant 1461 from the Council for Tobacco Research and a grant from The Skin Cancer Foundation to P.M.R. and by General Clinical Research Center Grant RR00102 from the National Institutes of Health to the Rockefeller University Hospital.

* Author to whom correspondence should be addressed.

‡ Present address: Tochigi Research Laboratories, Kao Corp., 2606, Akabane, Ichikaimachi, Haga, Tochigi, 321-34 Japan.

0006-2960/87/0426-2001$01.50/0

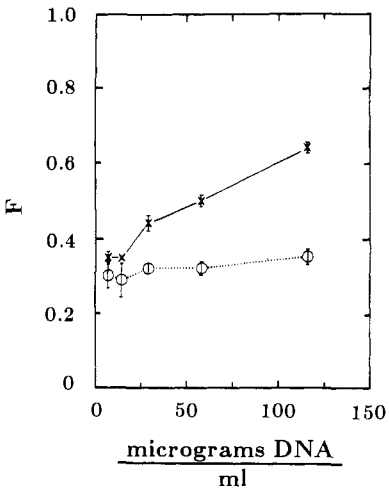

FIGURE 1: Effect of [DNA] on psoralen photo-cross-linkage: solid lines and ×'s, TMP; dotted lines and circles, 8-MOP. Phage λ DNA diluted serially in TE buffer to the concentrations shown on the abscissa was incubated with 1.8×10^{-8} mol/L TMP or with 1.1×10^{-5} mol/L 8-MOP and exposed to 4 kJ/m^2 360-nm light. Samples were diluted in TE buffer to 14.4 μg/mL (or left at 7.2 μg/mL), denatured in alkali, and tested for S1 sensitivity as described under Materials and Methods. Un-cross-linked DNA of the same concentration was denatured and digested with S1 as a control.

et al., 1979b). When [PS] is proportional to [P], photobinding can be predicted from [P], which is much easier to measure than [PS]. We tested the linearity of eq 1, using 1.3×10^{-3} mmol/L for 8-methoxypsoralen (8-MOP) (Rodighiero & Dall'Acqua, 1984) and 5.6×10^{-5} mol/L for 4,5′,8-trimethylpsoralen (TMP) (Isaacs et al., 1977) as dissociation constants.

In Table I, [PS]/[S] was calculated from eq 1. Columns 3 (for 8-MOP) and 4 (for TMP) show a wide range where [PS]/[S] is linear with [P], [P] $\ll K_d$ for the psoralen–DNA complex, and the slopes of the predicted binding curves decrease with [S]. To test for the effect of [S] on [PS], phage λ DNA at concentrations ranging from 7.2 to 116 μg/mL was incubated with dilute 8-MOP or TMP and exposed to 4 kJ/m^2 360-nm light. Figure 1 shows that F increased with [S] for TMP treatment but not for 8-MOP treatment at [DNA] exceeding 14 μg/mL. Since F at 115 μg of DNA/mL was double F for very dilute DNA, there was about 0.3 TMP binding site per nucleotide pair. Table I, column 5, shows that at [S] $< 5 \times 10^{-5}$ mol/L (about 350 μg of DNA/mL) TMP dark binding sites are saturated when [TMP] exceeds 5×10^{-6} mol/L. For most experimental and therapeutic situations, [TMP] is below this level, but similar behavior is expected of hydrophilic psoralen derivatives with high binding affinity, like 4′-(aminomethyl)trimethylpsoralen (Isaacs et al., 1977).

Calculation of Cross-Link Frequency from the Fraction of Reversibly Bihelical DNA in a Mixture of Cross-Linked and Un-Cross-Linked Molecules. If cross-links are formed as random, independent events, their frequency follows the Poisson distribution (Haight, 1967). When there are n cross-links per length L, the probability P_r that r cross-links occur in L is $P_r = e^{-n}n^r/r!$. F, the fraction of DNA-spared cross-linkage, is e^{-n}. Thus

$$n = k[\mathrm{P}]L = -\ln F \tag{2}$$

where k is a cross-linking constant determined by the kinetics of psoralen photoaddition. F can be measured experimentally, and a plot of $-\ln F$ against [P] will be a straight line of slope kL.

Cellular DNA is usually sheared during purification (Mandell & Hershey, 1960; Reynolds, 1978), so its molecular weight distribution is neither uniform nor random (Rubenstein et al., 1961; Harrington & Zimm, 1965). Sheared DNA of average length $\bar{L}$ has i length groups. If p_j is the fraction of total DNA in the jth group, l_j is the length in base pairs of jth group molecules, and F_j is the fraction of DNA free from cross-links in l_j base pair long DNA

$$F = \sum_{j=1}^{i} p_j F_j = \sum_{j=1}^{i} p_j e^{-k[\mathrm{P}]l_j} \tag{3}$$

The slope in a plot of F against log [P] for DNA mixtures in eq 3 will be shallower than that for DNA of a single molecular weight because of the various length components. If we define

$$Y = -\ln F = -\ln \sum_{j=1}^{i} p_j e^{-kl_j[\mathrm{P}]}$$

a plot of Y against [P] will not give a straight line when $i \geq 2$, because DNA molecules differing in length differ in cross-link susceptibility. While it is not obvious from such a plot how to determine the cross-link frequency, the slope of the curve described in a plot of Y against [P] at any [P] is given by the derivative of Y with respect to [P]:

$$\frac{dY}{d[\mathrm{P}]} = -\frac{1}{\sum_{j=1}^{i} p_j e^{-kl_j[\mathrm{P}]}} \frac{d}{d[\mathrm{P}]} \sum_{j=1}^{i} p_j e^{-kl_j[\mathrm{P}]}$$

$$= \frac{k\sum_{j=1}^{i} p_j l_j e^{-kl_j[\mathrm{P}]}}{\sum_{j=1}^{i} p_j e^{-kl_j[\mathrm{P}]}}$$

The initial slope S is the derivative at [P] = 0:

$$S = \frac{dY}{d[\mathrm{P}]_{[\mathrm{P}]=0}} = \frac{k\sum_{j=1}^{i} p_j l_j}{\sum_{j=1}^{i} p_j} = k\bar{L} \tag{4}$$

Equation 2 can be averaged to yield the cross-link frequency per molecule of average length:

$$\bar{n} = k[\mathrm{P}]\bar{L} \tag{2a}$$

From eq 4 and 2a, $\bar{n}/\bar{L} = S[\mathrm{P}]/\bar{L}$.

When hydrodynamic shear or restriction digestion introduces no nicks, $\bar{L} = \bar{M}_w$ (Tanford, 1961). F can be defined either as that fraction of DNA not connected to a cross-link or as the fraction of material that is not reversibly bihelical. When the DNA contains strand breaks, these quantities differ. In our assays, only DNA that is connected to a cross-link in both strands is reversibly bihelical, so $\bar{L}$ is the average length of intact helix. Therefore

$$\bar{L}_u = S_u\bar{L}_k/S_k \tag{5}$$

where $\bar{L}$ is the mean distance between breaks in either strand and the subscripts refer to a standard DNA of known $\bar{L}$ and the unknown, nicked sample cross-linked under identical conditions. Equations 4 and 5 are valid for any pattern of breakage.

