CHARACTERIZATION OF PSORALEN-DNA PHOTOADDUCTS AT THE NUCLEOTIDE LEVEL IN pBR322 and SV40

by

Elaine A. Ostrander

A DISSERTATION

Presented to the Department of Microbiology and Immunology
and the Oregon Health Sciences University
School of Medicine
in partial fulfillment of
the requirements for the degreee of

Doctor of Philosophy

May 1987

APPROVED:

(Professor in Charge of Thesis)

(Chairman, Graduate Council)

If the history of protein crystallography is any guide, the present simplicity in DNA structure that comes mainly from ignorance will shortly be replaced by a bewildering complexity of new data, before it ultimately settles down again into the simplicity that means that we truly understand matters.

Richard Dickerson

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AGMK	African green monkey kidney
AMT	4'-aminomethyl-4,5,'8-trimethylpsoralen
bp	base pair (s)
dA	deoxyadenine
dG	deoxyguanine
DMS	dimethyl sulfate
D. melanogaster	*Drosophila melanogaster*
E. coli	*Escherichia coli*
HMT	4'-hydroxymethyl-4,5',8-trimethylpsoralen
LTR	long terminal repeat
mRNA	messenger RNA
5-MIP	5-methylisopsoralen or 5-methylangelicin
8-MOP	8-methoxypsoralen
nm	nanometer
PUVA	psoralen/long wavelenth ultraviolet light
rRNA	ribosomal RNA
SV40	Simian Virus 40
TAg	large tumor antigen
TMP	4,5'8-trimethylpsoralen
tRNA	transfer RNA
UV	ultraviolet light
UVA	long-wavelength ultraviolet light (320-400nm)

ACKNOWLEDGEMENTS

I would like to express my appreciation to the many individuals in Lesley Hallick's lab who contributed so much to this project in terms of ideas, time, and technical expertise, in particular to Gordon Robinson who guided me though my first tentative encounters with DNA.

I am indebted to many of the faculty in the Microbiology and Genetics Departments, in particular Jules Hallum, Jorge Crosa, Katie Richardson, Mike Litt, and Robert Koler for advice freely given, but far too often ignored.

My special thanks to Chuck Faust, for teaching me the difference between patience and persistence, and to Gary Wiesehahn and Steve Isaacs for showing me the value of excellence.

To my colleagues and friends Sophia Kondoleon and Cheryl Maslen, I owe a special debt of gratitude which I can never adequately repay. Their enthusiasm, humor, and friendship has been a sustaining force during this endeavor.

Finally, to my thesis advisor, Lesley Hallick, more than anything else I thank you for sharing with me the joy of science, and yes, I really would do it all over again.

ABSTRACT

Psoralens are tricyclic planar hydrocarbons that can intercalate into double stranded DNA, and then undergo photocycloaddition which results in the formation of a covalent adduct between the psoralen and a pyrimidine of DNA. Because they can easily pass through cellular and viral membranes without disruption of the biological entity, and because they bind preferentially to exposed regions of DNA, psoralens have become an indispensable tool for studying many types of nucleoprotein complexes.

In this thesis, psoralen-DNA interactions have been characterized at the nucleotide level by taking advantage of the inhibitory nature that both psoralen monoadducts and crosslinks have on restriction enzymes and exonucleases. Initially, the site specificity of 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT) and 5-methylisopsoralen (5-MIP) were investigated by fluorescent densitometry to examine the frequency of cleavage by the restriction enzyme Bgl I on photoreacted pBR322. This enzyme has an 11 base pair recognition sequence which varies from site to site due to a five base pair neutral region within the recognition sequence. When conditions were altered to favor the formation of particular photoadducts, a pattern emerged which suggested that the placement of psoralen adducts on purified DNA was not random, but rather the type of psoralen derivative used, as well as sequences surrounding the photoaddition site, played a role in determining adduct position.

Findings from these experiments were used to design a fine structure mapping technique that allowed the localization of 5-MIP adducts to be made precisely at the nucleotide level. This approach takes advantage of the fact that 5-MIP adducts inhibit the processive nature of λ exonuclease, a 5' exonuclease required for general recombination in the bacteriophage λ. This technique has been used to probe psoralen accessibility patterns in the SV40 origin of replication and surrounding transcription regulatory elements. Data from these experiments pinpoint a region of about 180 base pairs which is hypersensitive for psoralen photoaddition. This region includes the 72 base pair repeats which function as enhancers of viral transcription, the major late promoter and mRNA start sites, and the major late DNase I hypersensitive sites. In vivo analysis of intracellular minichromosome complexes 60 hours post-infection and of packaged virus demonstrates this same pattern. These data suggest that a sequence-directed structural alteration exits within and to the late side of the SV40 regulatory region which has been previously undetected.

I. LITERATURE REVIEW

A. Psoralen Photochemistry

1. Introduction

Psoralens are three ring heterocycylic compounds formed by the linear fusion of a furan ring and a coumarin (see figure 1). They are found naturally in a variety of plants, most abundantly the Umbelliferae (celery and carrot), Rutacea (citrus fruits), and Leguminosae (peanuts) families where they are believed to act as natural insecticides and antifungal agents (Mustafa et al., 1967; Barenbaum et al., 1978). Their clinical applications have been appreciated in modern medicine since 1948 when 8-methoxypsoralen was used in conjunction with long wavelength ultraviolet light (UVA) to treat the pigment disorder vitiligo (El Mofty et al., 1948). They are currently used to treat a variety of dermatologic abnormalities including psoriasis, vitiligo, and mycosis fungoides (Pathak et al., 1969; Parrish et al., 1974; Perone et al., 1972; Honigsmann et al., 1967).

Because psoralens can readily pass through cellular and nuclear membranes as well as viral capsids without disruption of the biological entity, they are probably the most important class of photochemical reagents for the investigation of nucleic acid structure and function. They have been applied to the study of RNA nucleoprotein complexes, chromatin, DNA torsional tension, and secondary structure of viral RNA genomes; they have also been used as reagents in DNA repair studies (for review see Cimino et al., 1985). Like ethidium bromide, psoralens have the ability to intercalate between the stacked bases of double-stranded

nucleic acids where they then undergo a photochemical reaction resulting in the formation of a covalent bond between the psoralen and nearby pyrimidine of DNA (Musajo et al., 1966; Krauch et al., 1967; Pathak et al., 1969; Cole, 1970). Isaacs et al. (1977) proposed the following mechanism for the reaction of psoralen with double-stranded nucleic acids (see also figure 2).

$$P + S \rightleftharpoons PS$$

$$PS + h\nu \rightarrow A$$

$$A + h\nu \rightarrow X$$

$$P + h\nu \rightarrow B$$

In this reaction P is psoralen molecule which intercalates into the available site (S) to form a noncovalent intermediate termed PS. In the presence of long wavelength UV light (hv, 320-380 nm), either the 4'5' (furan) or the 3,4 (pyrone) double bond of the psoralen can react with the 5,6 double bond on the pyrimidine to form a cyclobutane ring which characterizes the covalent adduct (A) (Musajo et al., 1966; Krauch et al., 1967; Musajo et al., 1965). This reaction usually occurs with thymine in DNA and uracil in RNA, although a minor reaction with cytosine also occurs (Straub et al., 1981; Kanne et al., 1982a; Kanne et al., 1982b). There is a single report of 8-methyoxypsoralen (8-MOP) forming adducts with adenine but this reaction is generally considered to be quite minor (Ou, 1978a). If two pyrimidines are located in adjacent positions on opposite DNA strands, absorption of a second photon of ultraviolet light can convert a psoralen monoadduct into a diadduct (X), thus crosslinking the double helix (see figure 3) (Dall'Acqua et al.,

1970; Cole, 1971; Dall'Acqua et al., 1972). Additional UV light converts the unintercalated psoralen to breakdown products (B). Psoralen reactivity with both protein (Yoshikawa et al., 1979; Lerman et al., 1980; Veronese et al., 1981; Granger et al., 1982) and lipid (Kittler et al., 1983) occurs only incidentally.

Photoreversal of psoralen adducts from DNA has been demonstrated and this technique has been used extensively in the analysis of RNA secondary structure (Thompson et al., 1981; Thompson et al., 1983; Bachellerie et al., 1982; Turner et al., 1982; Thompson et al., 1982). The photochemistry of the reaction has been most recently characterized by Cimino et al. (1986) who examined the photoreversal of HMT crosslinks formed between two thymines on opposite strands of a DNA oligomer, as well as from an isolated T-HMT-T diadduct. They found that both types of crosslinks were photoreversed by exposure to 240-313 nm light, but that preferential photoreversal occurred at the furan end of the psoralen in the T-HMT-T diadduct, but at the pyrone end in the crosslinked oligonucleotides.

A class of compounds closely related to the psoralens are the isopsoralens or angelicins which are formed by the angular fusion of a furan and a coumarin. Due to misalignment between the reactive double bonds of the isopsoralen and the pyrimidine, subsequent reaction of this molecule with the remaining double bond of a pyrimidine on the opposite strand is not possible, hence these derivatives do not form crosslinks with DNA in dilute aqueous solutions (see figure 1) (Bordin et al., 1979a; Bordin et al., 1979b). There are several reports, however, which

suggest that they can crosslink certain types of folded DNA, such as that found within the head of bacteriophage lambda (Lown et al., 1978; Kittler et al., 1980).

2. Mechanism of the Reaction.

The mechanism of psoralen binding to DNA is fairly controversial and has been studied intensely for a number of years. The molecule contains both singlet (two electrons located in separate orbitals having opposite spin) and triplet (two electrons located in separate orbitals with the same spin) excitation states. Based on both theoretical and spectroscopic evidence, Song et al. (1971) suggested that the triplet excitation state which was substantially localized in the 3,4 pyrone double bond, rather then the singlet excitation state which was primarily located in the 4',5'furan double bond, would be the most reactive. Thus, the vast majority of adducts formed should be pyrone side monoadducts. Using NMR spectroscopy however, Kanne et al. (1982a) showed that furan side adducts actually predominate in DNA reacted with 4,5',8-trimethylpsoralen (TMP), 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT), 8-MOP, and psoralen. The contradiction is resolved in part by the data of Beaumont et al. (1979) who measured the rate constant for the reaction of triplet states of 8-MOP and TMP using a laser flash photolysis technique. These psoralen derivatives are among the most reactive in vivo and for that reason 8-methoxypsoralen (8-MOP), which does not cause erythema as does TMP, is used in the PUVA program to treat skin disorders, rather than the less effective psoralen. Beaumont's

group found no correlation between rate of reactivation of the triplet states of the psoralens and their actual reactivity with DNA. In fact, the rate of reaction for the triplet state of psoralen was approximately 400 times greater than that for trimethylpsoralen, and 140 times greater than the rate of 8-MOP which is highly reactive in vivo. They suggest that steric considerations between the psoralen molecule and the DNA helix are actually the critical parameters in determining psoralen reactivity.

It was also believed for some time that pyrone side monoadducts were the precursors to crosslink formation (Matulin et al., 1973). Many investigators argued, however, that since crosslinks are formed by irradiation at 320-380nm, the second photoreaction must occur from a species capable of absorbing light in this range (Lown et al., 1978; Ou et al., 1978b). The reported spectrum of a pyrone side monoadduct shows little or no absorbance at wavelengths longer than 300 nm, while the absorption spectrum of a 4'5' monoadduct is more intense than that of psoralen; hence furan side adducts would be expected to undergo a second photoreaction to form crosslinks (Lown et al., 1978). Eventually, Kanne et al. (1982) showed that only furan side adducts can go on to form crosslinks. They propose that the 4'5' photoadduct was formed via the singlet state after absorption of the first photon, and that subsequent absorption of a second photon by the photoadduct could lead to crosslinking via the triplet state.

3. Controlling The Photoreaction.

Considerable progress has been made in controlling the reaction by which psoralen binds DNA, and hence the type and proportion of adduct formed. Dozens of psoralen derivatives are currently available on a commercial basis, many of which represent modifications of parent compounds that are either more reactive or more soluble than the parent, or favor the formation of a particular type of adduct. Overall, methylation tends to increase the reactivity of both psoralens and isopsoralens (Isaacs et al., 1982). Also, the presence or absence of a methyl group at position 4 has been shown to have a critical role in controlling the proportion of pyrone-side monoadducts (Kanne et al., 1984). Psoralens that contain a 4 methyl group, such as the HMT used in this thesis, generally form a low percentage of pyrone side adducts. This has been attributed to steric interference between the methyl group of the psoralen and the 5 methyl group of thymine with which the psoralen is predominantly reactive.

Johnston et al. (1977) postulated that two approaches were possible for exclusive production of monoadducts. One approach was to use a very low light dosage during the irradiation, thus making it unlikely that any single psoralen molecule would receive the two successive photons necessary for crosslink formation. The difficulty with this approach is that the final concentration of monoadducts on the DNA is apt to be quite low. The other approach is to present the light in a short pulse so that there is not enough time for a second photon to be absorbed. Johnston et al. demonstrated the feasibility of this approach using the psoralen

derivative 4'-aminomethyl-4,5',8-trimethylpsoralen (AMT), and a 15 nanosecond light pulse directed from an ultraviolet 347 nm laser.

A second method was proposed shortly thereafter by Chaterjee and Cantor (1978) who were concerned that the number of total monoadducts was still relatively low using the laser method, and that the performance of experiments was dependent on the availability of a laser. They demonstrated that the production of 4'5' monoadducts could be significantly enhanced by irradiation with wavelengths above 380nm. In the experiments described in manuscript 1, modifications of both methods were used to generate molecules with less than 5% crosslinks. The data summarized in the manuscript was compiled from data generated by using the Chaterjee and Cantor method.

4. Psoralens as Probes of Nucleic Acid Structure.

a. Mapping RNA Adducts.

Because of their crosslinking abilities psoralens have been used successfully to study the secondary structure of many types of RNA (for review see Cimino et al., 1985). Initially Wollenzien et al. (1978) used HMT in conjunction with electron microscopy to examine the frequency of hairpin structures in Drosophila melanogaster 26S and 16S rRNA, following irradiation and denaturation. This type of approach is basically qualitative. The first sequence resolution of psoralen crosslinks in RNA was reported the following year by Rabin and Crothers (1979). Their method involved treating E. coli 5S rRNA with AMT, and then digesting with RNase T_1. This was followed by two dimensional gel

electrophoresis which revealed T_1 resistant fragments not found in the untreated samples. The putative crosslinks were then photoreversed with 250nm UV, and the photoreversal product was isolated and characterized by conventional RNA sequencing methods. Using a modification of this technique, Thompson and Hearst (1983) have been able to localize a large number of psoralen crosslinks in E. coli 16s rRNA to within 15 bp of their actual binding site. Their results not only support the generally accepted models for 16S rRNA structure, but also demonstrate the dynamic nature of rRNA by elucidating many long-range interactions which appear to exist in dynamic equilibrium with local secondary interactions. The current thinking is that these equilibria probably reflect conformational changes which occur in RNA in the course of the normal functioning of ribosomes (Cimino et al., 1985). The most precise method for analyzing rRNA was devised by Youvan et al. (1982). In this approach, 16s rRNA is irradiated with psoralen, and then primed with a 3'-proximal restriction endonuclease-generated fragment. This substrate is then copied with reverse transcriptase which is inhibited by psoralen crosslinks. When analyzed on a gel, the abbreviated cDNAs indicate the positions of the psoralen diadducts.

Fine structure analysis of psoralen photoadducts was used to study the structure of tRNA as well. Garret-Wheeler et al. (1984) used an analysis similar to the one of Rabin and Crothers to analyze yeast $tRNA^{phe}$. Most recently, Hui et al. (1985) have outlined the simplest and most reproducible method reported to date. In their approach, end-labeled DNA that overlaps the approximate position of a crosslink is

hybridized to RNA and then treated with mung bean nuclease. It is assumed that the psoralen will crosslink areas of RNA secondary structure, and that the overlapping oligonucleotide will be unable to hybridize to the double-stranded RNA at this point. The resulting "bubble" is particularly sensitive to digestion by mung bean endonuclease which preferentially digests single-stranded DNA. The resulting digest is then analyzed on a denaturing sequencing gel. Unfortunately this method requires prior knowledge of the approximate positioning of the psoralen and does not allow sequencing of large regions in a single experiment.

b. Mapping Psoralen Adducts in DNA.

A variety of attempts have been made to map psoralen adducts at both the restriction fragment and nucleotide levels in DNA. At the onset of this thesis project in 1981, psoralens could be localized on DNA at the restriction fragment level only, using a method described by Robinson and Hallick (1983). Briefly, this method involved performing in vivo irradiations with tritium-labeled psoralen, purifying the DNA, then digesting the DNA with multiple restriction enzymes. The resulting fragments were then separated by electrophoresis, and the percent tritium determined for each band. This data was compared to control DNA which was similarly digested and analyzed on the same gel, but was photoirradiated in vitro. The ratio of radioactivity in the corresponding fragments is indicative of the relative psoralen accessibility of that fragment in the experimental in vivo sample. As

will be described in the following section, this method was used to study the structure of intracellular SV40 minichromatin as well as packaged viral DNA.

Subsequently, Becker and Wang (1984) developed a method for detecting protein-DNA interactions utilizing UV light and suggested its application for mapping psoralen photoadducts. Their chemical approach exploits a common feature of photoreacted bases which allows the breaking of the DNA backbone only at the sites of photodamage. Basically, in many types of photoproducts one or more of the 5,6 double bonds of a pyrimidine becomes saturated during the photoreaction. Since carbonyl groups conjugated to double bonds are poorly reduced by $NaBH_4$, the carbonyl group at position four of photochemically modified bases is selectively reduced, resulting in opening of the thymine ring. The glycosidic bond of the photochemically modified (and now reduced) thymine is much more labile to acid hydrolysis than unmodified thymines. Once the glycosidic bond is hydrolyzed by acid, the phosphate backbone can be broken at the sugar ring by subsequent treatment with alkali. Whether this method is actually applicable to sequencing psoralen crosslinks and/or monoadducts is not yet clear.