Materials and Methods

Materials. Agarose for gel electrophoresis, phage λ DNA and its *Hin*dIII restriction fragments, and *Eco*RI endonuclease were obtained from Bethesda Research Laboratories. DNA

contained less than one strand break per three to four molecules, as estimated by alkaline agarose gel electrophoresis. S1 nuclease, purchased from New England Nuclear, was used only when no double-strand activity was detectable under our digestion conditions.

DNA Preparation. Plasmid pUC19 DNA (1.8×10^6 daltons = 2686 nucleotide pairs) was chloramphenicol amplified in broth cultures of *Escherichia coli* HB101 carrying this plasmid (Maniatis et al., 1982) and purified as described by Garger et al. (1983). Contaminating RNA was removed with boiled ribonuclease A, phenol extraction, ethanol precipitation, and Sephacryl S-500 (Pharmacia) gel filtration. Closed circular plasmid DNA was digested with *Eco*RI endonuclease and purified by gel filtration through Sephacryl S-500, and the intactness of the final product was confirmed by alkaline agarose gel electrophoresis. *E. coli* DNA labeled metabolically with [^{3}H]thymidine and DNA from Schneider-2 cultured *Drosophila melanogaster* cells were prepared by detergent and organic extractions as described elsewhere (N. Matsuo and P. M. Ross, submitted for publication).

Psoralen plus Light Treatment of DNA. 4,5′,8-Trimethylpsoralen (TMP; from J. Paul Elder & Co.) was recrystallized from ethanol 3 times, dissolved and serially diluted 2-fold in ethanol, and stored in glass vials in a dark drawer. [TMP] was measured by spectrophotometry (ϵ_{249} = 31 008 M^{-1} cm^{-1}). DNA diluted to 20 or 30 μg/mL in TE buffer [10^{-2} mol/L tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 7.5, 10^{-3} mol/L disodium ethylenediaminetetraacetate (Na_2EDTA)] was mixed with 0.01 volume of TMP, held 10 min at room temperature, and exposed to light from a battery of four 48-in. fluorescent lamps (Sylvania Lifeline FR40712/PUVA) emitting primarily at 360 nm. A 2 cm thick layer of 10% Cu_2SO_4 in 0.01 mol/L H_2SO_4 in a polystyrene box interposed between light source and sample absorbed short wavelengths. The incident fluence, measured with an actinometrically calibrated Black-ray near-ultraviolet light (UVA) meter (Model NOJ 221, Ultraviolet Products, Inc., Pasadena, CA), ranged from 5 to 10 W/m^2. Total dose was 4 kJ/m^2 in all experiments described here. Incident and effective doses were the same when tested.

Measurement of Cross-Linked DNA Concentration by Nuclease Sensitivity of DNA Free from Interstrand Cross-Links. In most experiments, 65 μL of sample containing approximately 1 μg of DNA in TE buffer (10 mM Tris-HCl, pH 7.5, containing 1 mM Na_2EDTA) was denatured by adding 7 μL of 1 M NaOH and holding at 37 °C for 30 min. This solution was neutralized with 7 μL of 1 mol/L acetic acid. Twenty microliters of buffer containing 0.3 mol/L sodium acetate, pH 4.6, 0.25 mol/L NaCl, 12.5% glycerol, 10 mM $ZnSO_4$, and 20 units of S1 nuclease was added to the denatured DNA sample. One unit of S1 nuclease degrades 1 μg of single-stranded DNA to acid-soluble material in 1 min at 37 °C. After 20 min of incubation at 37 °C, 1 mL of 1 μg/mL ethidium bromide in TE buffer was added to the mixture and residual double-stranded DNA was measured by fluorescence (λ_{ex} = 320 nm; λ_{em} = 600 nm) in borosilicate glass tubes. Fluorescence intensity increased in linear proportion to DNA concentration up to 2 μg/mL. Phage λ DNA was used as a standard. Our fluorometric assay is simpler than the acid precipitation methods commonly used (Ben-Hur et al., 1979; Fujiwara, 1982). Similar results are obtained by the two methods of measurement. The fraction F of DNA remaining free from cross-links was determined from the total fluorescence (T) of native DNA digested with S1, the background fluorescence (B) of un-cross-linked, denatured, and digested DNA, and the fluorescence of each denatured sample (D) as $F = 1 - (D - B)/(T - B)$.

Measurement of S. Data in the range of low cross-link frequency were plotted as $-\ln F$ against [P]. The initial slope of each plot was determined either by linear regression or by the average of measurements by each of three independent observers.

Determination of Average Molecular Weight of Sheared DNA. ^{3}H-Labeled *E. coli* DNA syringe sheared to various lengths and denatured in 0.17 mol/L NaOH was resolved by alkaline agarose gel electrophoresis in 0.01 M Na_2EDTA at pH 12.5–13.0 (McDonnell et al., 1977). After electrophoresis, the gel was stained with 5 μg/mL ethidium bromide in 0.05 M Tris-HCl, pH 7.0, for 1 h and destained in the same buffer without ethidium bromide. Fluorescence (λ_{ex} = 300 nm) was photographed through red and UV filters, and the gel was then cut into lanes. Each lane was sliced into 5-mm pieces which were placed into vials, supplemented with 1 mL of water, and heated to 100 °C until the gel melted. Radioactivity was measured by scintillation counting in Hydrofluor (National Diagnostics). Weight-average molecular weight was calculated as $\bar{M}_w = \sum M_i C_i / \sum C_i$, where C_i is the radioactivity or mass of each component or fraction and M_i is the molecular weight at that fraction. $\bar{M}$, the median molecular weight, is the molecular weight at 50% of the total mass. Although $\bar{M}_w$ usually exceeds $\bar{M}$, we find $\bar{M}$ useful for comparing samples because it is so easily measured (see below).

γ Irradiation. DNA solutions were exposed under air in glass tubes to a ^{137}Cs source at 1.5 krad/min incident dose.

Results

Cross-Link Frequency in Uniform Length DNA. To test our experimental system, we measured the fraction of uniform length DNA cross-linked by varied [P] at fixed light exposure. The single *Eco*RI site in pUC19 is located about 500 nucleotides from the origin of replication of the plasmid (Norrander et al., 1983). We were concerned that the short stretch of RNA normally found at the origin of replication in plasmids under relaxed replication control prepared by chloramphenicol amplification might affect our results. However, there was no evidence by agarose gel electrophoresis for alkaline breakage in closed circular plasmid DNA exposed to our denaturation conditions (not shown).

*Eco*RI-cut pUC19 DNA was treated with various concentrations of 4,5′,8-trimethylpsoralen (TMP), exposed to 4 kJ/m^2 of 360-nm light from the PUVA lamp, denatured in alkali, and digested with S1 nuclease. Residual, cross-linked DNA was detected by ethidium bromide fluorescence. The S1-sensitive fraction is plotted in Figure 2 (lower curve) as $-\ln F$ against [TMP]. As expected for uniform length DNA (eq 2), this plot is linear to the highest [P] shown, at which 85% of the DNA contained cross-links. The value of k is 3767 L/(mol·base pair). Cross-link frequency at 10^{-7} mol/L TMP, for example, is 3.77×10^{-4} per base pair or 1 cross-link per 2653 base pairs. k may also be thought of as the rate constant of cross-link formation under the conditions of irradiation, in this case 3.77 per kilobase per micromolar TMP.

We also tested λ DNA, a linear molecule 48 502 nucleotide pairs in length (Sanger et al., 1982). Measurements (upper curve, Figure 2) were as reproducible as those with pUC19 DNA. However, there was a decline in slope at [TMP] > 3×10^{-8} mol/L. The regression line calculated to estimate the cross-link frequency is biased by the higher values of $-\ln F$ and so is systematically low. The λ DNA contained fewer than 20% broken strands by alkaline gel electrophoresis; the poor fit illustrates the sensitivity of our measurement system and

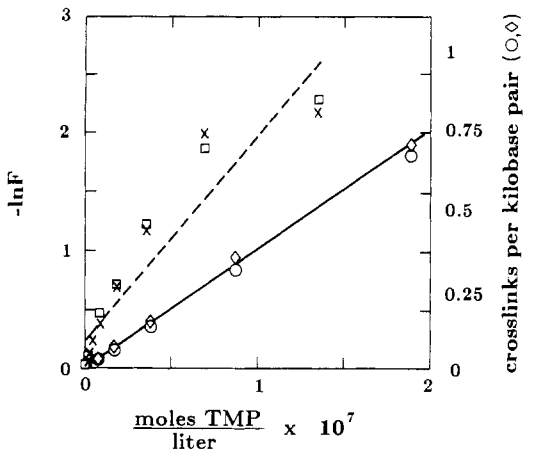

FIGURE 2: Reversible bihelicity caused by psoralen plus light in (diamonds and circles) linearized plasmid pUC19 DNA (length = 2686 base pairs) or (×'s and squares) phage λ *cI857 S7* DNA (length = 48 502 base pairs). *Eco*RI-digested plasmid pUC19 diluted to 20 μg/mL, or whole phage λ DNA at 30 μg/mL in TE buffer (pH 7.5), was supplemented with TMP, exposed to 360-nm light at 10 W/m^2, and tested for reversible bihelicity as described under Materials and Methods. The symbols at each concentration are duplicate measurements. Smooth curves, k calculated from data at F near 0.8 by using $k = -\ln F/[P]L$ (eq 2).

the need for a way to analyze such data.

Cross-Link Frequency in DNA Mixtures of Known Length Dispersion. The fragment length distributions in restriction enzyme digests of bacteriophage DNA are both nonrandom and well-defined. We exploited this for a critical test of whether DNA in solution behaves according to eq 4. *Hin*dIII restriction enzyme cleaves λ DNA into fragments 23 130, 9416, 6682, 4361, 2322, 2027, 564, and 125 base pairs long in equimolar concentration, with $\bar{M}_w$ = 14 374 base pairs. We compared cross-linking in a *Hin*dIII digest of λ DNA and intact λ DNA exposed separately to TMP and light. Both DNA preparations were diluted to 30 μg/mL and supplemented with TMP, exposed to 360-nm light, denatured in alkali, and brought to pH 4.5. Figure 3a shows the fraction of DNA remaining S1 sensitive after the above treatment. Data and also the curves calculated from k for λ DNA data near F = 0.8 are shown; k is 980 L/(mol·base pair).

We replotted the data of Figure 3a as $-\ln F$ against [TMP] in Figure 3b. Equation 4 predicts that the ratio of the initial slopes should be the same as the ratio of the weight-average molecular weights. The dashed lines from the calculated curves of Figure 3a and the solid regression curves fit the data in the range shown. $S = 5 \times 10^7$ L/mol for λ DNA; for λ *Hin*dIII fragments $S = 1.4 \times 10^7$ L/mol. The ratio of the S values is 3.57, exceeding by 6% the calculated ratio of weight-average molecular weights, 3.37. The independent estimate of k from the initial slope of the restriction digest experiment was 1043, 6% higher than the value for intact phage DNA. Within this range of error, the results indicate that eq 4 is valid for DNA nonrandom in length distribution.

A length effect such as excluded volume (Wetmur & Davidson, 1968) could explain the 6% difference in the observed and theoretical ratios of S values if the intact DNA was less accessible to TMP than the shorter molecules. To test this possibility, we compared reversible bihelicity in DNA fragmented before or after cross-linking. Interstrand TMP cross-links do not affect the action of restriction endonuclease *Eco*RI (P. M. Ross and H.-S. Yu, unpublished results). *Eco*RI nuclease cleaves λ DNA into fragments 21 226, 7421, 5804, 5643, 4878, and 3530 nucleotide pairs in length ($\bar{M}_w$ = 12 523 base pairs). We next treated intact λ DNA and also *Eco*RI