To study the interaction between the DNA replication apparatus and DNA lesions, Piette and Hearst (1983) developed a limited system that enabled them to identify the position of HMT crosslinks *in vitro*. They used a double-stranded circular DNA template ϕX174 which contains a single nick at a specific site. The DNA is reacted with HMT, and then used as a substrate for nick-translation with *E. coli* DNA polymerase I

holoenzyme. By using the dideoxy chain-terminating sequencing procedure, they found it was possible to map the termination sites on the photoreacted template, which occur at nucleotides preceding psoralen crosslinking sites. By repeating the experiment with 5-MIP, they further determined that DNA polymerase I is not stopped by psoralen monoadducts.

Interestingly, in the preceding study Piette and Hearst found that 5'-TA and 5'-AT were equally good substrates for crosslink formation. This result is in conflict with crystallographic and solution studies (summarized by Sobell et al., 1982), which indicate that simple intercalating agents such as psoralens show a marked binding preference for 5'-CG and 5'-TA sequences in short self-complementary oligomers. This preference is believed to reflect the ease of unstacking pyrimidine-purine sequences vs. purine-pyrimidine sequences during intercalation. This idea is supported by a 1984 report from Gamper et al. in which the efficiency of HMT adduct formation was examined for Kpn I and Bam HI linkers. The double-stranded Kpn I linker (CGGTACCG) was crosslinked 100 fold more efficiently by HMT than the double-stranded Bam HI linker (CGGATCCG). Furthermore, when the differential reactivity of HMT towards the two sequences was observed in SV40 DNA, the Kpn I reaction site was still more susceptible to HMT modification than the Bam HI site by 10 fold. The reduction in preference is believed to be related to sequences surrounding the intercalation sites.

Recently, Zhen et al. (1986) also concluded that 5'-TA dinucleotides are more reactive with crosslinking agents than 5'-AT sequences. Their analysis is based on mapping the positions of Bal 31 exonuclease

inhibition on DNA crosslinked with HMT. They argue that base tilt angles obtained from X-ray crystallographic studies (Dickerson and Drew, 1981) show that 5'-TA dinucleotides are opened approximately 8° toward the minor groove, while 3'-TA sequences are opened approximately 8° toward the major groove. Since psoralen intercalation leading to the proper orientation is believed to take place with the psoralen hydrophobic edge entering from the minor groove, intercalation at 5'-TA sites should be favored over intercalation at 5'-AT sites (Hansen et al., 1985). Curiously, Zhen's data showed that only 70% of a total of 20 5'-TA dinucleotides within the analyzed DNA fragment appeared to constitute efficient crosslinking sites. They speculate that the influence of more distant bases on the configuration of the TA base pair, in terms of both base pair twist and tilt angle, may be responsible for this difference in efficiency. Their results do not, however, suggest any obvious 3 or 4 bp consensus sequence which plays a role in determing the efficiency of crosslinking. Interpretation of this data is made difficult by the fact that Bal 31 is not inhibited by psoralen monoadducts, consequently their data reflects a preference for HMT crosslinking and does not provide information about preferences for intercalation or monoadduct formation.

In this thesis, I have taken an enzymatic approach to mapping psoralen photoadducts on restriction fragments spanning the SV40 regulatory region (see manuscript 2). This approach, which readily lends itself to <u>in vivo</u> applications, takes advantage of several features of λ exonuclease, an enzyme required for general genetic recombination in the bacteriophage λ (Shulman et al., 1970; Little et al., 1967a). It is a

processive exonuclease which cleaves 5'-mononucleotides from the 5' phosphoryl termini of double stranded DNA but not RNA (Little et al., 1967b; Thomas et al., 1978; Carter et al., 1971). It was chosen for these experiments because it preferentially initiates at 5' recessed ends and not at nicks or gaps (Little et al., 1967b; Carter et al., 1971, Masumune et al., 1971; Sriprakas et al., 1975).

5. Psoralens as Probes of Chromatin Structure

Eucaryotic DNA is typically composed of many superimposed levels of nucleoprotein organization, all of which play a role in condensing the DNA in the nucleus (for review see Eissenberg et al., 1985). At its simplest level, chromatin appears as a series of "beads on a string" in which the DNA is condensed approximately 6 fold by the generation of nucleosomes. Each nucleosome consists of two molecules each of the basic histones H2A, H2B, H3 and H4, around which 140-160 bp of DNA is wrapped twice. The nucleosome beads are connected by interbead "linker regions" that vary in length from a few base pairs (bp) to as many as 70. In addition, one molecule of histone H1 is typically bound to DNA sequences entering and leaving the nucleosome and plays a key role in stabilization of the structure.

The characteristic structure of chromatin was deduced, in part, by studies with micrococcal nuclease which primarily digests the unconstrained DNA located in the linker region (Noll, 1974). Unfortunately, the type of information obtained from these types of experiments is limited for multiple reasons. First, the native chromatin

is disrupted during the reaction and its relationship to the structure of the isolated monosome is unclear. Second, the enzymatic cleavage itself may induce nucleosome migration. Finally, because of the absolute requirement of micrococcal nuclease for Ca^{2+}, the question of the effect of divalent cation concentration on native chromatin structure cannot be addressed (Wiesehahn et al., 1977).

Hansen et al. (1976) first proposed that psoralens be used as probes of chromatin structure for the study of Drosophila embryo nuclei. In these experiments, nuclei were photoreacted with TMP, the DNA was isolated and denatured using heat and formaldehyde, and then analyzed by electron microscopy. Tightly crosslinked regions appeared double-stranded while regions protected from crosslinking appeared as open loops of single-stranded arms. The use of this technique on heavily irradiated chromatin revealed protected regions which occurred in tandem and had a basic length of 160 to 200 pairs. The basic structure of chromatin had been deduced previously using electron microscopy to examine chromatin extended by low ionic strength, as well as from sizing DNA fragments resistant to in situ nuclease digestion. The periodicity predicted from these studies ranged from 175 to 200 bp, which was virtually identical to the conclusions reached by Hansen et al. (1976).

These results were extended to a more detailed study of Drosophila melanogaster nuclei (Wiesehahn, 1977). In these experiments, nuclei were photoreacted to different extents with ^{3}H-TMP and subsequently treated with micrococcal nuclease. The specific activity of the DNA which was resistant to digestion and that of undigested DNA were then analyzed.

The level of radioactivity in the digested sample was much lower than that of the undigested sample, indicating that the nuclease preferentially digested the ^{3}H-TMP-containing regions of the DNA, strongly suggesting that psoralens primarily reacted with linker DNA.

Hallick et al. (1980) used the same derivative to probe the organization of bacterial DNA in situ. This analysis was motivated by a previous report from Griffith (1976) that lysis of E. coli directly onto electron microscope grids revealed a chromosome structure that resembled the regular beaded structure of eucaryotic chromatin. In addition, several reports indicated that a low molecular weight basic DNA-binding protein of E. coli, termed HU, had a beaded appearance in glutaraldehyde-fixed preparations visualized in the electron microscope which resembled the minichromosome appearance of SV40 (Germond et al., 1975). Hallick showed, however, that E. coli DNA was not protected in vivo from photochemical addition of TMP. In saturation experiments with multiple additions of TMP, the final level of ^{3}H-TMP bound to the bacterial DNA was indistinguishable from the level obtained with purified DNA. In contrast, eucaryotic DNA is protected from photoaddition at a level approximately 10 times higher than that seen with naked DNA. This indicates that E. coli DNA is not packaged into conventional nucleosomes.

Using multiple approaches, including psoralen photochemistry, considerable progress has been made in understanding the chromatin structure of many eucaryotic systems. Because SV40 is small, easy to grow in large quantities, and has a "beads on a string structure" which mimics that of more complex systems, SV40 is the system of choice in many

laboratories for studying chromatin structure. Out of such studies a wealth of information regarding the structure of both the DNA helix and chromatin have been obtained. Since this virus was the biological system of choice for this thesis, and information regarding its life cycle, nucleoprotein structure and transcriptional regulation is relevant to findings presented herein, a discussion of these topics follows.

B. SV40

1. Introduction

Simian Virus 40 (SV40), together with polyoma virus, and JC and BK viruses, belongs to a group of small DNA tumor viruses known as the papovaviruses. SV40 was first described by Sweet and Hilleman (1960) as a noncytopathic contaminant of rhesus monkey cells used in the preparation of human poliomyelitis vaccine. Interest in the virus heightened quickly thereafter when it was shown that large amounts of SV40 would induce tumors when inoculated into neonatal hamsters (Eddy et al., 1962; Girardi et al., 1962). In susceptible or "permissive" cells, such as monkey cells, the virus undergoes a lytic infection which leads to the production of new infectious virus particles and eventually cell death. Alternatively, the virus can stably transform "nonpermissive" cells at low levels; the virus continues to divide under conditions which inhibit normal cell division. SV40 can also transform a variety of rodent and primate cells, which may then induce tumors when transplanted into susceptible rodents.

Although the initial impetus for studying SV40 was to use it as a model for tumor viruses, considerable information about the organization and regulation of eucaryotic cells has resulted from intensive studies of this virus family over the last 25 years. Because the viral genome is small and only encodes 6 genes, it must rely a great deal on eucaryotic host machinery for both replication and transcription. This has led to the productive use of SV40 in studies of replication, transcription, mRNA processing, and promoter and enhancer functions.

2. Virus Description and Life Cycle

The SV40 virion is an icosahedral particle with a diameter of about 45nm (Tooze et al., 1981), whose DNA sequence has been published (Fiers et al., 1978; Reddy et al., 1978). It has a double-stranded DNA genome of 5243 base pairs with the Bgl I site at the origin of replication assigned a nucleotide number of 1.

Immediately following infection of a permissive host, the small (t) and large tumor antigen (TAg) encoded on the "early" side of the viral genome are expressed for about 96 hours (Tegtmeyer et al., 1975; Prives et al., 1978). Transcription of these genes by eucaryotic RNA polymerase II initiates from sites very close to the viral origin of replication. During this early phase of the lytic cycle, replication does not occur, (Tegtmeyer, 1972) and expression of the viral capsid proteins occurs only at low levels (Cowan et al., 1973). When the intracellular concentration of the large TAg reaches a certain level, usually at about 12 hours post infection, viral replication is initiated (Graessmann et al., 1978).

Following the onset of replication, the transcription of early genes is gradually reduced, while the "late" genes are turned on. The late genes encode the VP1, VP2 and VP3 proteins which make up the viral coat protein. Additionally, they code for a small protein known as the agnoprotein which plays a key role in viral infectivity.

The entire viral life cycle is elegantly orchestrated by a 400 bp regulatory region that bridges the sequences responsible for early and late transcription (See figure 4). Although these sequences will be discussed in considerably more detail in sections 3 and 4, it is important to realize this region includes the origin of replication, the initiation sites and promoters for both early and late transcription, and the two 72 bp repeats which have been identified as enhancers of transcription. Also in this region are the three T antigen binding sites I, II and III which have key roles in both replication and transcription. The viral regulatory sequences are particularly intriguing since at multiple stages of the viral life cycle a subpopulation of molecules has been identified which displays an altered chromatin structure encompassing this region.

3. Replication of SV40

Replication of the virus begins approximately 12 hours post-infection at a unique site (which coincides precisely with the Bgl I site) and proceeds bidirectionally (Kelly and Nathans, 1977; Li and Kelly, 1984)). Termination occurs when the two forks collide, without any apparent sequence specificity. Using temperature sensitive mutants,

early experiments by Tegtmeyer showed that SV40 large TAg is required for initiating new rounds of viral replication, but not to complete a cycle which has already begun, since African Green Monkey Kidney (AGMK) cells infected with TAg mutants at the restrictive temperature produced no detectable infectious DNA (Tegtmeyer, 1972). It has been shown that the regulation of replication is controlled, in part, by the binding of TAg to DNA sequences near the origin of replication (Myers and Tjian, 1980). In these experiments a 311 bp fragment with the viral origin of replication and surrounding sequences was inserted into a bacterial plasmid. A series of deletion mutants was then generated in the viral origin region of this plasmid using in vitro mutagenesis. When replication of mutant DNAs was compared with their ability to bind purified TAg, it was clear that the mutant DNAs which were deficient in ability to bind TAg also exhibited reduced levels of replication. This suggests that a direct interaction between TAg and DNA sequences at the origin of replication is necessary for efficient initiation of replication.

The exact binding sites of TAg which are necessary for replication have been deduced by a number of methylation protection and footprinting experiments (for example see Myers et al., 1981b) These studies show that there are three distinct sites within the SV40 regulatory region to which TAg binds with varying affinity. Affinity for site I, which extends approximately 30 bp from nucleotide 5185 to nucleotide 5215 is the highest. Affinity for site II, which extends from the boundary of site I at nucleotide 5215 through nucleotide 15 is considerably less.

This region contains the early RNA start sites and a 17 bp AT-rich promoter. TAg has the lowest affinity for site III which falls on the late side of transcription and extends to approximately nucleotide 39. Dimethyl Sulfate (DMS) footprinting was used to identify 13 pentanucleotide contact sites which are distributed approximately equally within these three primary binding sites, and have a consensus sequence of 5'-(G>T)(A>G)GGC-3' (DeLucia, 1983).

Located among the three TAg binding sites is a 65 bp segment between nucleotides 5208 and 30 termed the minimal origin of replication. This region contains a 27 base pair inverted repeat, the 17 base pair A/T rich subregion, and TAg binding site II (DiMaio and Nathans, 1982). Studies characterizing the minimal origin of replication have determined that binding of TAg to site II is absolutely required for replication while binding to site I is required for enhanced replication efficiency (Tegtmeyer, 1972; Shortle et al., 1979; DiMaio and Nathans, 1980; Stillman et al., 1985; DeLucia et al., 1986). Very recently, an SV40 in vitro replication system has been developed which shows that replication is dependent not only on TAg, but also the DNA polymerase α-primase complex, topoisomerase I, and a single-stranded DNA binding component (Wobbe et al., 1987).

4. SV40 Transcription

The regulatory region for SV40 transcription is about 420 bp long and is contained within the Hind III-Hpa II sites located between nucleotides 5171 and 346 (Tooze et al., 1981) (see figure 4). This

region is positioned between the coding sequences of two sets of genes that are transcribed divergently at early and late times post infection (for review see Das et al., 1985). Sequences corresponding to the 5' ends of early and late mRNAs are located within this region, as are the promoters for early and late transcription. SV40 early pre-mRNA has two discrete donor 5' splice sites and one common acceptor 3' splice site that are utilized to form the large TAg and small t mRNAs.

Myers et al. (1981b) have shown that TAg binding at the origin represses early RNA synthesis in vitro. They demonstrated that this repression does not occur by an attenuation mechanism where T antigen:DNA complexes simply block the forward movement of polymerase, but rather by a complex autoregulatory mechanism. Binding of TAg at site III is absolutely necessary to completely repress early transcription, since TAg binding at that position blocks the critical 21 bp repeat sequences of the early promoter. Since the affinity for site III is much less than that for sites I or II, site III is only bound when there is an excess of TAg and early transcription can be repressed.

The actual transcription of early sequences requires three spatially distinct components. These are the TATA box or Goldberg-Hogness box located in the 17 base pair A+T rich sequence, three copies of a G+C rich 21 base pair repeat which actually plays a role in both early and late promoter functions, and two perfect 72 bp repeats which comprise the transcriptional enhancers.

The three 21 bp repeats contain two each of G+C rich motifs having the consensus sequence GGGCGGRR (R = purine). Analysis of a series of

deletion mutants has lead Fromm and Berg to argue that any two of the six motifs are sufficient for early transcription (Fromm and Berg, 1982). Everett et al., (1981) have shown that a four-fold drop in transcription occurs under the same circumstances. Interestingly, this same study also demonstrated that the 21 bp sequences, like the 72 bp repeats, function bidirectionally. It is not surprising, then that these sequences play a role in late as well as early transcription.

The SV40 72 bp repeats were the first such sequences to be termed enhancers, which are operationally defined as cis-acting short stretches of DNA that can potentiate the transcription of a diverse set of genes in a manner independent of orientation (Gruss et al., 1981; Benoist et al., 1981). When one copy of the two repeats is deleted, the rate of transcription drops only slightly, whereas deletion of both completely abolishes transcription. Enhancers have since been identified in a variety of other viruses including JC virus, bovine papilloma virus, adenovirus, and the LTRs of many retroviruses (Fried et al., 1986). They also exist in many other well characterized eucaryotic systems such as the immunoglobulin heavy chain genes (Gillies et al., 1983; Banerji et al., 1983; Queen et al., 1983)

Four basic models have been proposed to account for the mechanism of action of enhancers (summarized by Courey et al., 1986). First, it was suggested that they might act as entry sites for transcription factors which later move to a promoter where they initiate transcription. This is supported by the general observation that the further an enhancer is from the promoter, the more inefficiently it functions. Second,

enhancers may act to organize chromatin into a transcriptionally active conformational state. This is supported by the work of Jongstra et al. (1984) who showed that SV40 enhancer sequences can change the pattern of DNAse I hypersensitivity. Third, they may function by targeting the transcribing complex to a nuclear matrix-associated region where other necessary transcription factors are located. And finally, a related suggestion, they may act as binding sites for topoisomerases. This is supported by the findings of Yang et al. (1985) who showed that topoisomerase II binding sites in SV40 are located very near the enhancer sequences.

Plon and Wang (1986) have recently published the results of an elegant set of experiments investigating the proposal that enhancers may act as binding sites for sequence-specific gyrase-like molecules. They constructed a novel heteroduplex molecule termed a "tailed circle", in which one of the strands contains an extra palindromic sequence which folds back on itself into a hairpin. The human β globin gene was placed in the circle, while the SV40 enhancer was centered in the hairpin. Since the globin gene and the enhancer are not topologically linked, any action by a topoisomerase on the enhancer sequences would not affect the topology of the globin gene. When this molecule was transfected into HeLa cells, they observed that the SV40 enhancer was still able to stimulate transcription of the globin gene, thus arguing that enhancers do not function as sequence-specific binding sites for gyrase.

Additional information on the mechanism of enhancer activity resulted from the work of Sassone-Corsi et al. (1985), who first

identified trans-acting factor(s) which might be responsible for SV40 enhancer activity *in vitro*. These factors were clearly unique from other factors which bind to the TATA box or the early promoters. By *in vitro* analysis, they showed that the enhancer and the trans-acting factor(s) appear to form a stable complex which requires both the 5' and 3' end of the enhancer sequences. This implies that either there is a single cellular binding factor which is very large, that there are multiple binding factors, or that the DNA sequence is distorted in some way during the binding process. DNAse I hypersensitivity experiments with HeLa cell extracts and enhancer sequences lead to the identification of multiple domains within the enhancer sequences to which trans-acting factors bind, thus suggesting that the mechanism of action involves the formation of a large nucleoprotein complex (Wildeman et al. 1986).

One final consideration are the findings of Nordheim and Rich (1983) who identified three 8-bp segments of alternating purine-pyrimidine tracts within the enhancer sequences which they propose assume a Z-DNA conformation on molecules which are highly supercoiled. Using anti-Z-DNA antibody, they mapped these tracts to nucleotides 258-265, 199-205, and 128-133. Interestingly, DNAase I hypersensitive sites localized by others appear to bracket these same three segments.

Far less is actually known about the regulation of late transcription. The late region encodes mRNAs generated by a series of processing events that ultimately results in two size classes, 19S and 16S. 19S mRNAs are translated to the minor capsid proteins VP2 and VP3, as well as the major capsid protein VP1. The 16S mRNAs are translated

primarily to VP1 (Prives and Shure, 1979). Regulatory elements which may be involved in controlling the rate or content of expressed genes are not well defined. Late transcription is initiated at a number of sites which span a region of about 220 bp. The major site is positioned at nucleotide 325, and a second minor site at position 185. For the initiation site at nucleotide 325, sequences 25 bp upstream from the site appear to function like a TATA box. The six G+C motifs located far upstream are also required for full expression of both the 325 and the 185 start sites (Hartzell et al., 1984).

Recent work by Ernoult-Lange et al. (1987) explored the role of the 72 base pair repeats in late transcription. These investigators constructed an extensive series of deletion mutants that encompassed both the GC rich repeats and the enhancer region. Using this approach they identified two distinct domains in the enhancer that are important for late promoter activity. Domain I located between nucleotides 232 and 265 is primarily active before replication begins, and comprises the late proximal end of the 72 bp repeats and extends 15 bp to the late side of the enhancers. Domain II extends from nucleotides 179 to 204, and is primarily active following replication. Previous work indicated that protein-protein interactions occur between factors bound to the GC rich region and the enhancer. Together with the data of Ernoult-Lange et al. this suggests that prior to replication the GC rich region promotes transcription, chiefly in the early direction, whereas after replication this effect is reduced and domain II stimulates transcription in the late direction.

In addition to the viral coat proteins, a small 61 amino acid, highly basic polypeptide encoded within the 5' leader of the late 16S mRNA is also produced during late transcription (Jay et al., 1981). For some time the function of this protein, termed the agnoprotein, was unclear. Using single-base substitution mutations, Resnick and Shenk (1986) have recently established that the production of this protein is not important for early gene expression, DNA replication, synthesis of late viral products, or the kinetics of virion assembly. Immunofluorescent data, however, supports the notion that it is required for cell to cell spread of the virus, and that it facilitates the normal positioning of the major capsid polypeptide within the nucleus. Specifically, Resnick and Shenk speculate that the agnoprotein is responsible for the mechanical breakdown of infected cells. If that is true, it is unclear why the protein would be equipped with both single and double-stranded DNA binding activity (Jay et al., 1981). An alternative suggestion is that the agnoprotein is responsible for the intranuclear organization of mature virus particles in a manner which favors efficient release of the virus.

Additional cellular factors which are important in both early and late transcription were identified in 1983 by Dynan and Tjian. Using in vitro transcription systems with crude extracts in an attempt to determine the basis of transcriptional selectivity, they identified two cellular factors which have key roles. One of these, Sp1, is promoter-specific and activates the SV40 early and late promoters. The other, Sp2, is a general transcription factor which was required for

transcription of all the promoters tested. It was suggested that Sp1 might work to facilitate interaction of the transcription machinery with weak promoters, that it blocks the action of some cellular inhibitors, or that it interacts with the enhancer sequences located far upstream.

5. Overview of Large and Small Tumor Antigen Functions

Extensive information is available about regulation and expression of the SV40 early genes. The large T antigen is a remarkable protein which carries out multiple functions, many of which are only superficially understood. Most of the information available to date comes from studies using TAg temperature sensitive early mutants. Results indicate that cells infected at the restrictive temperature overproduce early RNA, do not initiate replication, and do not initiate synthesis of late mRNA (Reed et al., 1976; Myers et al., 1981b; Fromm and Berg 1982). In addition, TAg has an ATPase function which is probably required for replication (Clark et al., 1984). It also has the ability to bind adenylated nucleotides which relates to its ability to initiate transcription (Bradley et al., 1984). Finally, large TAg is critical for the transformation capabilities of the virus.

The function of small t antigen is poorly understood, but it is believed to relate to the initiation of cellular transformation in established cell lines, but not in primary cultures or rapidly dividing cells (Sleigh et al., 1978).

6. Mechanism of Large TAg Binding to the Regulatory Region

The mechanism by which TAg recognizes its binding sites and coordinates gene expression has only recently become apparent. A systematic examination of the nucleotide sequence of TAg binding at site I demonstrates that the essential binding sequences consist of 17 bp of DNA that contain two directly repeated pentanucleotides separated by a 7 bp spacer of the sequence 5'-TTTTTTG-3' (Ryder et al., 1986). Tjian et al. (1978) showed that single base pair substitutions within the spacer impair binding which implies that these sequences have a specific function. This is somewhat confusing since there is no evidence for any contacts between TAg and spacer nucleotides. Ryder and his colleagues recently suggested that the principal function of the spacer might be to alter the structure of the DNA helix (Ryder et al., 1986). Upon examination, it was found that a restriction fragment containing wild type region I displayed temperature-dependent electrophoretic anomalies, thus demonstrating that these fragments have a conformationally altered structure. This distortion is absent when single base substitutions are made within the spacer region. Similar observations made previously on restriction fragments isolated from Trypanosome kinetoplast DNA have been operationally defined as bent DNA (Marini et al., 1982). Wu and Crothers (1984) reported the presence of a poly A-poly T tract which they demonstrated has an altered structure that leads to bending at each of its junctions with B DNA. Ryder et al. (1986) proposed that the bend is centered around the spacer which, through an unknown mechanism,

stabilizes and coordinates the binding of monomer units of TAg with the neighboring pentanucleotides.

Recently Deb et al. (1986) showed that a second sequence-directed bend exists in the SV40 regulatory region, this one centered in the homopolymeric dA.dT region between nucleotides 22-30. It is probably more than coincidence that these sequences have also been defined as part of the minimal origin of replication, and that mutations which change the bending of the wild-type DNA also affect the efficiency of replication. While these studies cannot precisely define the mechanistic role of DNA bending in replication, the authors suggest that the homopolymeric tract may be a recognition sequence for the binding of viral or cellular proteins, or that DNA bending may actually position the dA.dT tract, or some protein bound to it, in proper helical alignment for replication. Also, it is formally possible that the bend simply facilitates strand separation during replication. However, this suggestion would be in contradiction with the observation that poly(dA).poly(dT) tracts are actually more stable than alternating copolymer poly dA.dT tracts.

7. Other TAg Functions

a. ATPase and Helicase Activities.

Since TAg has ATPase activity, there was early speculation that it also had DNA unwinding capabilities. Comprehensive studies using a SV40/Adenovirus hybrid TAg (D2 protein) on wild type SV40 DNA and plasmids with multiple copies of the SV40 origin region did not support this conclusion (Myers et al., 1981a). Recently, Dean et al. (1987)

again addressed the question and determined that TAg, in conjunction with ATP and a cellular topoisomerase, can catalyze the conversion of a plasmid containing the SV40 origin of replication to a highly underwound covalently closed circle. In addition to ATP and a cellular topoisomerase, the reaction also requires single-stranded DNA binding proteins and Mg^{+2}.

b. Role in Viral Transformation

Many investigators have shown that the TAg oncogene is capable of immortalizing primary cells as well as transforming primary and established cells (for review see Fried et al., 1986). Using mutation analysis, it was shown that the SV40 TAg is required for the initiation and maintenance of the transformed state (Brugge and Butel, 1975; Tegtmeyer, 1975; Kimura and Itagaki, 1975), but the manner in which this is accomplished is unclear. Neither the DNA binding, DNA replication nor ATPase activities of TAg are required for transformation (Pipas et al., 1983; Cole et al., 1986). Ketner and Kelly (1976) investigated the pattern of integration of SV40 sequences in transformed cells using restriction endonucleases. When DNA from five independent transformants of BALB/c 3T3 mouse cell lines were digested with Hpa II or Bam HI and then analyzed by agarose gel electrophoresis they found that the integration mechanism was random with regard to both chromosomal location and viral junction sequences for every cell line.

Complete understanding of the mechanism by which TAg facilitates transformation is complicated by reports that TAg forms extremely stable

complexes with the cellular p53 protein (Lane and Crawford, 1979). Recently, the possibility that distinct structural subclasses of large TAg bind the p53 protein has been addressed. Montenarh et al. (1986) observed that newly synthesized cellular p53 bound to all structural subclasses of TAg. They were able to detect various intermediates of the T-p53 complexes in sucrose gradients, but these were unremarkable and formed into typical mature aggregates within two hours.

8. SV40 Chromatin Structure

The intracellular nucleoprotein or "minichromosome" complexes of SV40 have a structure similar to that of eucaryotic chromatin, with approximately 21 core particles around each of which is wrapped 60 ± 40 bp of viral DNA (Lutter et al., 1979; Shelton et al., 1980). Numerous studies, however, have reported that the nucleosome positioning along the SV40 viral sequences is not random, at least in a subpopulation of minichromosomes. Electron microscopy studies in 1980 by both Jakobavits et al. and Saragosti et al. reported the absence of nucleosomes in a fraction of SV40 chromatin in the region between the origin of replication and the sequences coding for the late leader mRNA. Both groups determined that approximately 25% of viral minichromosome molecules had a visible gap of about 320±90 bp. To locate the gap relative to known restriction sites, minichromosome molecules were digested with single-cut restriction enzymes. They determined that enzymes which cut in the origin region, such as Bgl I, yielded about 70% linear molecules, whereas those which cut elsewhere in the genome, such

as Eco RI or Bam HI yielded only about 25% linear molecules. This allowed them to unequivocally locate the exposed stretch of DNA as starting very close to the Bgl I site and extending into the late region for about 320 base pairs.

Many other studies had shown previously that the same region was particularly sensitive to restriction endonucleases (Scott et al., 1978; Sundin et al., 1979; Varshavsky et al., 1978; Varshavsky et al., 1979; Waldeck et al., 1978). An elegant set of experiments using double origin mutants was done by Wigmore et al. (1980) who compared four partially duplicated mutants with respect to their nuclease sensitivity. They found that all origin segments were preferentially and equally cleaved by endogenous endonuclease, indicating that the altered nucleoprotein structure is sequence-dependent and unrelated to the positioning or number of such sequences within the viral genome. It was suggested that this altered nucleoprotein structure is critical for virion assembly, transcription, or for release of newly formed viral particles.

Robinson and Hallick (1982) examined the state of the viral minichromosome 46 hours post-infection using psoralen photoaddition at which time viral replication is at its maximum. Again it was determined that approximately 25% of molecules contained an open origin. More recently the SV40 genome examined in extracellular and intracellular virus using psoralen photoaddition was also shown to contain a nucleosome free origin (Kondoleon et al., in preparation). There is substantial evidence that the SV40 genome in virus particles is structurally different from the minichromosome in infected cells. Moyne et al. (1982)

have shown that the nucleoprotein complexes obtained from virions by mild dissociation at neutral pH are not arranged as typical nucleosomes. Instead they appear as circular structures with a mean of 93±17 beads with an observed diameter of 7nm. They suggest that the major capsid protein VP1 is partially involved in unfolding the nucleosomal structure.

The studies of Kondoleon et al. also demonstrated that the state of the chromatin is altered by common virus storage techniques such as freezing at -20°C. Samples which were previously frozen do not exhibit a preferentially psoralen accessible region at the origin, while those that are stored at 4° do. While this data does not immediately imply a mechanism for the altered psoralen specificity, this finding has interesting implications for understanding the physical mechanism of how nucleosome phasing and rearrangement occurs. Most recently, Milavetz (1986) analyzed the chromatin structure in encapsidation intermediates and determined that a nucleosome-free origin exists in non-encapsidated and partially encapsidated SV40 chromosomes and previrions.

C. Summary

Clearly the SV40 regulatory region is a multifunctional set of sequences whose mechanism of action is, to date, only superficially understood. This region is characterized by sequences with unusual helical perturbations, as well as a remarkable chromatin structure. Psoralen photoaddition provides a sensitive and unique mechanism for the investigation of DNA structure-function relationships in any system, and was therefore chosen to investigate the SV40 regulatory region both at

the DNA and chromatin levels. At the initiation of these thesis studies in 1981, psoralen-DNA interactions were still very poorly understood at the sequence level. The first manuscript characterizes such interactions by examining cleavage inhibition patterns of the restriction enzyme Bgl I following photoaddition of pBR322 under conditions which favor the formation of particular types of photoadducts. Findings from these experiments were used to successfully design a fine structure mapping technique which allowed the localization of psoralen monoadducts to be made precisely at the nucleotide level. This technique, which is described in the second manuscript was used to examine the psoralen accessibility of the SV40 origin region and surrounding sequences. Finally, in part III, application of this technique to the in vivo characterization of SV40 chromatin complexes is discussed.

BIBLIOGRAPHY

Bachellerie, J.P., and Hearst, J.E. (1982). Specificity of the photoreaction of 4'-(hydroxymethyl)-4,5',8-trimethylpsoralen with ribonucleic acid. Identification of reactive sites in Escherichia coli phenylalanine-accepting transfer ribonucleic acid. Biochemistry 21, 1357-1363.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33, 729-740.

Barenbaum, M. (1978). Toxicity of a furanocoumarin to armyworms: A case of biosynthetic escape from insect herbivores. Science 201, 532-534.

Beaumont, P.C., Parsons, B.J., Phillips, G.O., and Allen, J.C. (1979). A laser flash photolysis study of the reactivities of the triplet states of 8-methoxypsoralen and 4,5',8-trimethylpsoralen with nucleic acid bases in solution. Biochim. Biophys. Acta. 562, 214-221.

Becker, M.M., and Wang, J.C. (1984). Use of light for footprinting DNA in vivo. Nature 309, 682-687.

Benoist, C., and Chambon P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature 290, 304-310.

Bordin, F., Carlassare, F., Baccichetti, F., Guiotto, A., and

Rodighiero, P., et al. (1979a). 4,5'-dimethylangelicin: A new DNA-photobinding monofunctional agent. Photochem. Photobiol. 29, 1063.

Bordin, G., Carlassare, F., Baccichetti, F., Guiotto, A., and Rodighiero, P. (1979b). 5-methylangelicin: A new highly photosensitizing angular furocoumarin. Experentia 35, 1567-1568.

Bradley, M.K., Hudson, J., Villanueva, M., and Livingston, D.M. (1984). Specific in vitro adenylation of the simian virus 40 large T antigen. Proc. Natl. Acad. Sci. USA 81, 6574-6578.

Brugge, J.S., and Butel, J.S. (1975). Involvement of the simian virus 40 gene. A function in maintenance of transformation. J. Virol. 15, 619-635.

Carter, D.M., and Radding, C.M. (1971). The role of exonuclease and × protein in phage genetic recombination. J. Biol. Chem. 246, 2502-2512.

Chatterjee, P.K. and Cantor, C.R. (1978). Photochemical production of psoralen-DNA monoadducts capable of subsequent photocrosslinking. Nucl. Acids Res. 5, 3619-3633.

Cimino, G.D., Gamper, H.B., Isaacs, S.T., and Hearst, J.E. (1985). Psoralens as photoactive probes of nucleic acid structure and function: Organic chemistry, photochemistry and biochemistry. Ann. Rev. Biochem. 54, 1151-1193.

Cimino, G.D., Shi, Y.B., and Hearst, J.E. (1986). Wavelength dependence for the photoreversal of a psoralen-DNA crosslink. Biochemistry 25, 3013-3020.

Clark, R., Tevethia, M.J., and Tjian R. (1984). The ATPase activity of SV40 large T antigen. Cancer Cells 2, 363.

Cole, C.N., Tornow, J., Clark, R., and Tjian, R. (1986). Properties of the simian virus 40 (SV40) large T antigen encoded by SV40 mutants with deletions in gene A. J. Virol. 57, 539-546.

Cole, R.S. (1970). Light-induced cross-linking of DNA in the presence of a furocoumarin (psoralen). Biochim. Biophys. Acta 217, 30-39.

Cole, R.S. (1971). Psoralen monoadducts and interstrand cross-links in DNA. Biochim. Biophys. Acta 251, 30-39.

Courey, A.J., Plon, S.E., and Wang, J.C. (1986). The use of psoralen-modified DNA to probe the mechanism of enhancer action. Cell 45, 567-574.