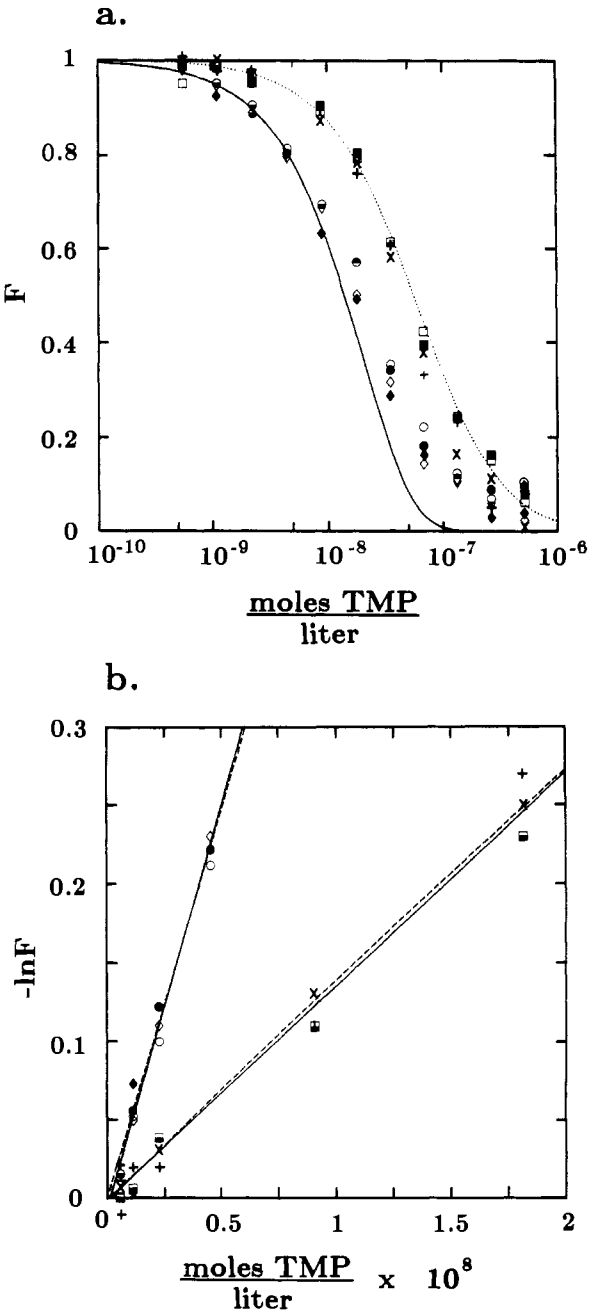

FIGURE 3: Reversible bihelicity due to psoralen plus light treatment of whole phage λ DNA or *Hin*dIII restriction fragments of λ DNA. (a) F as a function of [TMP]. Solid line, curve calculated as described in Figure 2 for the data on λ DNA; dotted line, similar curve for the data on *Hin*dIII digestion fragments, but calculated by using eq 3. (b) $-\ln F$ as a function of [TMP]. Solid lines, linear regression curve fit to the data; dashed lines, calculated curves from (a) replotted as $-\ln F$. Paired symbols are for duplicate measurements on samples from the same batch of cross-linked DNA: (♦, ◇, ○, and ●) measurements on whole λ DNA; (+, × and □, ■) measurements on *Hin*dIII restriction fragments of phage λ DNA; (◇, ♦ and +, ×) data from separate fluorometric readings on the same processed sample; (○, ● and □, ■) data for two samples digested in separate tubes, processed several days after the same just described. This tests the stability and reproducibility of each step. Whole λ DNA or its *Hin*dIII digest, diluted in TE buffer (pH 7.5) to 30 μg/mL, was supplemented with TMP and exposed to 4 kJ/m^2 360-nm light. Procedures were otherwise as described under Materials and Methods.

fragments of λ DNA with TMP and light, digested the intact DNA with the restriction enzyme, and then measured duplex DNA in each sample. The cross-linking curves overlapped [k = 1400 L/(mol·base pair) for DNA cross-linked before digestion; k = 1348 and 1250 L/(mol·base pair) in separate

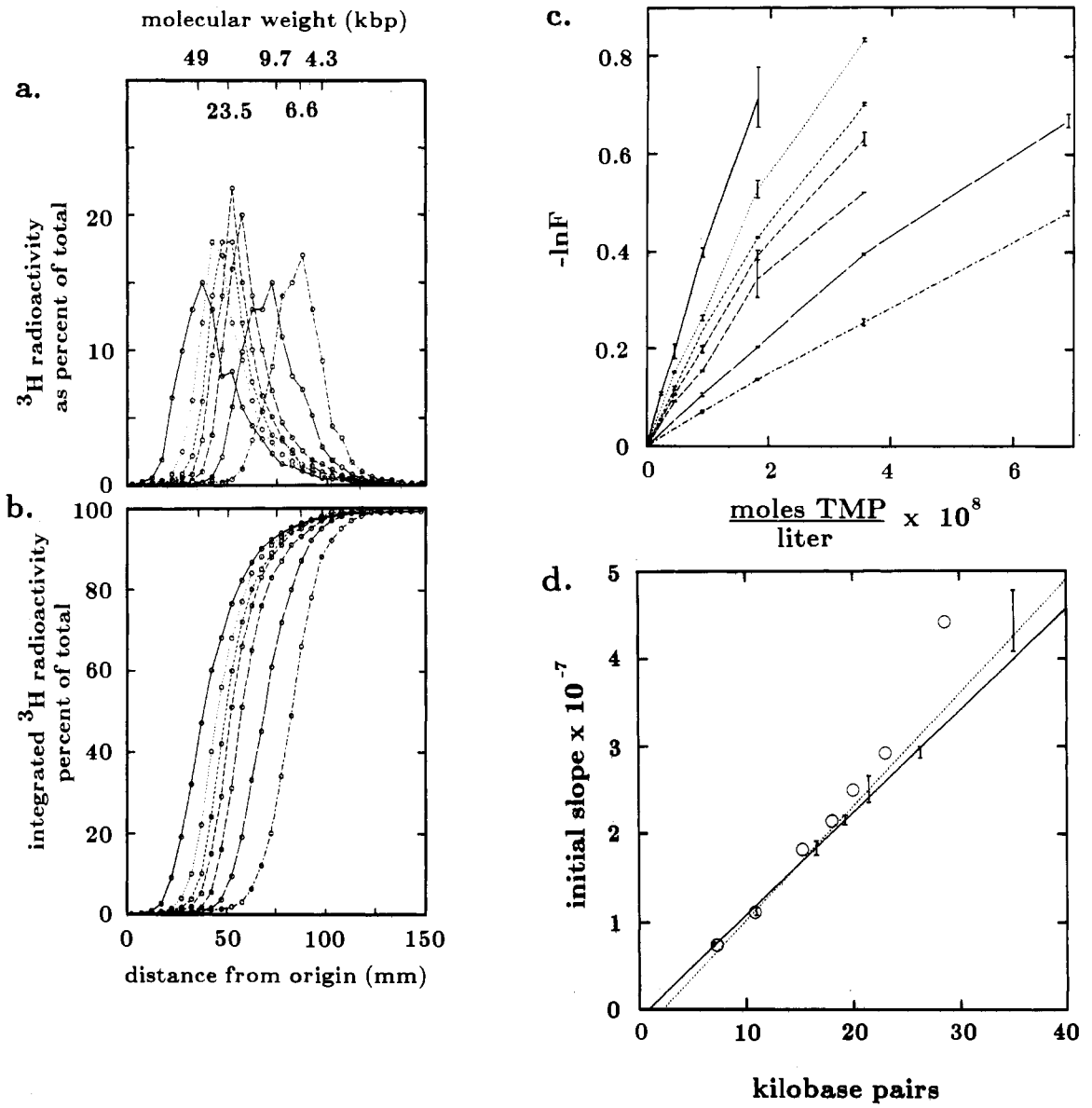

FIGURE 4: Effect of shear on the alkaline electrophoretic mobility and cross-linkage of *E. coli* DNA. Syringe shear decreased the average DNA strand length at least 5-fold to about 7 kilobases and had a like effect on the initial slope. (a) Radioactivity in each slice as a percent of the total in the lane is plotted against the distance of the slice from the origin. (b) Percent radioactivity integrated over the distance migrated from the origin. The lower abscissa shows the distance migrated through the gel; the upper abscissa gives the positions of marker λ DNA and *Hin*dIII digestion fragments of λ DNA. Samples were in increasing level of shear: (—) 1, (···) 2, (- - -) 3, (– –) 4, (– · –) 5, (— —) 6, (– · · –) 7. (c) $-\ln F$ as a function of [TMP] added to sheared, *E. coli* DNA before exposure to 360-nm light. Larger DNA molecules were more readily cross-linked and therefore have a steeper slope. Lines connecting the average measurements correspond to those in (b). Error bars show the range of data for two separate digests of the same batch of cross-linked DNA. (d) S as a function of the molecular weight of the DNA. The initial slope of each line in (c) was measured for data in the range $-\ln F < 0.3$. $\bar{M}$ and $\bar{M}_w$ are from Table I. Lines and error bars are for $\bar{M}$; circles are for $\bar{M}_w$. ^{3}H-Labeled *E. coli* at a concentration of 380 μg/mL was sheared by repeated passage through a 27-gauge hypodermic needle. Seven samples (3 mg each) removed after progressively greater shear were resolved by alkaline agarose gel electrophoresis to measure molecular weight or cross-linked and digested with S1 nuclease to measure F and S by fluorometry, as described under Materials and Methods.

curves for DNA digested before cross-linking]. A two-tailed t test applied to data throughout the range of measurement yielded a product–moment correlation of 0.99. We therefore reject the view that excluded volume influenced results in this or the previous experiment.