Cowan, K., Tegtmeyer, P., and Anthony, D.D (1973). Relationship of replication and transcription of simian virus 40 DNA. Proc. Natl. Acad. Sci. USA 70, 1927-1930.

Dall'Acqua, F., Marciani, S., and Rodighiero, G. (1970). Inter-strand cross-linkages occurring in the photoreaction between psoralen and DNA. FEBS Lett. 9, 121-123.

Dall'Acqua, F., Marciani, S., Vedaldi, D. and Rodighiero, G. (1972). Formation of inter-strand cross-linkings on DNA of guinea pig skin after application of psoralen and irradiation at 365nm. FEBS Lett. 27, 192-194.

Das, G.C., Niyogi, S.K., and Salzman, N.P. (1985). SV40 promoters and their regulation. Progress in Nucl. Acids Res. 32, 217-236.

Davis, R.W., Simon, M., and Davidson, N. (1971). Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. Methods. Enzymol. 21: 413-428.

Dean, F.B., Bullock, P., Murakami, Y., Wobbe, C.R., Weissbach, L., and Hurwitz, J. (1987). Simian virus 40 (SV40) DNA replication: SV40 large T antigen unwinds DNA containing the SV40 origin of replication. Proc. Natl. Acad. Sci. USA 84, 16-20.

Deb, S., DeLucia, A., Koff, A., Tsui, S., and Tegtmeyer, P. (1986). The adenine-thymine domain of the simian virus 40 core origin directs DNA bending and coordinately regulates DNA replication. Mol. and Cell. Biol. 6, 4578-4584.

DeLucia, A., Lewton, B., Tjian, R., and Tegtmeyer, P. (1983). Topography of simian virus 40 A Protein-DNA complexes: Arrangement of pentanucleotide interaction sites at the origin of replication. J. Virol. 46, 143-150.

DeLucia, A.L., Deb. S., Partin, K., and Tegtmeyer, P. (1986). Functional interactions of the SV40 core origin of replication with flanking regulatory sequences. J. Virol. 57, 138-144.

Dickerson, R.E., and Drew, H.R. (1981). Structure of a β-DNA dodecamer II. Influence of base sequence on helix structure. J. Mol. Biol. 149, 761-786.

DiMaio, D., and Nathans, D., (1980). Cold sensitive regulatory mutants of simian virus 40. J. Mol. Biol. 140, 129-142.

DiMaio, D., and Nathans, D. (1982). Regulatory mutants of simian virus

40: Effect of mutations at a T antigen binding site on DNA replication and expression of viral genes. J. Mol. Biol. 156, 531-548.

Dynan, W.S., and Tjian R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell 32, 669-680.

Eddy, B.E., Borman, G.S., Grubbs, G.E., and Young, R.D. (1962). Identification of the oncogenic substance in rhesus monkey kidney cell cultures as simian virus 40. Virology 17, 65-75.

Eissenberg, J.C., Cartwright, I.L., Thomas, G.H., and Elgin, S.C.R. (1985). Selected topics in chromatin structure. Ann. Rev. Genet. 19, 485-536.

El Mofty, A.M. (1948). A preliminary clinical report on the treatment of leucoderma with Ammi Majus Linn. J. Roy. Egypt Med. Assoc. 31, 651-665.

Ernoult-Lange, M., Omilli, F., O'Reilly, D.R., and May, E. (1987). Characterization of the simian virus 40 late promoter: relative importance of sequences within the 72-base-pair repeats differs before and after viral DNA replication. J. Virol. 61, 167-176.

Everett, R.D.. Baty, D., and Chambon, P. (1983). The repeated G-C rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucl. Acids Res. 11, 2447-2464.

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., van der Voorde, A.,

van Heuverswyn, H., van Herreweghe, J., Volckaert, V., and Ysebaert, M. (1978). Complete nucleotide sequence of SV40 DNA. Nature 273, 113-120.

Fried, M., and Prives, C. (1986). The biology of simian virus 40 and polyomavirus. Cancer Cells. 4,1-16.

Fromm, M., and Berg, P. (1982). Deletion mapping of DNA required for SV40 early region promoter region function in vivo. J. Mol. Appl. Genet. 1, 457-481.

Gamper, H., Piette, J., and Hearst, J.E. (1984). Efficient formation of a crosslinkable HMT monoadduct at the Kpn I recognition site. Photochem. and Photobiol. 40, 29-34.

Garrett-Wheeler, E., Lockard, R.E., and Kumar, A. (1984). Mapping of psoralen cross-linked nucleotides in RNA. Nucl. Acids Res. 12, 3405-3423.

Germond, J.E., Hirt, B., Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Folding of the DNA double helix in chromatin-like structures from simian virus 40. Proc. Natl. Acad. Sci. USA 72, 1843-1847.

Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983). A tissue specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33, 717-728.

Girardi, A.J., Sweet, B.H., Slotnick V.B., and Hilleman, M.R. (1962). Development of tumors in hamsters inoculated in the neonatal period with vacuolating virus, SV40. Proc. Soc. Exp. Biol. Med. 109, 649.

Graessmann, A., Graessmann, M., Guhl, E., Mueller, C. (1978). Quantitative correlation between SV40 T-antigen synthesis and late viral gene expression in permissive and non-permissive cells. J. Cell. Biol. 77, R1-R8.

Granger, M., Toulme, F., and Helene, C. (1982). Photodynamic inhibition of Escherichia coli DNA polymerase I by 8-methoxypsoralen plus ultraviolet irradiation. Photochem. Photobiol. 36, 175-180.

Griffith, J. (1976). Visualization of procaryotic DNA in a regularly condensed chromatin-like fiber. Proc. Natl. Acad. Sci. USA 73, 563-567.

Gruss, P., Dhar, R., and Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. USA 78, 943-947.

Hallick, L.M., Hanson, C.V., Cacciapuoti, J.O., and Hearst, J.E. (1980). Photoaddition of trimethylpsoralen as a probe for the intracellular organization of Escherichia coli DNA. Nuc. Acids Res. 8, 611-622.

Hansen, C.V., Shen, C-K.J., Hearst, J. (1976). Cross-linking of DNA in situ as a probe for chromatin structure. Science 193, 62-64.

Hansen, J.B., Bjerring, P., Buchardt, O., et al. (1985). Psoralenamines. 3. Synthesis, Pharmacological Behavior, and DNA Binding of 5-(aminomethyl)-8-methoxy-, 5-[[(3-aminopropyl)oxy]methyl]-, and 8-[3-aminopropyl)oxy]psoralen derivatives. J. Med. Chem. 28, 1001-1010.

Hartzell, S.W., Byrne, B.J., and Subramanian, K. (1984). Mapping the late promoter of simian virus 40. Proc. Natl. Acad. Sci. USA 81, 23-27.

Honigsmann, A.K., Konrad, K., Schnait, F.G., and Wolf, K. (1967). Presented at the 7th Int. Cong. Photobiol., Rome. Book of Abstracts.

Hui, C.-F. and Cantor, C.R. (1985). Mapping the location of psoralen crosslinks on RNA by mung bean nuclease sensitivity of RNA-DNA hybrids. Proc. Natl. Acad. Sci. USA 82, 1381-1385.

Isaacs, S.T., Shen, C.-K.J., Hearst, J.E., and Rapoport, H. (1977). Synthesis and characterization of new psoralen derivatives with superior photoreactivity with DNA and RNA. Biochemistry 16, 1058-1064.

Isaacs, S.T., Chun, C., Hyde, J.E., Rapoport, H., and Hearst, J.E. (1982). Trends in Photobiology, ed. C. Helene, M. Charlier, Th. Montenay-Garestier, G. Laustriat, pp. 279-294. New York.

Jakobovits, E.B., Bratosin, A., and Aloni, Y. (1980). A nucleosome-free region in SV40 minichromosomes. Nature 285, 263-265.

Jay, G., Nomura, S., Anderson, C.W., and Khoury, G. (1981). Identification of the SV40 agnogene product: a DNA binding protein. Nature 291, 346-349.

Johnston, B.H., Johnson, M.A., Moore, C.B., and Hearst J.E. (1977). Psoralen-DNA photoreaction: Controlled production of mono- and diadducts with nanosecond ultraviolet laser pulses. Science 197, 906-908.

Jongstra, J., Reudelhuber, T.L., Oudet, P., et. al. (1984). Induction of altered chromatin structures by simian virus 40 enhancer and promoter elements. Nature 307, 708-714.

Kanne, D., Straub, K., Hearst, J.E., and Rapoport, H. (1982a). Isolation and characterization of pyrimidine-psoralen-pyrimidine photo-diadducts from DNA. J. Am. Chem. Soc. 104, 6754-6764.

Kanne, D., Straub, K., Rapoport, H., and Hearst, J.E. (1982b). Psoralen-DNA photoreaction. Characterization of the monoaddition products from 8-methoxypsoralen and 4,5',8-trimethylpsoralen. Biochemistry 21, 861-871.

Kanne, D., Rapoport, H., and Hearst, J.E. (1984). 8-methoxypsoralen-nucleic acid protoreaction. Effect of methylsubstitution on pyrone vs. furan photoaddition. J. Med. Chem. 27, 531-534.

Kelly, T.J., and Nathans, D. (1977). The genome of SV40. Adv. Virus Res. 21, 85-173.

Ketner, G., and Kelly, T.J. (1976). Integrated simian virus 40 sequences in transformed cell DNA: Analysis using restriction endonucleases. Proc. Natl. Acad. Sci. USA 73, 1102-1106.

Kimura, G., and Itagaki, A. (1975). Initiation and maintenance of cell transformation by simian virus 40. Proc. Natl. Acad. Sci. USA 72, 673-677.

Kittler, L., Hradecna, Z., and Suhnel, J. (1980). Crosslink formation

in phage lambda DNA in situ photochemically induced by the furocoumarin derivative angelicin. Biochim. Biophys. Acta 607, 215-220.

Kittler, L., Midden, W.R., and Wang, S.Y. (1983). Photoreactions of lipids sensitized by furocoumarins. Photochem. Photobiol. Suppl. 37, S16.

Kondoleon, S.K., Kurkinen, N.A., and Hallick, L.M. (1986). The SV40 nucleosome-free region is detected throughout the virus life cycle. (In preparation).

Krauch, C.H., Kramer, D.M., and Wacker, A. (1967). Zum Wirkungsmechanismum photodynamischer furocoumarine. Photoreaktion von psoralen-(4-^{14}C) mit DNA, RNA, homopolynucleotiden und nucleosiden. Photochem. Photobiol. 6 341-354.

Lane, D.P., and Crawford, L.V. (1979). T antigen is bound to a host protein in SV40-transformed cells. Nature 278, 261-263.

Lawley, P.D., and Brookes, P. (1967). Interstrand cross-linking of DNA by difunctional alkylating agents. J. Mol. Biol. 25, 143-160.

Lerman, S., Megaw, J., and Willis, I. (1980). The photoreaction of 8-methoxypsoralen with tryptophan lens proteins. Photochem. Photobiol. 31, 233-242.

Li, J.J. and Kelly, T.J. (1984). Simian virus 40 DNA replication in vitro. Proc. Natl. Acad. Sci. USA 81, 6973-6977.

Little, J.W., Lehman, I.R., and Kaiser, A.D. (1967a). An exonuclease induced by bacteriophage λ. Part I. J. Biol. Chem. 242, 672-678.

Little, J.W. (1967b). An exonuclease induced by bacteriophage λ. Part II. J. Biol. Chem. 242, 679-686.

Lown, J.W., and Sim, S.-K. (1978). Photoreaction of psoralen and other furocourmarins with nucleic acids Bioorganic Chem. 7, 85-95.

Lutter, L.C., (1979). Precise location of DNase I cutting sites in the nucleosome core determined by high resolution gel electrophoresis. Nucl. Acids Res. 6, 41-56.

Marini, J.C., Levene, S.D., Crothers, D.M., and Englund, P.T. (1982). Bent helical structure in kinetoplast DNA. Proc. Natl. Acad. Sci. USA. 79, 7664-7688.

Masamune, Y., Fleischman, R.A., and Richardson, C.C. (1971). Enzymatic removal and replacement of nucleotides at single strand breaks in deoxyribonucleic acid. J. Biol. Chem. 246, 2502-2510.

Matulin, W.W., Song, P.-S. (1973). Excited states of skin-sensitizing coumarins and psoralens. Spectroscopic studies. J. Am. Chem. Soc. 95, 5122-5129.

Milavetz, B. (1986). Analysis of the origin-specific nucleosome free region in SV40 encapsidation intermediates. Virology 153, 310-313.

Montenarh, M., Kohler, M., and Henning, R. (1986). Complex formation of simian virus 40 large T antigen with cellular protein p53. J. Virol. 60, 415-422.

Moyne, G., Harper, F., Saragosti, S., and Yaniv, M. (1982). Absence of nucleosomes in a histone-containing nucleoprotein complex obtained by dissociation of purified SV40 virions. Cell 30, 123-130.

Musajo, L., Rodighiero, G., and Dall'Acqua, F. (1965). Evidences of a photoreaction of the photosensitizing furocoumarins with DNA and with pyrimidine nucleosides and nucleotides. Experientia 21, 24-26.

Musajo, L., Rodighiero, G., Breccia, R., Dall'Acqua, F., and Malesani, G. (1966). Skin-photosensitizing furocoumarins; photochemical interaction between DNA and $-O^{14}CH_3$ bergapten (5-methoxy-psoralen). Photochem. Photobiol. 5, 739-745.

Mustafa, A. (1967). Furopyrans and Furopyrones, in The Chemistry of Heterocyclic Compounds. Vol. 23, 15. New York: Wiley.

Myers, R.M., and Tjian, R. (1980). Construction and analysis of simian virus 40 origins defective in tumor antigen binding and DNA replication. Proc. Natl. Acad. Sci. USA 77, 6491-6495.

Myers, R.M., Kligman, M., and Tjian, R. (1981a). Does simian virus 40 T antigen unwind DNA? J. of Biol. Chem. 256, 10156-10160.

Myers, R.M., Rio, D.C., Robbins, A.K., and Tjian, R. (1981b). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell 25, 373-384.

Noll, M. (1974). Subunit structure of chromatin. Nature 251, 249-251.

Norheim, A., and Rich, A. (1983). Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature 303, 674-679.

Ou, C.N., Tsai, C.H., Tapley, K.J., and Song, P.S. (1978a). Photobinding of 8-methoxypsoralen and 5,7-dimethylcoumarin to DNA and its effect on template activity. Biochemistry 17, 1047-1054.

Ou, C.N., and Song, P.S. (1978b). Photobinding of 8-methoxypsroalen to transfer RNA and 5-fluorouracil-enriched transfer RNA. Biochemistry 17, 1054-1059.

Parrish, J.A., Fitzpatrick, T.B., Tanenbaum, L., and Pathak, M.A. (1974). Photochemotherapy of psoriasis with oral methoxypsoralen and long wavelength ultraviolet light. N. Eng. J. Med. 291, 1207-1211.

Pathak, M.A. and Kramer, D.M. (1969). Photosensitization of skin in vivo by furocoumarins (psoralens). Biochim. Biophys. Acta 195, 197-206.

Perone, V.B. (1972). The natural occurrence and use of the toxic coumarins. in Microbioal Toxins Vol. 8, 71-81. Academic Press: New York.

Piette, J.G., and Hearst, J.E. (1983). Termination sites of the in vitro nick-translation reaction on DNA photoreacted with psoralen. Proc. Natl. Acad. Sci. USA 80, 5540-5544.

Pipas, J.M., Peden, K.W.C., and Nathans, D. (1983). Mutational analysis of simian virus 40 T antigen: Isolation and characterization of mutants with deletions in the T-antigen gene. Mol. Cell. Biol. 3, 203-213.

Plon, S.E., and Wang, J.C. (1986). Transcription of the human β globin gene is stimulated by an SV40 enhancer to which it is physically linked but topologically uncoupled. Cell 45, 575-580.

Privis, C., Gluzman, Y., and Winocour, E. (1978). Cellular and cell-free synthesis of simian virus 40 T-antigens in permissive and transformed cells. J. Virol. 25, 587-595.

Privis, C., and Shure, H. (1979). Cell-free translation of simian virus 40 16S and 19S L-strand specific mRNA classes to simian virus 40 major VP-1 and minor VP-2 and VP-3 capsid proteins. J. Virol. 29, 1204-1212.

Queen, C. and Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. Cell 33, 741-748.

Rabin, D. and Crothers, D.M. (1979). Analysis of RNA secondary structure by photochemical reversal of psoralen crosslinks. Nucl. Acids Res. 7, 689-703.

Reed, S.I., Startk, G.R., and Alwine J.C. (1976). Autoregulation of simian virus 40 gene A by T antigen. Proc. Natl. Acad. Sci. USA 73, 3083-3087.

Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L., and Weissman, S.M. (1978). The genome of simian virus 40. Science 200, 494-502.

Resnick, J., and Shenk, T. (1986). Simian virus 40 agnoprotein facilitates normal nuclear location of the major capsid polypeptide and cell-to-cell spread of virus. J. Virol. 60, 1098-1106.

Robinson, G.W. and Hallick, L.M. (1982). Mapping the in vivo arrangement of nucleosomes on simian virus 40 chromatin by the photoaddition of radioactive hydroxymethyltrimethylpsoralen. J. Virol. 41, 78-87.

Ryder, K., Silver, S., DeLucia, A., Fanning E., and Tegtmeyer, P. (1986). An altered DNA conformation in origin region I is a determinant for the binding of SV40 large T antigen. Cell 44, 719-725.

Saragosti, S., Moyne, G., and Yaniv, M. (1980). Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. Cell 20, 65-73.

Sassone-Corsi, P., Wildeman, A., and Chambon, P. (1985). A trans-acting factor is responsible for the simian virus 40 enhancer activity in vitro. Nature 313, 458-463.