Cross-Link Frequency and Molecular Weight of Sheared DNA. Although the fragment size frequencies in most sheared DNA preparations do not follow the Gaussian distribution, $\bar{M}_w$ is readily determined. We next cross-linked sheared DNA for which we measured the average molecular weight independently and could therefore predict the relative cross-link frequency with eq 2a from k. DNA labeled uniformly with [^{3}H]thymidine was variously sheared and then resolved by electrophoresis through an alkaline agarose gel. DNA distributions were recorded first by ethidium bromide fluorescence (Table II). Then, each lane was sliced into pieces, and the distribution of radioactive DNA was measured by liquid scintillation counting. Profiles of the radioactivity are shown in Figure 4a. Figure 4b shows the counts integrated over distance migrated, as a percent of the total. The distance from the origin to 50% maximum integrated counts for any lane corresponds to $\bar{M}$ for that sample. The lengths obtained by comparison of the radioactivity distributions with the distances migrated by a marker *Hin*dIII digest of phage λ DNA (upper abscissa, Figure 4a) were used as an independent measurement of molecular weight in the next experiment.

Samples of the sheared DNA were incubated in the dark with TMP and then exposed to 4 kJ/m^2 360-nm light and

Table II: Single-Strand Molecular Weight, Initial Slope, and Cross-Linking Constant for *E. coli* DNA Subjected to Various Levels of Hydrodynamic Shear

sample no.	molecular weight[a] $\bar{M}_w$[b]	$\bar{M}$ by ^{3}H radioactivity[c]	$\bar{M}$ estimated by eye[d]	$\bar{M}$ from densitometric scan of negative[e]	$S \times 10^{-7}$, initial slope[f]	k, cross-linking constant, $k = S/\bar{M}_w$[g]
1	28.5	35.0	30.8 ± 6.3	34.0	4.43	1554
2	23.0	26.3	27.0 ± 2.4	24.0	2.93	1274
3	20.1	21.5	22.2 ± 1.4	19.5	2.51	1249
4	18.0	19.3	19.6 ± 1.1	19.5	2.16	1200
5	15.23	16.6	17.0 ± 1.4	15.3	1.84	1208
6	10.8	11.0	11.4 ± 0.83	10.1	1.12	1037
7	7.4	7.03	6.8 ± 0.25	7.1	0.755	1020
						1220 ± 164[h]
						1100 ± 68[i]

[a] Expressed in kilobases for DNA resolved by electrophoresis through a 0.6% alkaline agarose gel. [b] Weight-average molecular weight, from data of Figure 4b; average of two determinations. [c] Median molecular weight, same samples; average of two determinations. [d] Mean ± SD of two estimations by each of three observers. [e] Single determination with a Beckman DU-8 scanning spectrophotometer. [f] In L/mol, average of a single determination by each of three observers. [g] In L/(mol·base pair); $\bar{M}_w$ from column 2. [h] Average ± SD; calculation is for all data, including the unsheared sample. [i] Average ± SD calculated for all data except unsheared sample.

denatured, and the fraction of DNA containing cross-links was measured by ethidium bromide fluorescence in the pH 4.5 S1 digest. The results are plotted in Figure 4c as $-\ln F$ against [P]. The initial slope of each plot in Figure 4c is shown in Table II along with the average molecular weights and k calculated from the relation $S = k\bar{M}_w$. To test this relationship, the initial slopes shown in Table II were plotted as a function of the sample molecular weights in Figure 4d. The points are connected by straight lines calculated by regression analysis. The correct line should have a slope equal to k, the cross-linking constant. The solid regression curve [k = 1099 L/(mol·base pair)] was calculated for $\bar{M}$, by omitting the highest molecular weight sample, while the dashed line [k = 1120 L/(mol·base pair)] was calculated by including that sample. There was departure from linearity for DNA exceeding 30 kilobases in length. Comparison with sucrose gradient sedimentation data (other experiment not shown) revealed the deviation at high $\bar{M}_w$ was due only to systematic inaccuracy in electrophoretic measurement of $\bar{M}_w$. The data in Figure 4 thus confirm eq 4 experimentally for DNA having continuous molecular weight distribution.

For the experimental conditions described here, values of F were the same whether calculated from acid-precipitable radioactivity or ethidium bromide fluorescence data. It follows that the distribution of fluorescence in the gel should match the distribution of ^{3}H radioactivity. We estimated the average molecular weight of the DNA from the fluorescence photographs of the agarose gels (Table II). The visual estimates were identical with densitometric scans of photographic negatives of ethidium bromide fluorescence in the gels and with the distributions of ^{3}H radioactivity throughout the range of accurate molecular weight measurement (Table II, columns 2 and 3). Visual inspection of ethidium bromide fluorescence may therefore be used for routine measurement of $\bar{M}$ in polydisperse DNA resolved by agarose gel electrophoresis.

Distance between Interruptions in Irradiated DNA. In the previous experiments, the substrate DNA lacked single-strand tails. γ irradiation of purified DNA in dilute aqueous solution causes a broad spectrum of DNA damage, including single-strand breaks, alkali-labile sites, double-strand breaks, and a variety of alkali-stable adducts and damaged bases [see Hutterman et al. (1978) for reviews]. To measure $\bar{L}$ in irradiated samples, DNA purified from cultured *Drosophila* cells was incubated with various [TMP] and exposed to 360-nm light, then placed near a ^{137}Cs source, and sampled at intervals. Ethidium bromide fluorescence in alkaline and neutral agarose electrophoretic gels (data not shown) revealed that fewer than one in eight breaks and alkali-labile sites caused by γ radiation in this DNA was double-stranded. Each sample was held in alkali for denaturation and to cause chain scission at alkali-labile sites. The single-stranded DNA was degraded with S1 nuclease. In Figure 5a, the γ-irradiated samples increased less rapidly in $-\ln F$ with increased [TMP] than unirradiated DNA since strand breaks reduced the length of DNA made reversibly bihelical at any frequency of cross-links per nucleotide.

The linear relationship between $1/S$ and exposure to γ radiation in Figure 5b is expected from eq 5 if γ breakage is first order with exposure. To calculate the average distance between radiation-induced breaks, we syringe sheared DNA from the original sample used for this experiment to $\bar{M}$ = 19000 base pairs (alkaline agarose gel not shown). The ratio of S of the sheared to the intact sample was 9.05 (Figure 5a), so the median distance between breaks in either or both strands of the original DNA sample was 172 kilobase pairs. The right ordinate in Figure 5b shows the average frequency of γ nicks and breaks, calculated by substituting $\bar{L}$ ($=\bar{M} \approx \bar{M}_w$; Table II) and S for the sheared DNA into eq 5, assuming $\bar{L}_n = \bar{L}/2$, which is justified if γ breakage is random (Charlesby, 1960). Apparently, the initial slope permits accurate determination of $\bar{L}$ in the range 10–170 kilobases.