Scott, W.A., and Wigmore, D.J. (1978). Sites in simian virus 40 chromatin which are preferentially cleaved by endonucleases. Cell 15, 1511-1518.

Shelton, E.R., Wassarman, P.M., and DePamphilis, M.L. (1980). Structure, spacing and phasing of nucleosomes on isolated forms of mature SV40 chromosome. J. Biol. Chem. 255, 771-782.

Shortle, D., Margolskee, R., and Nathans, D. (1979). Mutational analysis of the simian virus 40 replicon: pseudorevertants of mutants with a defective replication origin. Proc. Natl. Acad. Sci. USA 76, 6128-6131.

Shulman, M.J., Hallick, L.M., Echols, H., and Signer, E.R. (1970). Properties of recombination-deficient mutants of bacteriophage lambda. J. Mol. Biol. 52, 501-520.

Sleigh, M.J., Topp, W.C., Hanich, R., and Sambrook J.F. (1978). Mutants of SV40 with an altered small t protein are reduced in their ability to transform cells. Cell 14, 79-88.

Sobell, H.M., Sakore, T.D., Jain, S.C., Banerjee, A., Bhandary, K.K., Reddy, B.S., and Lozanosky, E.D. (1982). Cold Spring Harbor Symposium on Quantitative Biology 47, 293-314.

Song, P.-S., Harter, M.L., Moore, T.A., and Herndon, W.C. (1971). Luminescence spectra and photocycloaddition of the excited coumarins to DNA bases. Photochem. and Photobiol. 14, 521-530.

Sriprakash, K.S., Lundh, N., Huh, M.M.-O., and Radding C.M. (1975). The specificity of λ exonuclease. J. Biol. Chem. 250, 5438-5445.

Stillman, B.W., and Gluzman, Y. (1985). Replication and supercoiling of simian virus 40 DNA in cell extracts from human cells. Mol. Cell. Biol. 5, 2051-2060.

Straub, K., Kanne, D., Hearst, J.E., and Rapoport, H. (1981). Isolation and characterization of pyrimidine-psoralen photoadducts from DNA. J. Am. Chem. Soc. 103, 2347-2355.

Sundin, O., and Varshavsky, A. (1979). Staphylococcal nuclease makes a single non-random cut in the simian virus 40 viral minichromosome J. Mol. Biol. 132, 535-546.

Sweet, B.H., and Hilleman, M.R. (1960). The vacuolating virus, SV40. Proc. Soc. Exp. Biol. Med. 105, 420-427.

Tegtmeyer, P. (1972). Simian virus 40 DNA synthesis: The viral replicon. J. Virol. 10, 591-598.

Tegtmeyer, P. (1975). Function of simian virus 40 gene A in transforming infection. J. Virol. 15, 613-618.

Thomas, K.R., and Olivera, B.M. (1978). Processivity of DNA exonucleases. J. Biol. Chem. 253, 424-429.

Thompson, J.F., Wegnez, M.R., and Hearst, J.E. (1981). Determination of secondary structure of Drosophila melanogaster 5S RNA by hydroxymethyltrimethylpsoralen. J. Mol. Biol. 147, 417-436.

Thompson, J.F., Bachellerie, J.-P., Hall, K., and Hearst, J.E. (1982). Dependence of 4'-(hydroxymethyl)-4,5',8-trimethylpsoralen photoaddition on the conformation of ribonucleic acid. Biochemistry 21, 1363-1368.

Thompson, J.F. and Hearst, J.E. (1983). Structure of E. coli 16S RNA elucidated by psoralen crosslinking. Cell 32, 1355-1365.

Tjian, R. (1978). Protein-DNA interactions at the origin of simian virus 40 DNA replication. Cold Spring Harbor Symp. Quant. Biol. 43, 655-662.

Tooze, J. (1980). DNA Tumor Viruses part 2, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor.

Turner, S., Thompson, J.F., Hearst, J.E., and Noller, H.F. (1982). Identification of a site of psoralen crosslinking in E. coli 16S ribosomal RNA. Nucl. Acids Res. 10, 2839-2849.

Varshavsky, A.J., Sundin, O.H., and Bohn, M.J. (1978). SV40 viral minichromosome: preferential exposure of the origin of replication as probed by restriction endonucleases. Nucl. Acids Res. 5, 3469-3477.

Varshavsky, A.J., Sundin, O., and Bohn, M., (1979). A stretch of "late" SV40 viral DNA about 400 bp long which includes the origin of replication is specifically exposed in SV40 minichromosomes. Cell 16, 453-466.

Veronese, F.M., Schiavon, O., Bevilacqua, R., Bordin, F., and Rodighiero,

G. (1981). The effect of psoralens and angelicins on proteins in the presence of UV-A irradiation. Photochem. Photobiol. 34, 351-354.

Waldeck, W., Fohring, B., Chowdhury, K., Gruss, P., and Sauer, G. (1978). Origin of DNA replication in papovavirus chromatin is recognized by endogenous endonuclease. Proc. Natl. Acad. Sci. USA 75, 5964-5968.

Wildeman, A.G., Zenke, M., Schatz, C., et al. (1986). Specific protein binding to the simian virus 40 enhancer in vitro. Mol. and Cell. Biol. 6, 2098-2015.

Wiesehahn, G.P., Hyde, J.E., and Hearst, J.E. (1977). The photoaddition of trimethylpsoralen to Drosophila melanogaster nuclei: a probe for chromatin substructure. Biochemistry 16, 925-932.

Wigmore, D.J., Eaton, R.W., and Scott, W.A. (1980). Endonuclease-sensitive regions in SV40 chromatin from cells infected with duplicated mutants. Virology 104, 462-473.

Wobbe, C.R., Weissbach, L., Borowiec, J.A., Dean, F.B., Murakami, Y., Bullock, P, and Hurwitz, J. (1987). Replication of simian virus 40 origin-containing DNA in vitro with purified proteins. Proc. Natl. Acad. Sci. USA 84, 18341839-

Wollenzien, P.L., Youvan, D.C., and Hearst, J.E. (1978). Structure of psoralen-crosslinked ribosomal RNA from Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 75, 1642-1646.

Wu, H.-M, and Crothers, D.M. (1984). The locus of sequence-directed and

protein-induced DNA bending. Nature 308, 509-513.

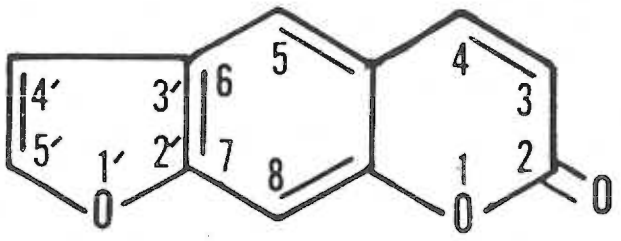
Yang, L., Rowe, T.C., Nelson, E.M., and Liu, L.F. (1985). In vivo mapping of DNA topoisomerase II-specific cleavage sites on SV40 chromatin. Cell 41, 127-132.

Yoshikawa, K., Mori, N., Sakakibara, N., Mizuno, N., and Song, P.S. (1979). Photo-conjugation of 8-methoxypsoralen with proteins. Photochem. Photobiol. 29, 1127-1133.

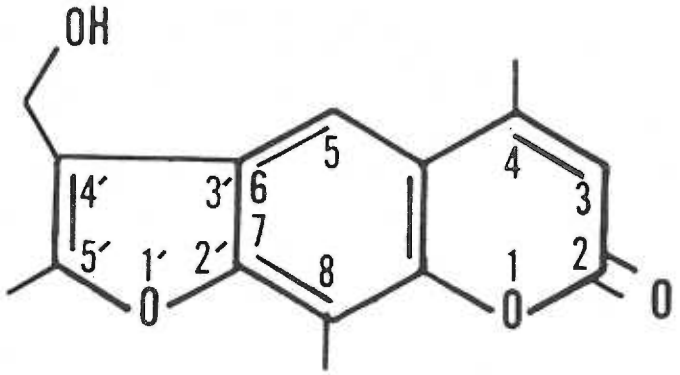
Youvan, D.C. and Hearst, J.E. (1982). Sequencing psoralen photochemically reactive sites in Escherichia coli 16 S rRNA. Anal. Biochem. 119, 86-89.

Zhen, W.-P., Buchardt, O., Nielsen, H., and Nielsen, P.E. (1986). Site specificity of psoralen-DNA interstrand cross-linking determined by nuclease Bal 31 digestion. Biochemistry 25, 6598-6603.

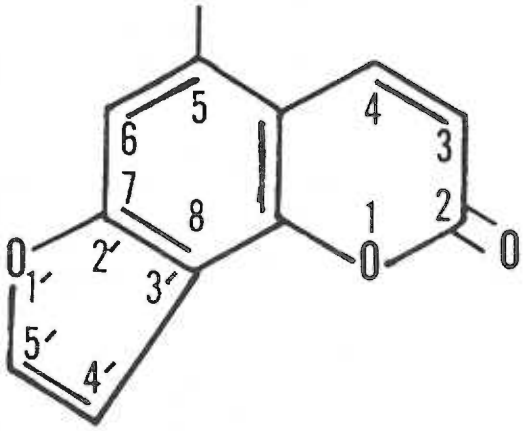
Figure 1. Chemical structures of the psoralen derivatives 4'-hydroxymethyl-4,5',trimethylpsoralen (HMT), and the isopsoralen derivative 5-methylisopsoralen (5-MIP). Absorption in the range of 320-380 nm light results in covalent photocycloaddition of the 4'5'- or 3,4- double bond (furan side or pyrone side respectively) to the 5,6-double bond of a pyrimidine.



PSORALEN



4′-HYDROXYMETHYL-4,5′,8-TRIMETHYLPSORALEN (HMT)



5 - METHYLISOPSORALEN (5 - MIP)

Figure 2. Scheme Showing the Formation of Psoralen Monoadducts and Crosslinks. The psoralen molecule, represented by the black rectangle, intercalates between the stacked bases of the helix. Upon treatment with long wavelength UV light (320-380 nm) a covalent adduct forms between the psoralen molecule and pyrimidine of the DNA. With the addition of another photon, a crosslink forms which effectively ties the two strands of the helix together.

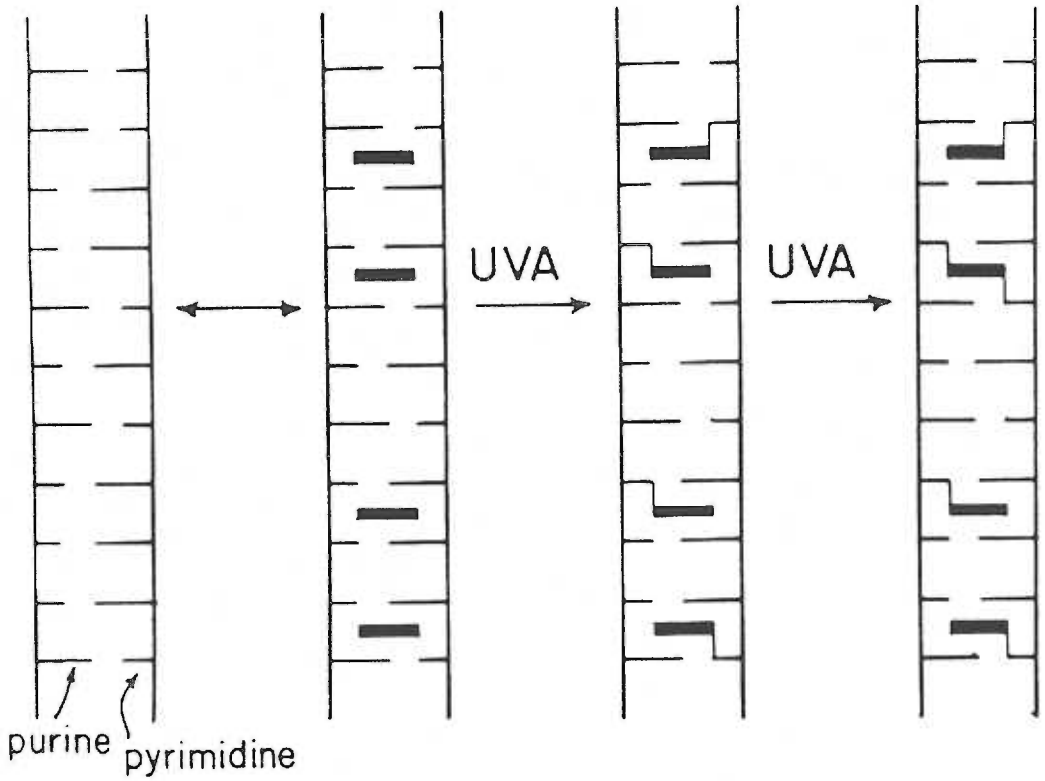

UVA
UVA
purine
pyrimidine

Figure 3. Electron micrograph of HMT crosslinked SV40 DNA (Ostrander, E.A., Stenzel-Poore, M.P., and Hallick, L.M., unpublished data). Supercoiled SV40 DNA was irradiated to a level of about 80 adducts per genome, and then denatured in 0.5 M glyoxal-70% formamide-0.01 M sodium phosphate (pH7.0)-1mM EDTA for 60 min at 40°C. The sample was spread in formamide according to the method of Davis et al. (1971). Grids were rotary shadowed with 80% Pt-Pd and microscopy carried out with a Philips 201 electron microscope. The "bubbles" indicate regions where there were no HMT crosslinks and the two strands of the helix are denatured.

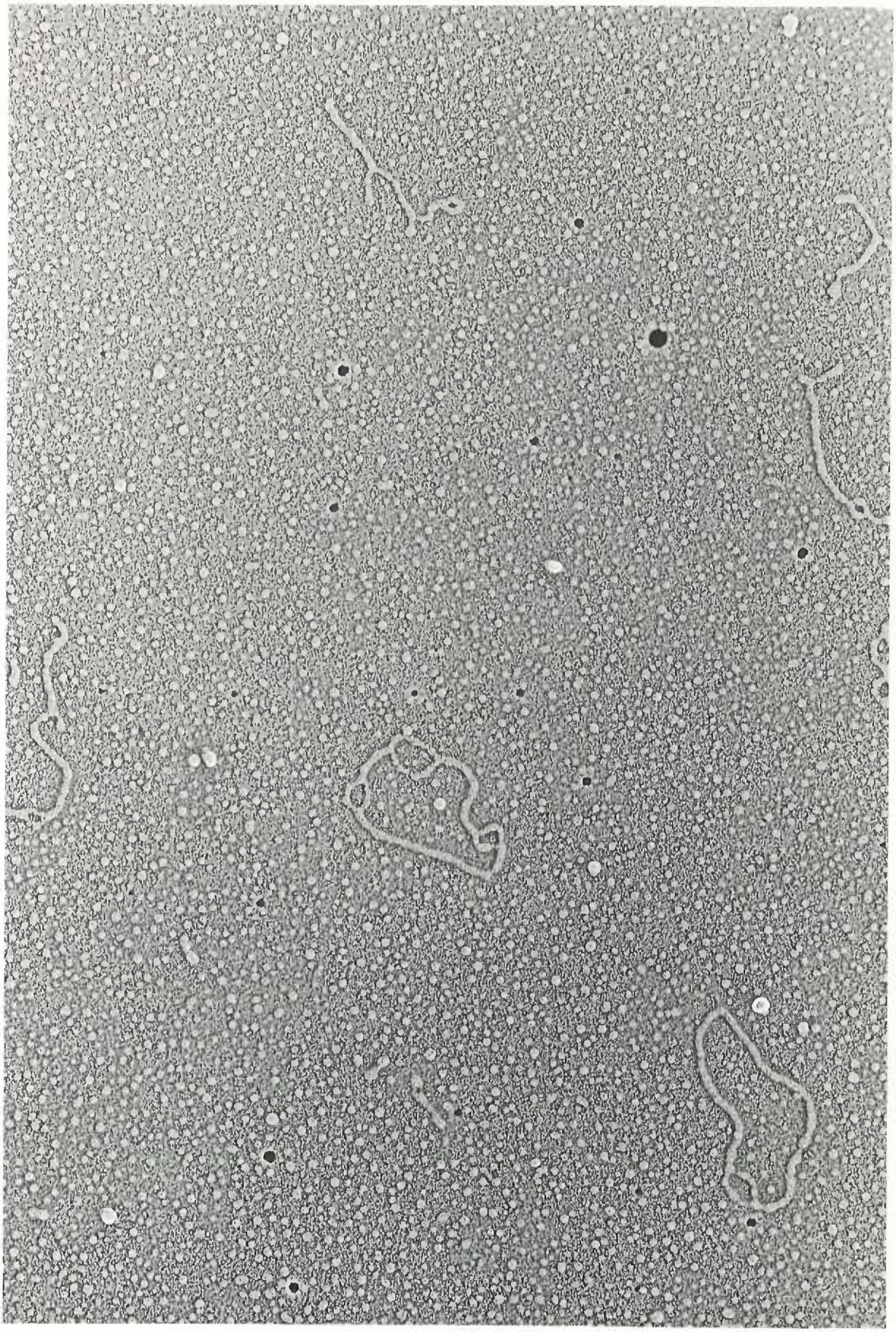
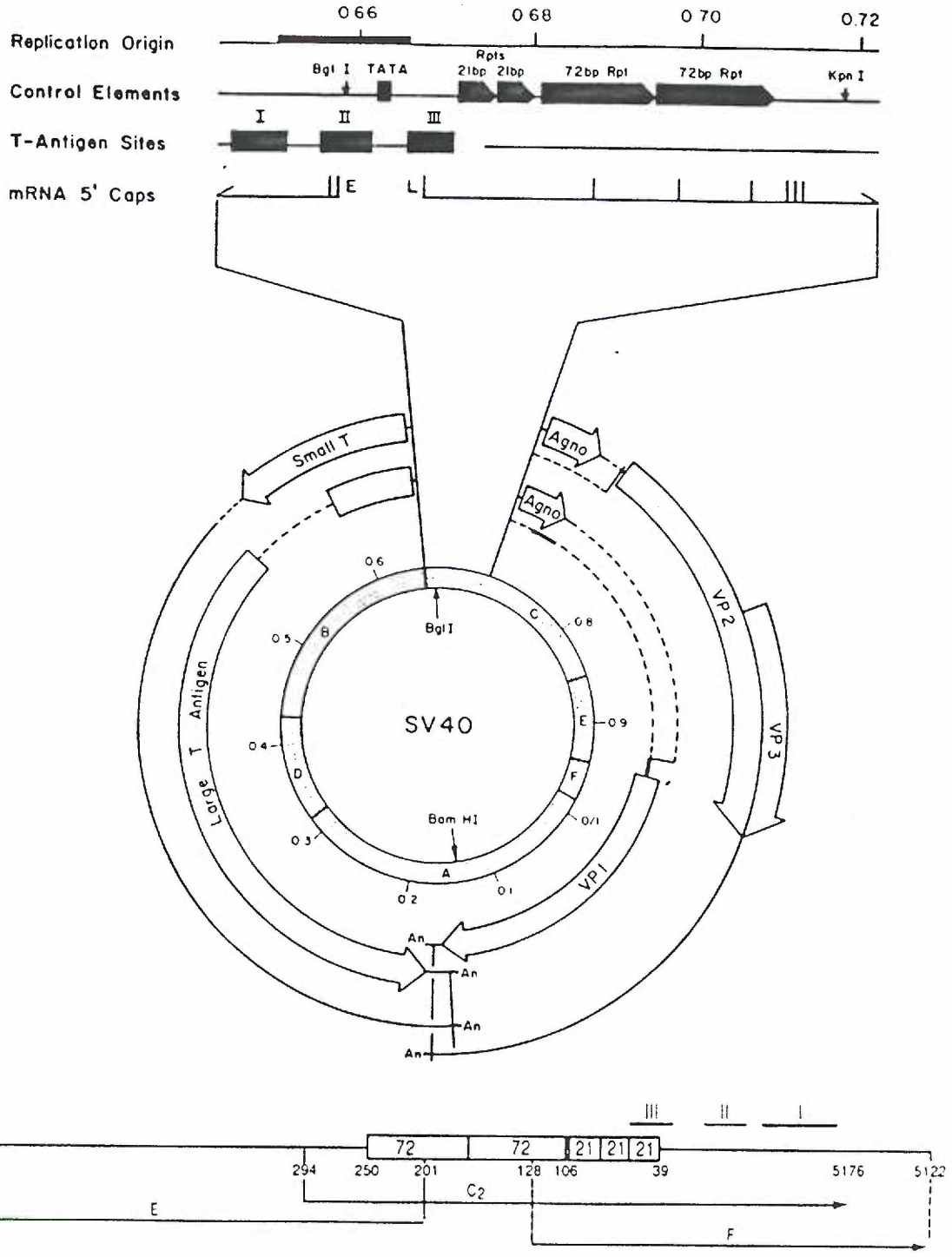


Figure 4. Structural Organization of the SV40 Genome. The major structural features of the genome as well as the recognition sites for Bgl I, Kpn I and Bam HI are indicated by map units. The six Hind III recognition sites are marked within the shaded circle for orientation. Early and late transcription proceeds on different strands in counterclockwise and clockwise directions, respectively. Small and large T antigens are coded for by the early strand, while the viral capsid proteins and agnoprotein are encoded on the late strand. Dotted lines indicate sequences which are spliced out during mRNA processing. A 400 base pair region which bridges the early and late mRNA start sites is expanded to show the regulatory features of the genome. This region includes the origin of replication (Bgl I site), the 21 bp repeats which are part of both the early and late promoter, 72 bp repeats which function as enhancers of transcription, TAg binding sites I, II and III, and the mRNA caps sites. The linear diagram at the bottom of the figure indicates nucleotide numbers for the boundaries of the enhancers, 21 bp repeats, T antigen binding sites, and some of the major restriction enzyme sites.


0.66
0.68
0.70
0.72
Replication Origin
Rpts
Bgl I
TATA
21bp
21bp
72bp Rpt
72bp Rpt
Kpn I
Control Elements
I
II
III
T-Antigen Sites
mRNA 5' Caps
E
L
Small T
Agno
Agno
VP2
VP3
VP1
Large T Antigen
SV40
Bgl I
Bam HI
0.6
0.5
0.4
0.3
0.2
0.1
0.8
0.9
0/1
A
B
C
D
E
F
An
An
An
An
III
II
I
72
72
21
21
21
560
294
250
201
128
106
39
5176
5122
C2
E
F

Manuscript 1.

The Site-Specific Inhibition of Bgl I Cleavage by Psoralen Photoadducts

ABSTRACT

We have investigated the site specificity of furocoumarins by using fluorescent densitometry to examine the frequency of cleavage by the restriction enzyme Bgl I. This enzyme has an 11 base pair (bp) recognition sequence which varies slightly from site to site because it includes a 5 base pair neutral region. Cleavage at all three Bgl I recognition sites in pBR322 was inhibited by the photoaddition of the psoralen derivative 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT) which forms both crosslinks and monoadducts in a dose-dependent manner. One site, which contains two thymidines in a crosslinkable configuration, was observed to be markedly more sensitive to HMT photoadducts. In contrast Bgl I cleavage at all sites was relatively resistant to the 5-methylisopsoralen (5-MIP), which forms only monoadducts. When HMT-reacted DNA was generated with widely different ratios of monoadducts to crosslinks (3% and 40% crosslinks), essentially the same level and pattern of inhibition was observed in both cases. Taken together, the data imply that differences in inhibition seen at the three cutting sites of Bgl I with HMT are attributable to DNA sequence and the role it plays in adduct positioning.

INTRODUCTION

Psoralen molecules provide a valuable tool for the study of *in vivo* structures because they readily penetrate intact cells and viruses without disruption of membranes or viral capsids, and can then bind preferentially to the linker region of nucleosomal DNA. This preferential binding has led to their wide use as probes of chromatin structure (Hanson *et al.*, 1976; Wiesehahn *et al.*, 1977; Cech and Pardue, 1977; Hallick *et al.*, 1978). These compounds intercalate between the base pairs (bp) of the DNA helix, and in the presence of long wavelength UV light become covalently bound to pyrimidines through a cyclobutane linkage. Many of the psoralen derivatives can form both monoadducts, by covalently binding to the pyrimidines of one strand, and crosslinks, by binding pyrimidines on opposite strands in adjacent positions (Cole, 1970; Krauch *et al.*, 1967; Musajo *et al.*, 1966; Pathak and Kramer, 1969; Straub *et al.*, 1981). It has been shown that greater than 90% of the adducts are formed with thymine (Straub *et al.*, 1981; Kanne *et al.*, 1982a; Kanne *et al.*, 1982b); however, it has not been clear whether the positioning of psoralen adducts in DNA was random with regard to DNA sequence. Carlson *et al.* (1982) have reported that psoralen adducts can block restriction enzyme sites in a preferential manner, although no correlation was made between a sequence at or near the sites and the cutting frequency observed. Recent work by Gamper *et al.* (1984) has shown that the 5'-TpA sequence is greatly favored over 5'-ApT for crosslink formation in DNA linker fragments using

4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT). Tessman et al. (1985) have made a similar observation with 8-methoxypsoralen.

The following experiments were designed to further investigate the site specificity of psoralen using the derivatives HMT and 5-methylisopsoralen (5-MIP). The structure of these compounds is shown in Fig. 1. In the presence of long-wavelength UV light, about 40% of the HMT adducts formed are crosslinks. In the case of 5-MIP, however, once a monoadduct has formed subsequent reaction of the remaining double bond with a pyrimidine on the opposite strand is not possible because the reactive bonds are not aligned (Bordin et al., 1979a; Bordin et al., 1979b). Fluorescent densitometry was used to examine preferential site-dependent cleavage by the restriction enzyme Bgl I. The preferences of this endonuclease are of special interest because it cuts at a large recognition site (11 bp) which contains a 5 bp neutral region (Table 1); thus, individual Bgl I recognition sites vary markedly in sequence. In these analyses the frequencies of Bgl I digestion at three well-separated sites on pBR322 DNA were calculated by quantitating the amount of DNA in each restriction fragment generated by partial digestion with Bgl I (Fig. 2). The same methodology was then used to compare the effects of HMT and 5-MIP adducts on the frequency of Bgl I cutting at each site following exhaustive digestion of the DNA.

We report here that HMT adducts block Bgl I in a dose-dependent manner. Maximum inhibition of Bgl I requires nearly one HMT adduct per recognition site with a strong preferential inhibition occurring at one of the three sites on pBR322 DNA. By comparing these results to similar

studies done with 5-MIP, we demonstrate that this preference appears to be due to both adduct type and selective adduct positioning. Finally, the effect of HMT and 5-MIP monoadducts was compared by using a high-intensity long-wavelength UV light source to generate samples with fewer than 3% crosslinks. The HMT monoadducts were then converted to crosslinks by a second broad-band UV irradiation, and the level of Bgl I inhibition was compared to that found in samples that did not receive a second dose of UV light.

MATERIALS AND METHODS

Enzymes and chemicals. Restriction endonucleases were purchased from Bethesda Research Laboratories, Inc. Agarose (Standard Low M_r, and Ultra Pure DNA grade) was obtained from Bio-Rad Laboratories. All analytical reagents were obtained from J. T. Baker Chemical Co. ^{3}H-4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT) and ^{3}H-5-methylisopsoralen (5-MIP) were synthesized as previously described (Isaacs et al., 1982; Isaacs et al., 1984) and had specific activities of 3.0×10^7, and 1.6×10^6 cpm/μg respectively as determined by spectrophotometric methods previously described (Robinson and Hallick, 1982).

Purification of pBR322 DNA. Supercoiled pBR322 DNA was obtained from E. coli strain RRI by a modification of the Holmes and Quigley rapid boiling method (Holmes and Quigley, 1981). The plasmid DNA was banded

twice in CsCl-ethidium bromide density gradients to obtain preparations with greater than 90% supercoiled molecules.

^{3}H-HMT and ^{3}H-5-MIP photoaddition. Samples containing 10-15 µg of Bam HI-linearized pBR322 DNA (prepared as described below) were diluted into 450 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.1). Various amounts of ^{3}H-HMT (range: 3.5 ng-10 µg) or ^{3}H-5-MIP (range: 5-18 µg) from aqueous dilutions were added and reaction volumes were adjusted to 500 µl with distilled water. The irradiation was done at 4°C for 2 hr by suspending 1.5 ml polypropylene tubes containing the samples between two General Electric F15T8 BLB fluorescent tubes (2.2-3.0 mW/cm^{2}/tube) spaced 7 cm apart. Samples were not purged of oxygen before irradiation. In cases where high numbers of psoralen adducts/bp of DNA were required, additional psoralen was added, and the samples reirradiated as above. Following irradiation, the salt concentration was raised to 0.25M with sodium acetate and each sample was extracted three times with chloroform-isoamyl alcohol (24:1) to remove noncovalently bound HMT or 5-MIP. The DNA was then precipitated at -20°C overnight in two volumes of ethanol. Control samples which were not irradiated following the addition of psoralen indicated that greater than 98% of noncovalently bound drug is removed by the above extraction procedure.

To prepare samples containing HMT primarily as monoadducts, 500 µg of linearized pBR322 DNA in 1100 µl of TE buffer was placed in a 1 X 1 X 5 cm quartz cuvette, cooled to 4°C and then irradiated in the presence of 40µg/ml ^{3}H-HMT with a 2500 W mercury-xenon light source. The incident light was filtered through a cobaltous nitrate liquid filter in line with

a 380 nm cut-off filter and a Baird-Atomic 389 nm interference filter with a band pass of 381 to 397 nm (10% of T_{max}). The average photon flux was 2 X 10^{15} photons/sec. Fresh aliquots of ^{3}H-HMT were added every 45-60 minutes. Three different time points (each representing one-third of the material) were taken and the DNA was extracted and precipitated as described above. The samples were resuspended in TE and aliquots from each sample were reirradiated in the absence of additional drug for 3h using the broad band UV irradiation device described initially. Samples were then ethanol precipitated, resuspended in TE and digested with an excess of Bgl I as described. Aliquots of each sample were removed for HPLC analysis to determine adduct type and level; values for the six samples (before and after reirradiation at three adduct levels) are given in Table 3.

Restriction enzyme digestion. Purified pBR322 DNA samples in TE buffer were digested with restriction enzymes according to the manufacturer's recommendations with the modification that Bgl I digestions were adjusted to 75 mM NaCl and 100 µg/ml sterile gelatin. All reactions were terminated by addition of tracking dye (40% sucrose, 10 mM Tris-HCl, pH 8.1, 200 µg/ml bromphenol blue, 10 mM EDTA, and 2% SDS).

Bam HI-linearized pBR322 DNA was prepared by digestion with an excess of enzyme for 4h at 37°C. The reaction was terminated by one buffer-saturated phenol and two ether extractions, then the DNA was ethanol precipitated and resuspended in TE buffer. Time-course digestions with Bgl I were carried out by incubating approximately 21 µg

of linear plasmid DNA with 60 units (U) Bgl I in a 200 µl reaction mixture; 25 µl aliquots were removed at the times indicated and the reaction was terminated as above.

Bam HI-linearized pBR322 DNA that had been photoreacted with ^{3}H-HMT or ^{3}H-5-MIP to various extents was also digested with Bgl I. For these samples 1 µg of photoreacted DNA was incubated for 4 hrs with 20 U Bgl I per 25 µl reaction mixture. These conditions were sufficient to completely digest four-fold higher amounts of non-photoreacted DNA. To insure that digestion was complete, aliquots of the above samples were taken, ethanol precipitated, resuspended and redigested with an excess of enzyme. The levels of digestion were then compared, and in all cases found to be essentially the same.

Gel electrophoresis. All DNA samples that were digested with Bgl I were subjected to electrophoresis on horizontal 0.75% agarose gels (34 by 18 by 0.3 cm, Canalco HSG Apparatus) in TBE buffer (0.089M Tris borate, 0.089M boric acid and 0.002M EDTA) containing 1 µg/ml ethidium bromide. The gels were submerged (3 mm buffer overlay) and electrophoresed at 90 volts for 18h. Gels were then photographed with a Polaroid camera, using a short-wave UV transilluminator (Ultra-Violet Products, Inc., San Gabriel, California) and scanned by direct densitometry.

Direct densitometric analysis. Ethidium bromide-stained gels were scanned in fluorescence mode using a Helena Quick Scan R and D densitometer, equipped with an extended red-range photomultiplier. Gels were supported in the instrument on a quartz tray (SDA 324, Schoeffel Instruments) above a mercury lamp (254 nm peak emission) that contained a

cobalt blue filter (to screen out wavelengths > 400 nm) and an excitation slit (7 mm by 0.7 mm). Ethidium bromide fluorescence was detected by the photomultiplier positioned directly over the gel, utilizing a second slit and a Wratten 23A filter (to screen out wavelengths < 585 nm). A black masking tray was used to prevent stray UV light from exciting gel lanes other than the one of interest. Individual gel lanes were aligned in the scanning path using a UV spot lamp (366 nm peak emission, UVL-56, Ultra Violet Products, Inc.). Peak areas from the chart paper were integrated using an electronic planimeter (Numonics, Lansdale, Pa).

Characterization of HMT photoadducts by HPLC. The percent of HMT present as furan-side HMT-thymidine monoadduct, pyrone-side HMT-thymidine monoadduct, and thymidine-HMT-thymidine crosslink was determined by HPLC analysis following enzymatic digestion of the DNA (Straub et al., 1981; Kanne et al., 1982; Tessman et al., 1985).

RESULTS

The densitometric method of DNA quantitation was utilized to examine preferential site-dependent cleavage by the restriction enzyme Bgl I. The sensitivity and linearity of the direct densitometric method will be reported elsewhere (Robinson et al., in preparation). These investigators demonstrated a linear relationship between peak area and DNA concentration over a 250-fold range (2-500 ng) of DNA. In restriction enzyme digestions of known DNAs, the DNA distribution

determined by densitometric analysis closely approximated expected values.

To carry out our experiments, it was first necessary to generate formulas to calculate the cutting (or noncutting) frequency at each Bgl I recognition site on pBR322. The general statement of this formula is:

$$\% \text{ uncut } X = \frac{Y}{Z} = \frac{(\text{total molecules containing uncut fragments})}{(\text{total pBR322 molecules measured})}$$

where X = site I, II or III. From Fig. 2 it can be seen that Bgl I digestion of pBR322 DNA linearized by Bam HI treatment can produce 10 possible fragments, four of which are limit digest fragments. The relative molar amount for each fragment was determined by scanning the gel, integrating the area under the curve representing each band and dividing by the number of base pairs in that DNA fragment. The numerator Y for the above formula was then calculated separately for each of the Bgl I sites by using the following summations:

(site I) CD + CDA + CDAB

(site II) CDA + DA + DAB + CDAB

(site III) AB + DAB + CDAB

Thus, the frequency with which Bgl I fails to cleave site I (which lies between fragments C and D) will be represented by the molar sum of all fragments generated when there is no cleavage between C and D. The

denominator Z for the above formula was the same for each of the sites and was calculated by taking the average of the four summations:

$$C + CD + CDA + CDAB$$

$$D + CD + DA + DAB + CDAB$$

$$A + AB + CDA + DA + CDAB$$

$$B + AB + DAB + CDAB$$

The percent of DNA cleaved was obtained by subtracting the calculated % uncut from 100. Formulas similar to those described above could have been used to calculate directly the percent of DNA cleaved, but these computations would have depended heavily on information derived from the DNA fragments CD, C and D which are more easily lost from the gel because of their small size.