Discussion

Unlike most binding analysis where the concentration of free ligand is held constant by dialysis, we addressed the common experimental condition of limited reaction volume. When [P] is well below K_d, [PS] is proportional to [P], and the number and kind of binding site in DNA can be used to predict the fraction of sites occupied at any [P]. Our calculations predict a significant decrease in binding due to ligand depletion at [S] exceeding 50 μM for 8-MOP or 5 μM for TMP (Table I). Intuitively, [S] should be proportional to but less than [DNA] expressed in nucleotide pairs and should depend upon the steric hindrance caused by an intercalating psoralen molecule. Dall'Acqua and co-workers (1978, 1979a) found a single kind of DNA binding site, between 0.05 and 0.2 per base pair, although there was some evidence for sequence specificity in binding. Hyde and Hearst (1978) found evidence for secondary sites, but only at elevated [P], in the range of K_d. The data in Figure 1 show F doubled in dilute TMP at [DNA] = 115 μg/mL, as predicted from eq 1 if [S] = 0.3[DNA] (Table I). The similarity of our estimate of [S] from [F] to literature

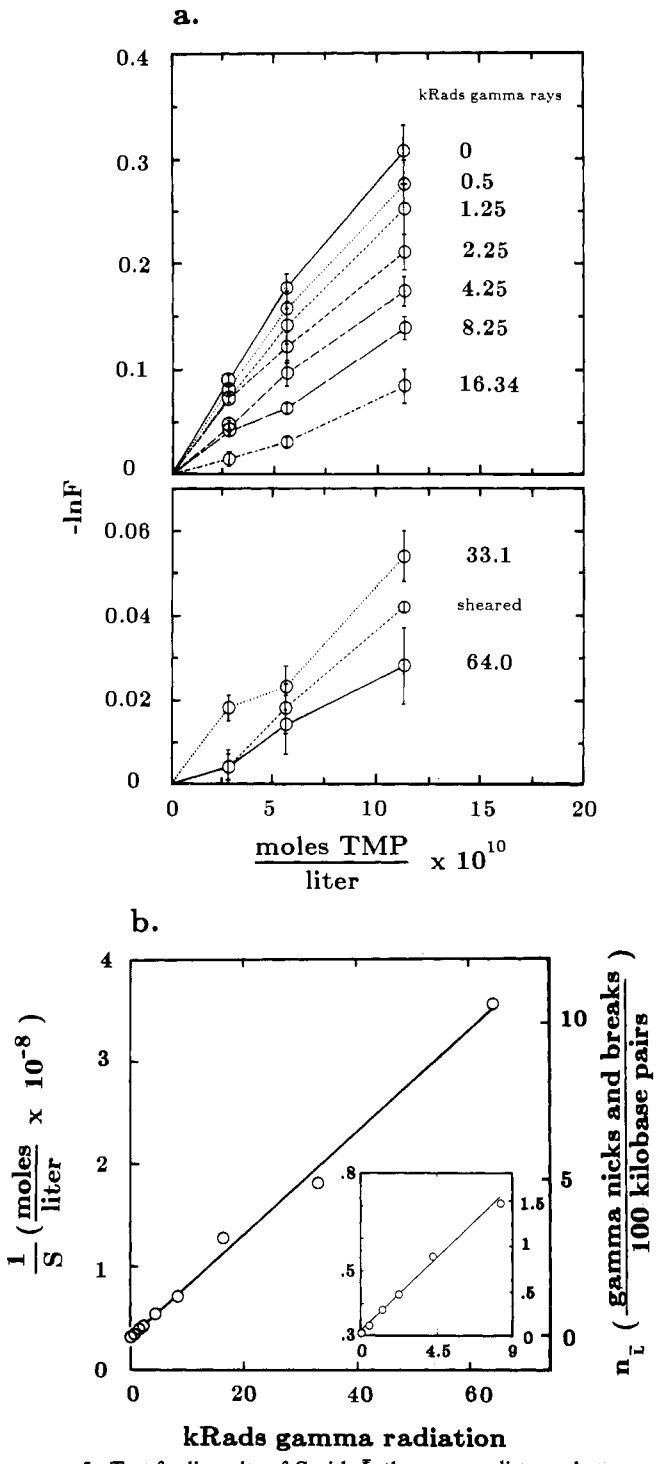

FIGURE 5: Test for linearity of S with $\bar{L}$, the average distance between breaks in either or both strands of the duplex. (a) Effect of γ radiation on $-\ln F$ of DNA containing TMP interstrand cross-links. (b) Reciprocal of the initial slope, $1/S$, shown as a function of DNA exposure to γ radiation. The right ordinate, n_L, was calculated from the relation $\bar{n}_L = 2/\bar{L}_n - 2/\bar{L}_0$, which gives the frequency of breaks due to irradiation. Purified *Drosophila* DNA (15 μg/mL) supplemented with the different concentrations of TMP shown on the abscissa was exposed to 360-nm light and then to the doses of γ radiation shown in the figure. The cross-linking profile for the DNA sample sheared to 19.5 kilobases but not exposed to γ radiation was used to determine the molecular weight of the unirradiated DNA (see text).

values from measurements in a very different kind of system supports the view that [PS] and cross-link frequency are linear with [P] in our experiments. The remaining data in this presentation were obtained with matched samples, so interstrand cross-linkage should depend only upon [TMP] at fixed light exposure.

We tested the experimental system and found cross-link number was first order with [P] to at least 1 cross-link per 1.5 kilobase pairs (Figure 2). At very low light dosage, more TMP would have been needed to get similar cross-linkage; binding site saturation could influence results at [TMP] 100-fold greater than in these experiments (Table I, column 5). The anomalous result with λ DNA (Figure 2) could conceivably result from a systematic error: Less than 5% of the incident light was absorbed in our samples, so effective dose of light is independent of those [TMP] used here. Moreover, such effects would be exaggerated in experiments with a shorter DNA requiring higher [TMP], yet with pUC19 DNA, $-\ln F/[P]$ remained constant to $-\ln F = 4$ (not shown). Quantum-chemical effects documented for laser pulses (Johnston et al., 1977) do not pertain at the fluences we used here. Since the phage purification included DNase treatment, we think infrequent nickage caused the curvature.

In the experiments described here, k ranged from 980 to 3770 L/(mol·base pair), although values below 1350 were obtained in four of the five experiments. The assay for cross-linkage was extremely reproducible for any sample (e.g., Figure 3), so variation in k is probably not due to measurement error. In Figure 3, S increased in proportion to $\bar{M}_w$ for mixtures composed of molecules of defined length in known proportions within 6% error. However, length of the DNA did not affect psoralen binding since λ DNA was cross-linked at similar efficiency both before and after restriction enzyme digestion. The variability could conceivably arise from undocumented fluence effects on intercalation as [P] varies during irradiation. However, the decomposition and photoaddition reactions occur at the same rate in all samples, preserving the relative [P] (Dall'Acqua et al., 1979b). Also, it is difficult to reconcile the data of Figures 3–5 with such a view; these data indicate that S was a linear function of $\bar{L}$ throughout the range of measurement. With the exception of the one unexplained very high value (Figure 2), the data indicate that variability in k was due to inaccurate measurement of $\bar{L}$ in the standard. Further data (not shown) indicate that k is aboout 1000 L/(mol·base pair) for sheared or restriction-digested DNA and is about 2000 L/(mol·base pair) for nicked DNA in dilute aqueous buffer exposed to near-UV as described here. Since k varied little within each experiment (Table II; Figure 5b), we conclude that our standard conditions were suitable for these tests.

Psoralen-catalyzed strand breakage would complicate our analysis. Psoralen plus UVA can generate singlet oxygen, which chemically activates psoralen tc bind to enzymes or peroxidize lipids (Grossweiner, 1984). On the other hand, Goyal and Grossweiner (1979) showed that DNA inhibits psoralen accessibility to oxygen. Also, Cohen et al. (1980) detected no TMP plus near-UV induced breaks by alkaline elution, a very sensitive method. Our data in Figures 2a, 4, and 5 show that such breaks occur less often than cross-links, if at all. In a more rigorous test, we exposed intact plasmid DNA in air- and psoralen-saturated water to 10 kJ/m^2 360-nm light. Superhlical DNA remained the same in the light-exposed samples as in unexposed controls (electrophoretic gels not shown). We conclude that oxidative effects were trivial in our experiments.

Summers and Szybalski (1967) used cross-link sensitivity for crude measurement of strand breakage in DNA. Initial slope analysis enabeled us to measure $\bar{L}$ with high sensitivity. Resolution was limited only by the accuracy of molecular weight measurement for the original sample. There were 1.9 $\times 10^{-3}$ nicks and breaks per kilobase pair per krad of γ ex-

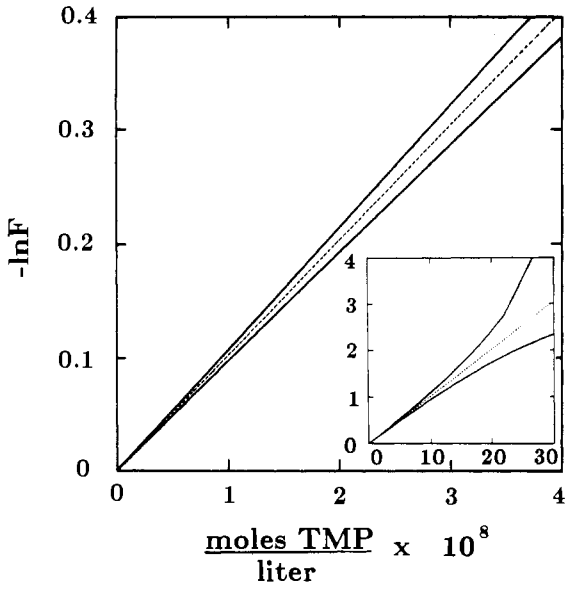

FIGURE 6: Effect of random 5% error in measurement of F on $-\ln F$. (Main figure) Data for $F > 0.7$. In this range, which we use for calculation of S, the error is small. (Inset) Data for $F > 0.02$. At F values below 0.13 ($-\ln F > 2$), there is a systematic overestimation of $-\ln F$ and molecular weight. The curve describing $-\ln F$ is replotted from Figure 3 for pUC19 DNA. The calculated curves are $-\ln [F + 0.05(1 - F)]$ for overestimation of F (lower curve) and $-\ln [F - 0.05(1 - F)]$ for underestimation of F (upper curve).

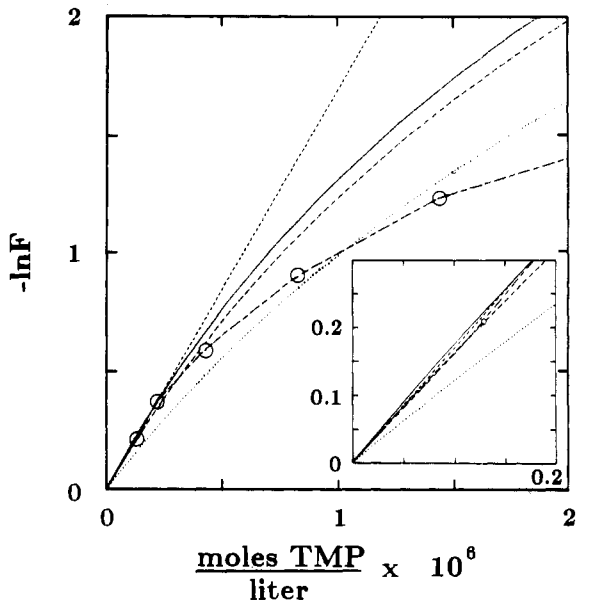

FIGURE 7: $-\ln F$ calculated from assumed molecular weight distributions is compared with experimental data and with initial slope. Initial slope analysis does not provide a molecular weight distribution but does provide useful information about the frequency of interstrand cross-links even in a DNA preparation of uncertain average molecular weight (see Discussion): (---) line connecting data points; (- - -) initial slope; (—) curve fit from assumed truncated normal distribution, $k\bar{M}_w$ calculated at $-\ln F = 0.3$; (···) similar curve, $k\bar{M}_w$ calculated at $-\ln F = 1.0$; (– –) similar curve, $k\bar{M}_w$ from initial slope. The equation used to calculate these curves is $F = [2/(2 + k[P]\bar{L})]^2$. This equation, derived originally by Becker et al. (1964), was arrived at independently by Jolley and Ormerod (1973) using similar assumptions. *D. melanogaster* DNA purified as in the experiment of Figure 4 and sheared to $\bar{M} = 50$ kilobases was cross-linked and tested for F as described under Materials and Methods.

posure over a very wide range of DNA exposures and hence of [TMP] (Figure 5). DNA cross-linked after breakage yielded no evidence for interaction of the radiation radicals with adducts in the DNA (not shown).

We had to choose a range of data to measure S. We calculated S for hypothetical mixtures of various uniform length DNA molecules, including one composed of equal parts of two DNA species of lengths L_1 and L_2 with $L_2 = 10L_1$, which had the shortest linear range of those tested but remained linear up to $-\ln F = 0.3$. We adopted the empirical approach of plotting our data to ascertain the linear range, which extended to at least 0.3 in our experiments, with DNA fragments free from nicks. Random fluctuations in sample volume were about 5% of each F value. Error in S will be less than this because several F values were used for each slope. Figure 6 shows that $-\ln F$ is progressively sensitive to 5% error in measurement of F. At F below 0.05, this measurement error causes systematic overestimation of $-\ln F$ and cross-link frequency. However, at $-\ln F < 0.3$, this error is small.