In the first series of experiments, Bam HI-linearized pBR322 DNA was cut with an excess of Bgl I. At the time points indicated in Fig. 3, aliquots of the digestion mixture were removed, the reaction was stopped, and the samples were electrophoresed as described in Materials and Methods. As can be seen from the data presented in Fig. 4, the rate of cutting by Bgl I was essentially identical at sites I and II, but slightly faster at site III.

In a second set of experiments, we investigated the inhibition of Bgl I cleavage by HMT photoadducts on the DNA. The reaction conditions and analyses were similar to those described above for the kinetic experiments with non-photoreacted DNA. Samples were prepared with

increasing ratios of HMT per base pair. Following extensive digestion with Bgl I, samples were electrophoresed, gels were scanned and the percent of DNA molecules uncut at each site was calculated. As can be seen from Figs. 5 and 6, cleavage at all sites was inhibited in a dose-dependent manner. The rate of inhibition was very similar for sites II and III, but was dramatically higher for site I. In DNA samples with saturating levels of HMT (1 per 4 base pairs), cleavage at site I was completely blocked, whereas it was only 50% inhibited at sites II and III. To ascertain that complete digestion had actually occurred under these reaction conditions, particularly in the samples containing high levels of HMT adducts, control experiments were carried out in which samples of linearized pBR322 DNA containing 1 HMT molecule per 4 base pairs were reacted with Bgl I for 3 hours, more enzyme was added, and incubation was continued for an additional 3 hours. When the results were compared, it was found that no additional cutting occurred after the first 3 hours (data not shown).

In order to determine whether the increased inhibition at site I was due to greater inhibition by crosslinks than by monoadducts, an analogous set of experiments with 5-MIP was performed (Fig. 7). Levels of 57, 85 and 91 5-MIP monoadducts/1000 base pairs were achieved in three unique samples. In each case, and at all three sites, the level of inhibition was clearly less than that seen with the corresponding HMT sample (Table 2 and Fig. 7). In addition, there was no significant difference in the levels of inhibition observed among the three sites with 5-MIP. This might indicate that the 5-MIP monoadduct perturbs the DNA structure less

than either the HMT crosslink or the HMT monoadducts. Alternatively, it may suggest that 5-MIP does not "fill up" available sites in the DNA in the same order as does HMT (i.e. 5-MIP has different site preferences). In the latter case, only at extremely high levels of 5-MIP adduct formation (approaching 250 adducts/1000 base pairs) would comparably high inhibition levels be seen. Unfortunately it was not possible to achieve such a high adduct load with 5-MIP, despite repeated efforts with multiple drug additions and reirradiation protocols.

Finally, in order to assess the role of HMT monoadducts on Bgl I inhibition, the light source and wavelength were manipulated to minimize the percent total crosslinks. Using a high-intensity long-wavelength UV light source providing incident light between 381-397 nm, levels of 86.2, 142.8, and 149.2 HMT adducts/1000 base pairs were achieved. The fraction of total HMT adducts found as crosslinks was determined by HPLC, and found to be between 2% and 4% for these three samples. The samples were then divided and one portion was reirradiated with the broad band long-wavelength black light (approximately 320-380nm) utilized in the previous experiments in order to drive the monoadducts into crosslinks. The percent of crosslinks in the reirradiated samples was between 40 and 50%. To our surprise, the level of Bgl I inhibition at each of the three sites, for any sample, was virtually identical to the corresponding sample that had not been reirradiated and contained only 2-4% crosslinks (Table 3). Again, a similar pattern of high inhibition at site I, with low levels at sites II and III, was observed. The extent of inhibition in both samples, at each adduct level, actually exceeded that found when

similar crosslink and adduct loads were achieved using the black light alone. The reason for this enhanced inhibition when the samples are irradiated with a monochromatic light source is unknown, but may be due to an altered distribution of adduct isomers, an altered sequence specificity at the long wavelength, or alternative types of photodamage which are associated with long wavelength irradiation (Pathak, 1984).

DISCUSSION

Utilizing the direct fluorescent method for DNA quantitation, we have measured the relative efficiency of cutting by the enzyme Bgl I at the three recognition sites of pBR322. Bam HI-linearization of the plasmid allows the initial Bgl I cut on the plasmid to be assigned unambiguously to one of the three sites, a feature that would not have been possible if a circular plasmid had been used. Control experiments in the absence of psoralens demonstrated a slight but reproducible preference for cleavage at site III (Figs. 3 and 4). Greater site preferences have been observed for other restriction enzymes such as Pst I on pSM I DNA (Armstrong and Bauer, 1982), Eco RI on λ DNA (Thomas and Davis, 1975), Hinc II on pBR322 DNA (Malcolm and Moffatt, 1981) and Hinf I on pBR322 DNA (Armstrong and Bauer, 1983) where the cutting ratios of most to least preferred sites ranged from 10 to 4. Additionally, Alves et al. (1984) have determined that nucleotide sequences adjacent to the recognition site influence the rate of cleavage by the restriction enzyme Eco RI of two decadeoxynucleotides. Armstrong and Bauer (1982 and 1984)

reached a similar conclusion and also suggested that the proximity of a given restriction site to the termini of the linear DNA molecule might influence the probability of cleavage. Drew and Travers (1985) have recently reported a correlation between flanking sequences and the probability of cleavage by the restriction enzymes Fnu DII and Hha I (recognition sequences CGCG and GCGC respectively). The favored configuration of the flanking nucleotides in both cases appeared to be a 5' pyrimidine and a 3' purine.

When linearized pBR322 DNA was photoreacted with the psoralen derivative HMT to different levels before exhaustive digestion with Bgl I, sites II and III displayed comparable inhibition profiles, whereas site I was much more strongly inhibited (see Figs. 5 and 6). Interestingly, both sites II and III have similar inhibition profiles, yet site II has one and site III has two thymidine residues (Table 1). Sites I and III both contain two thymidine residues at nearly the same location, but their inhibition profiles are very different. The formation of thymidine-HMT-thymidine crosslinks is only possible within site I where two thymidines are at adjacent and opposite positions, whereas only thymidine monoadducts can form in Sites II and III.

Recent work by Gamper et al. (1984) using Kpn I (CGGTACCG) and Bam HI (CGGATCCG) linkers with HMT has shown that the 5'-TpA sequence is favored for photoaddition approximately 100-fold over the 5'-ApT sequence. In contrast, when SV40 rather than purified linker DNA was utilized, they observed only a 10-fold difference in the inhibition of cleavage between the single Kpn I and Bam HI sites. Interestingly, HPLC

analysis on the linker samples indicated that in the case of the Kpn I mixture, 63% of the DNA molecules had a crosslink, 6% had an HMT monoadduct and 30% were unreacted. In the Bam HI reaction, 6% of the DNA linkers had a crosslink, 35% contained a monoadduct, and 59% were unreacted. This suggests that the GATC sequence is less reactive than the GTAC for overall HMT photoaddition as well as for crosslink formation. In our experiments increasing differences in the inhibition level between site I and sites II or III appear as the adduct level increases. None of the three sites contains the 5'-TpA sequence within the recognition region. We conclude that preferential photoaddition occurs at other sites such as the CATG sequence of site I.

Carlson et al. (1982) have also analyzed the effect of psoralen adducts on restriction site inhibition using the psoralen derivative 4,5',8-trimethylpsoralen and a variety of restriction enzymes. They report that the enzyme Hind III is extremely sensitive to low levels of adducts (approximately 50% inhibition at 5 adducts/1000bp) relative to the enzyme Eco RI which they indicate has an average sensitivity compared to other enzymes. In their analysis, individual Hind III sites exhibited different rates of inactivation which were ascribed to variations in the DNA sequences adjacent to the recognition sites. We suggest, however, that for the enzyme Bgl I the enhanced sensitivity of cleavage at site I is due to sequences within or very near the recognition site, since little preferential inhibition is observed until a photoaddition level of 1 adduct in 25-30 base pairs is achieved. This argument is supported by the fact that both sites I and II contain 5'-TpA sites directly outside

the 11 base pair canonical sequence whereas site III contains no TpA or ApT sites within 10 base pairs on either side of the canonical sequence.

The above experiments were repeated with the psoralen derivative 5-MIP, which lacks the ability to form crosslinks with DNA. Although the levels of 5-MIP/base pair were not as high as those achieved with the most heavily labeled samples in the HMT experiments (Table 2), it was nevertheless clear that the levels of inhibition were dramatically decreased over those seen with the corresponding HMT samples in all cases. Three explanations might account for this difference. Monoadducts *per se* may be less inhibitory to the restriction enzyme than crosslinks. Recent data, in fact, by Pearlman *et al.* (1985) have shown that the unwinding of the B form DNA helix due to a psoralen-crosslink is about 87.7°. This is considerably greater than the unwinding due to a 5-MIP monoadduct which has been shown to be 18 ± 3° (Isaacs *et al.*, 1984). Alternatively, since the monoadduct formed with 5-MIP is structurally distinct from that formed with HMT, the enzyme may be less affected. Finally, the site preferences observed with HMT may not apply to 5-MIP.

In order to determine the relative contribution of monoadducts and crosslinks to Bgl I inhibition, samples containing HMT primarily in the monoadduct form were generated with a monochromatic light source. These samples were divided and one portion reirradiated which increased the percent of crosslinks 10-20 fold. Following Bgl I digestion and gel analysis, it was clear that the levels of inhibition were virtually identical in the corresponding samples despite the dramatic increase in

the proportion of crosslinks following reirradation. In both cases, the pattern of high inhibition at site I, with significantly lower levels of inhibition at sites II and III, was maintained. This surprising result suggests either that HMT monoadducts have the potential to make a significant contribution to Bgl I enzyme inhibition, or that highly preferential crosslink sites filled during the initial irradiation have a profound effect on the structure of site I.

In conclusion, it is apparent that the differences in the level of inhibition at the three sites can be attributed to DNA sequence and the role it plays in selection of adduct position. In addition, different furocoumarins may have altered sequence specificities. Consistent with this prediction is the observation that the level of inhibition at each of the three sites in the samples generated with 5-MIP is essentially the same. The relatively low levels of inhibition that resulted from photoreacting the DNA with 5-MIP imply that monoadducts generated by different furocoumarin derivatives either bind to different sequences or have different structural effects on the DNA molecule itself. It will be necessary to map the adduct positions at the nucleotide level to distinguish between these two alternatives.

ACKNOWLEDGEMENTS

We would like to thank Sophia Kondoleon and Gary Wiesehahn for their assistance and suggestions during the course of the work, and John E. Hearst, in whose laboratory the high intensity irradiation and the HPLC

analyses were performed. This project was supported by PHS grants CA 24799 and GM 25151, PHS training grant GM 07336, by Tartar Fellowships to G.W.R. and E.A.O. and in part by BRSG Award 507 RR05412.

REFERENCES

Alves, J., A. Pingoud, W. Haupt, J. Langowski, F. Peters, G. Maass, and C. Wolff (1984) The influence of sequences adjacent to the recognition site on the cleavage of oligodeoxynucleotides by the EcoRI endonuclease. Eur. J. Biochem. 140, 83-92.

Armstrong, K., and W.R. Bauer (1982) Preferential site-dependent cleavage by restriction endonuclease Pst I. Nucl. Acids Res. 10, 993-1007.

Armstrong, K., and W.R. Bauer (1983) Site-dependent cleavage of pBR322 DNA by restriction endonuclease Hinf I. Nucl. Acids Res. 11, 4109-4126.

Bordin, F., F. Carlassare, F. Baccichetti, A. Guiotto, and P. Rodighiero (1979a) 5-methylangelicin: A new highly photosensitizing angular furocoumarin. Experientia 35, 1567.

Bordin, F., F. Carlassare, F. Baccichetti, A. Guiotto, P. Rodighiero, D. Vedaldi, and F. Dall'Acqua, (1979b) 4,5'-Dimethylangelicin: A new DNA-photobinding monofunctional agent. Photochem. Photobiol. 29, 1063.

Carlson, J.O., O. Pfenninger, R.R. Sinden, J.M. Lehman, and D.E. Pettijohn (1982) New procedure using a psoralen derivative for analysis of nucleosome associated DNA sequences in chromatin of living cells. Nucl. Acids Res. 10, 2043-2063.

Cech, T., and M.L. Pardue (1977) Cross-linking of DNA with trimethylpsoralen is a probe for chromatin structure. Cell 11, 631-640.

Cole, R.S. (1970) Light-induced cross-linking of DNA in the presence of a furocoumarin (psoralen). Biochim. Biophys. Acta. 217, 30-39.

Drew, H.R., and A.A. Travers (1985) Structural junctions in DNA: The influence of flanking sequence on nuclease digestion specificities. Nucl. Acids Res. 13, 4445-4467.

Gamper, H., J. Piette, and J.E. Hearst (1984) Efficient formation of a crosslinkable HMT monoadduct at the Kpn I recognition site. Photochem. Photobiol. 40, 29-34.

Hallick, L.M., H.A. Yokota, J.C. Bartholomew, and J.E. Hearst (1978) Photochemical addition of the cross-linking reagent 4,5',8-trimethylpsoralen (trioxsalen) to intracellular and viral Simian virus 40 DNA - histone complexes. J. Virol. 27, 127-135.

Hanson, C.V., C.-K.J. Shen, and J.E. Hearst (1976) Cross-linking of DNA in situ as a probe for chromatin structure. Science 193, 62-64.

Holmes, D.S., and M. Quigley (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114, 193-198.

Isaacs, S.T., H. Rapoport, and J.E. Hearst (1982) Synthesis of deuterium and tritium labelled psoralens. J. Labeled Compounds and Radiopharmaceuticals 19(3), 345-356.

Isaacs, S.T., G.P. Wiesehahn, and L.M. Hallick (1984) In vitro characterization of the reaction of four psoralen derivatives with DNA. J. National Cancer Institute 66, 21-30.

Kanne, D., K. Straub, H. Rapoport, and J.E. Hearst (1982a) Isolation and characterization of pyrimidine psoralen diadducts from DNA. Biochemistry 21, 861-871.

Kanne, D., K. Straub, H. Rapoport, and J.E. Hearst (1982b) Isolation and characterization of pyrimidine psoralen-pyrimidine photoadducts from DNA. J. Am. Chem. Soc., 104, 6754-6764.

Krauch, C.H., D.M. Kramer, and A. Wacker (1967) Zum wirkungsmechanismum photodynamischer furocumarine. Photoreaktion von psoralen - (4-^{14}C) mit DNS, RNS, homopolynucleotiden und nucleosiden. Photochem. Photobiol. 6, 341-354.

Malcolm, A.D.B. and J.R. Moffatt (1981) Differential reactivities at restriction enzyme sites. Biochim. Biophys. Acta. 655, 128-135.

Musajo, L., G. Rodighiero, A. Breccia, F. Dall'Acqua, and G. Malesani, (1966) Skin-photosensitizing furocoumarins: Photochemical interaction between DNA and O-$^{14}CH_3$ bergapten (5-methoxy-psoralen). Photochem. Photobiol. 5, 739-745.

Pathak, M.A. and D.M. Kramer (1969) Photosensitization of skin in vivo by furocoumarins (psoralens). Biochim. Biophys. Acta. 195, 197-206.

Pathak, M.A. and J.C. Prakashi (1984) Production of active oxygen species ($^1O_2$ and O_2) by psoralens and ultraviolet radiation (320-400nm). Biochim. Biophys. Acta. 798, 115-126.

Pearlman, D.A., Holbrook, S.R., Pirkle, D.H., and S-H. Kim (1985) Molecular models for DNA damaged by photoreaction. Science 227, 1304-1308.

Robinson, G.W. and L.M. Hallick, (1982) Mapping the in vivo arrangement of nucleosomes on Simian virus 40 chromatin by the photoaddition of radioactive hydroxymethylpsoralen. J. Virol. 41, 78-87.

Straub, K., D. Kanne, J.E. Hearst, and H. Rapoport (1981) Characterization of pyrimidine-psoralen photoadducts from DNA. J. Am. Chem. Soc. 103, 2347-2355.

Tessman, J.W., S.T. Isaacs, and J.E. Hearst (1985) Photochemistry of the furan-side 8-methoxypsoralen-thymidine monoadduct inside the DNA helix. Conversion to diadduct and to pyrone-side monoadduct. Biochemistry 24, 1669-1676.

Thomas, M. and R.W. Davis (1975) Studies on cleavage of bacteriophage lambda DNA with EcoRI restriction endonuclease. J. Mol. Biol. 91, 315-328.

Wiesehahn, G.P., J.E. Hyde, and J.E. Hearst (1977) The photoaddition of trimethylpsoralen to Drosophila melanogaster nuclei: A probe for chromatin substructure. Biochemistry 16, 925-932.

TABLE 1

Nucleotide Sequences Adjacent to Bgl I Sites of pBR322 DNA

Site Number	Nucleotide Sequence
Canonical	5'-GCCNNNNN↓GGC-3'
	3'-CGGNNNNN↑CCG-5'
I	GGCCATTATC-GCCGGCATGGC-GGCCGACGCG
	CCGGTAATAG-CGGCCGTACCG-CCGGCTGCGC
II	GGCGATTTAT-GCCGCCTCGGC-GAGCACATGG
	CCGCTAAATA-CGGCGGAGCCG-CTCGTGTACC
III	AAACCAGCCA-GCCGGAAGGGC-CGAGCGCAGA
	TTTGGTCGGT-CGGCCTTCCCG-GCTCGCGTCT

TABLE 2

Inhibition of Bgl I Cleavage by Psoralen Monoadducts

Adducts/ 1000 bp	Psoralen Derivative	% Inhibition [a]		
		I	II	III
57	5-MIP	3.64 ± 1.0	5.1 ± 2.6	2.2 ± .9
57	HMT	24	14	10
85	5-MIP	8.9 ± 1.4	9.2 ± 3.0	7.3 ± 1.9
85	HMT	35	20	15
91	5-MIP	10.0 ± 2.8	10.1 ± 3.1	8.9 ± 3.9
91	HMT	35	21	16

[a]The % inhibition (% uncut) for the 5-MIP samples is computed as described in the text. At least 4 gels were analyzed for each sample. The % inhibition for the HMT samples is interpolated from Figure 6 for the appropriate adduct level and is included for comparison. Irradiation conditions are specified in Materials and Methods.