Previous attempts to determine k from measurements of cross-linkage in polydisperse DNA mixtures have generally involved some form of curve fitting. Most [e.g., Iyer and Szybalski (1963), Summers and Szybalski (1967), Akhtar et al. (1975), Ben-Hur et al. (1979), and Konopa et al. (1983)] have adopted the approximation that $-\ln F$ is the frequency of cross-links throughout the range of measurement, concluding that when 63% of cellular DNA is cross-linked, there is an average of one cross-link per DNA molecule and this number can be divided by the average molecular weight to obtain the cross-link frequency. This approximation underestimates cross-linkage in polydisperse DNA. Lawley and Brookes (1967) found that salmon DNA required 20-fold more alkylation with sulfur mustard to achieve D_{37} in the cross-linking reacton than bacterial DNA of comparable length but which had a narrower length distribution. Becker et al. (1964) assumed the distribution of fragment lengths to fit a normal curve centered on the number-average molecular weight. Jolley and Ormerod (1973) and later Fujiwara (1982) used this approximation to calculate k for nitrogen mustard and mitomycin C cross-linking, respectively.

Figure 7 shows typical data from an experiment with highly purified *Drosophila* DNA of about 50 kilobase pairs average length. The k for a normal distribution calculated at D_{37} provides the best fit to the data over the range of measurement shown and is about half k from the intial slope. The higher the value of $-\ln F$ chosen for the intersection, the greater the underestimation of k. The cross-linking curve calculated by assuming Gaussian length distribution with k obtained from S progressively overestimates $-\ln F$ at higher values of [P], presumably because of the skewed molecular weight distribution in the sheared sample. The plot in Figure 7 thus illustrates a central weakness of the curve-fitting approach: Although it has been generally agreed in the literature that cross-link frequency increases in linear proportion to the concentration of any bivalent agent, the cross-linking constant must be arrived at by successive approximation; it is not obvious which data within the accurate range of measurement may be used to calculate the theoretical curve. We show (see Theory) that for DNA of any length dispersion the initial rate of increase in cross-linked material should be linear below some concentration of cross-linking agent and that this range is generally suitable to calculate the cross-linking constant, k.

In summary, we have derived a mathematical relation to calculate the cross-linking constant from measurements of cross-linkage in psoralen plus light-treated DNA for which the weight-average molecular weight is known. Empirical tests show the relation is valid for measurement of random cross-link frequency or distance between interruptions over an extraordinarily wide range of strand lengths in DNA of any molecular weight distribution.

ACKNOWLEDGMENTS

We thank Karen Ringereide for help with data processing and for preparing the figures and Paul Howard-Flanders and D. M. Carter for comments on the manuscript.

Registry No. TMP, 3902-71-4; 8-MOP, 298-81-7.

REFERENCES

Akhtar, A. K. H., Begleiter, A., Johnson, D., Lown, J. W., McLaughlin, L., & Sim, S.-K. (1975) *Can. J. Chem. 53*, 2891–2905.

Becker, E. F., Jr., Zimmerman, B. K., & Geiduschek, E. P. (1964) *J. Mol. Biol. 8*, 377–391.

Ben-Hur, E., Prager, A., & Riklis, E. (1979) *Photochem. Photobiol. 29*, 921–924.

Charlesby, A. (1960) *Atomic Radiation and Polymers*, Pergamon, Elmsford, NY.

Cohen, L. F., Ewig, R. A. G., Kohn, K. W., & Glaubiger, D. (1980) *Biochim. Biophys. Acta 610*, 56–63.

Cole, R. S. (1970) *Biochim. Biophys. Acta 217*, 30–39.

Dall'Acqua, F., Trbojevich, M., Marciani, S., Vedaldi, D., & Recher, M. (1978) *Chem.-Biol. Interact. 21*, 103–112.

Dall'Acqua, F., Vedaldi, D., Bordin, F., & Rodighiero, G. (1979a) *J. Invest. Dermatol. 73*, 191–197.

Dall'Acqua, F., Marciani Magno, S., Zambon, F., & Rodighiero, G. (1979b) *Photochem. Photobiol. 29*, 489–495.

Fujiwara, Y., (1982) in *DNA Repair, a Laboratory Manual of Research Procedures* (Freidberg, E. C., & Hanawalt, P. C., Eds.) Vol. 2, Marcel Dekker, New York.

Garger, S. J., Griffith, O. M., & Grill, L. K. (1983) *Biochem. Biophys. Res. Commun. 117*, 835–842.

Goyal, G. C., & Grossweiner, L. I. (1979) *Photochem. Photobiol. 29*, 847–850.

Grossweiner, L. I. (1984) *Natl. Cancer Inst. Monogr. 66*, 47–54.

Haight, F. A. (1967) *Handbook of the Poisson Distribution*, ORSA Handbook, Vol. 11, Wiley, New York.

Harrington, R. E., & Zimm, B. H. (1965) *J. Phys. Chem. 69*, 161–175.

Hütterman, J., Köhnlein, W., Tēoule, R., & Bertinchamps, A. J., Eds. (1978) *Effects of Ionizing Radiation on DNA*, Springer-Verlag, New York.

Hyde, J. E., & Hearst, J. E. (1978) *Biochemistry 17*, 1251–1257.

Isaacs, S. T., Che-Kun, J. S., Hearst, J. E., & Rapaport, H. (1977) *Biochemistry 16*, 1058–1064.

Iyer, V. N., & Szybalski, W. (1963) *Proc. Natl. Acad. Sci. U.S.A. 50*, 355–362.

Johnston, B. H., Johnson, M. S., Moore, C. B., & Hearst, J. E. (1977) *Science (Washington, D.C.) 197*, 906–908.

Jolley, G. M., & Ormerod, M. G. (1973) *Biochim. Biophys. Acta 398*, 242–251.

Kanne, K., Straub, K., Hearst, J. E., & Rapaport, H. (1982) *J. Am. Chem. Soc. 104*, 6754–6764.

Konopa, J., Pawlak, J. W., & Pawlak, K. (1983) *Chem.-Biol. Interact. 43*, 175–197.

Lawley, P. D., & Brookes, P. (1967) *J. Mol. Biol. 25*, 143–160.

Mandell, J. D., & Hershey, A. D. (1960) *Anal. Biochem. 1*, 66–77.

Maniatis, T., Foitsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

McDonnell, M. W., Simon, M. N., & Studier, F. W. (1977) *J. Mol. Biol. 110*, 119–146.

Norrander, J., Kempe, T., & Messing, J. (1983) *Gene 26*, 101–106.

Reynolds, R. J. (1978) *Mutat. Res. 50*, 43–56.

Rodighiero, G., & Dall'Acqua, F. (1984) *Natl. Cancer Inst. Monogr. 66*, 31–40.

Ross, P., & Howard-Flanders, P. (1977) *J. Mol. Biol. 117*, 137–158.

Rubenstein, I., Thomas, C. A., Jr., & Hershey, A. D. (1961) *Proc. Natl. Acad. Sci. U.S.A. 47*, 1113–1122.

Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., & Petersen, G. B. (1982) *J. Mol. Biol. 162*, 729–773.

Song, P.-S., & Tapley, K. J., Jr. (1979) *Photochem. Photobiol. 29*, 1177–1197.

Straub, K., Kanne, D., Hearst, J. F., & Rapaport, H. (1981) *J. Am. Chem. Soc. 103*, 2347–2355.

Summers, W., & Szybalski, W. (1967) *J. Mol. Biol. 26*, 107–123.

Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Wiley, New York.

Wetmur, J., & Davidson, N. (1968) *J. Mol. Biol. 31*, 349–370.