[a]Samples were irradiated at 381-397 nm with multiple additions of drug as described in Materials and Methods. They were then divided and an aliquot of each was reirradiated at 320-380 nm. The third entry in each set is from the experiment described in Figure 6 in which samples were irradiated one time with broad range UV light.

[b]Samples were analyzed by HPLC as described in the Materials and Methods. Sample recovery from HPLC analysis ranged from 87 to 95%. The percent pyrone-side monoadducts before reirradiation ranged from 2.4-2.5 and increased to 3-6.5% following reirradiation. The percent thymidine furan-side monoadducts prior to reirradiation ranged from 60-68.5% and decreased to a level of 17-21% in proportion to the increase in crosslinks after reirradiation. The level of cytidine furan-side monoadducts was 22-28% prior to reirradiation and decreased to 12-19% following reirradiation.

[c]The % inhibition (% uncut) is computed as described in the text. At least 4 gels were analyzed for each sample. The third entry in each set is data interpolated from Figure 6 for the appropriate adduct level and is included for comparison.

[d]A single sample prepared in a similar manner was analyzed by HPLC and found to contain 42% crosslinks.

[e]The adduct level was not redetermined on these samples. Based upon other experiments, it may have been reduced by approximately 5% due to photoreversal during the second irradiation (Isaacs et al., 1984).

TABLE 3

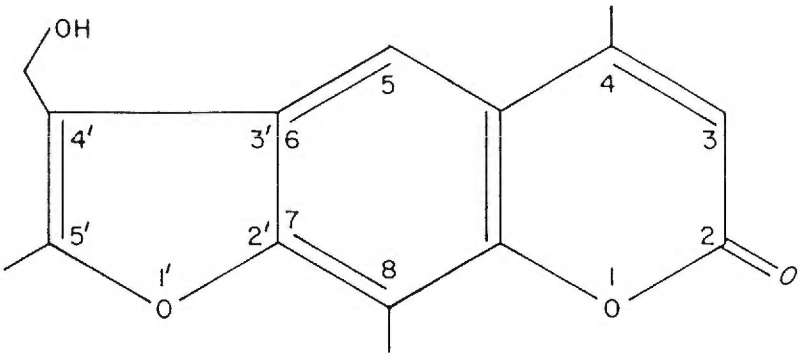
Inhibition of Bgl I Cleavage by

HMT Adducts: Monoadducts vs. Crosslinks

Adducts/ 1000 bp	Irradiation Wavelength[a]	% X-links[b]	% Inhibition[c] I	II	III
86.2	381-397	2.0	62.4±3.7	39.6±2.3	38.1±5.0
86.2[e]	381-397/320-380	42.6	64.6±4.3	42.6±4.8	32.2±2.7
86.2	320-380	d	35	20	15
142.8	381-397	2.3	84.4±2.6	53.0±4.5	55.8±4.8
142.8[e]	381-397/320-380	51.7	96.8±7.9	64.0±7.7	51.1±7.9
142.8	320-380	d	59	32	25
149.2	381-397	3.8	89.3±6.0	61.5±5.2	63.7±3.8
149.2[e]	381-397/320-380	43.1	94.7±6.8	61.3±8.0	52.8±9.7
149.2	320-380	d	61	34	26

Figure 1. Chemical structures of the psoralen derivative 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT) and the isopsoralen derivative 5-methylisopsoralen (5-MIP) are shown. Absorption of 320-380 nm light results in covalent photocycloaddition of the 4',5'- or the 3,4-double bond (furan-side or pyrone-side respectively) to the 5,6-double bond of a pyrimidine.

4′–Hydroxymethyl–4,5′,8–Trimethylpsoralen (HMT)



5–Methylangelicin (5–MIP)

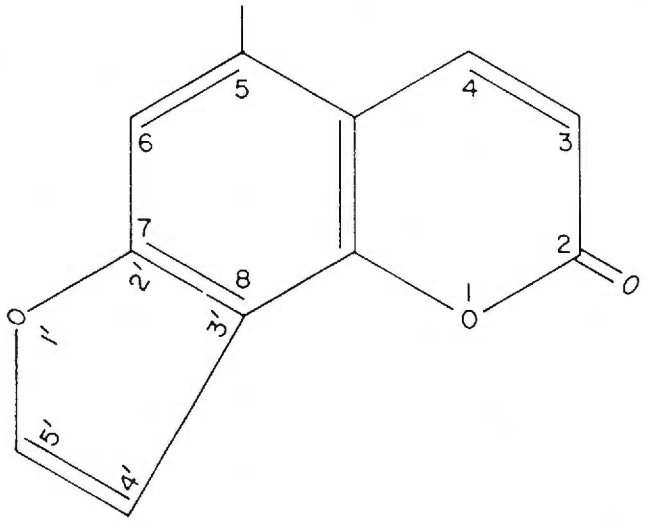


Figure 2. Bgl I cleavage diagram of Bam HI-linearized pBR322 DNA indicating the four limit (*) and six non-limit digest fragments resulting from incomplete digestion with Bgl I. The three cutting sites of Bgl I are designated I, II, and III. All DNA fragments are lettered according to the limit digest fragments that they contain, and are numbered to indicate base pairs of DNA (e.g. linear pBR322 is fragment CDAB and contains 4362 bp).

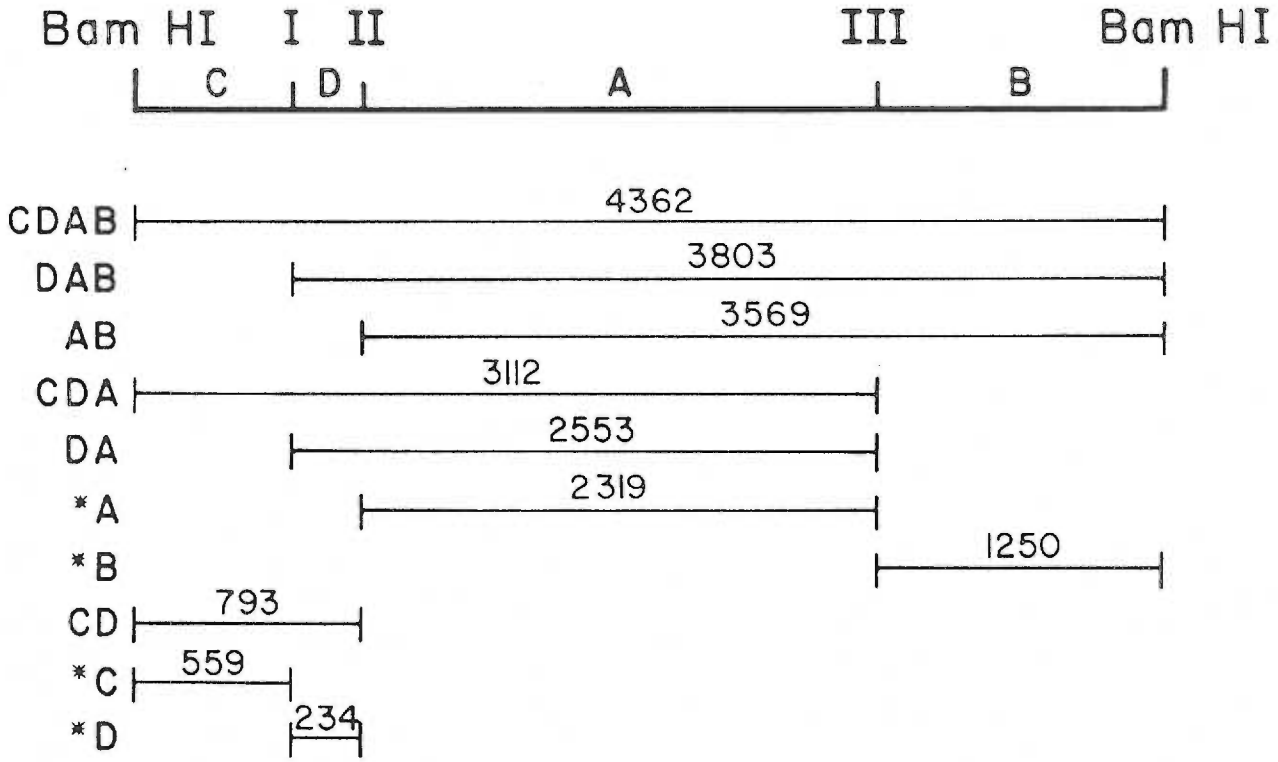

Bam HI
I
II
III
Bam HI
C
D
A
B
CDAB
4362
DAB
3803
AB
3569
CDA
3112
DA
2553
*A
2319
*B
1250
CD
793
*C
559
*D
234

Figure 3. Time course of Bgl I digestion of Bam HI-linearized pBR322 DNA. At the times indicated 3μg aliquots of the Bgl I digestion mixture were removed, the reactions were terminated, and 1.5μg samples were electrophoresed in duplicate on agarose gels as described in Materials and Methods. The 120 minute sample was a limit digest sample from this reaction. Lettered fragments refer to those of Figure 2. The time of each digestion in minutes is indicated above the corresponding lanes of the gel.

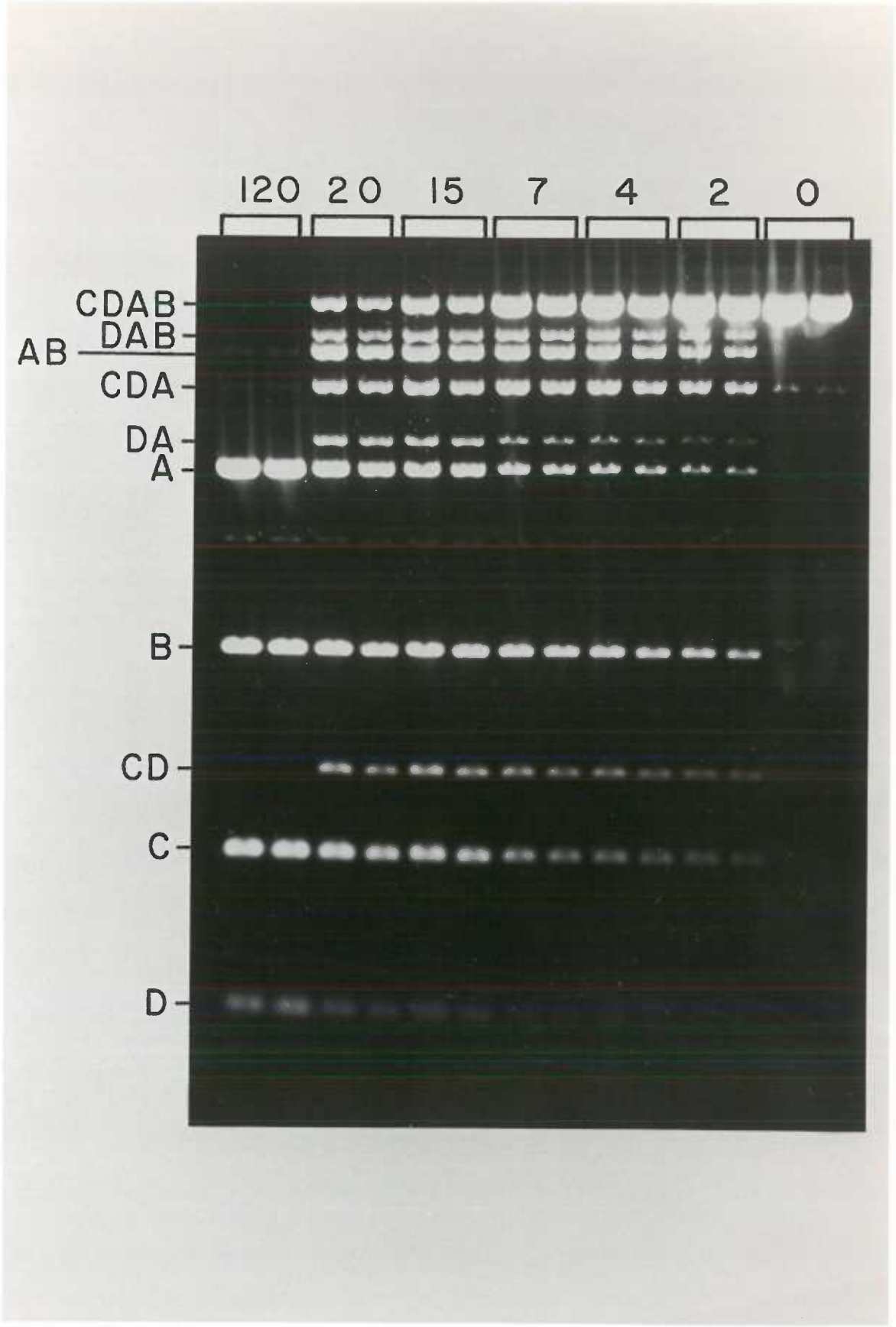

120
20
15
7
4
2
0
CDAB
DAB
AB
CDA
DA
A
B
CD
C
D

Figure 4. Kinetic analysis of Bgl I cleavage at the three cutting sites present on Bam HI-linearized pBR322 DNA. The percent of sites uncut by Bgl I at different times was determined by measuring amounts of DNA present in different bands from an agarose gel (Fig. 3) and by applying the formulas described in the text. Bgl I sites I(Δ), II(●), and III(0) are indicated. Points represent averages of two determinations and were fit to the lines by linear regression analysis with r-values of 0.999, 0.999, and 0.998 respectively.

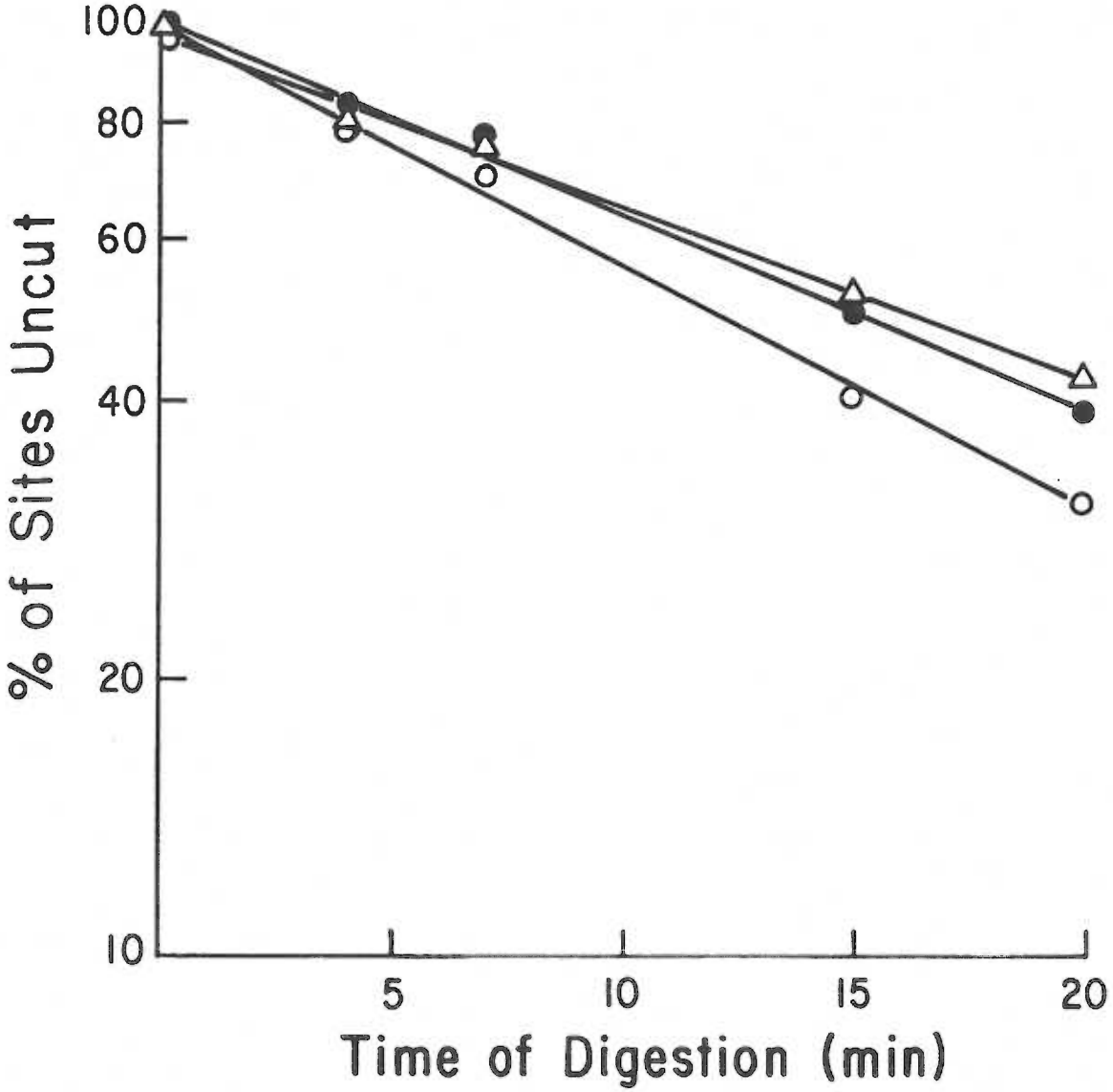

100
80
60
40
20
10
% of Sites Uncut
5
10
15
20
Time of Digestion (min)

Figure 5. Inhibition of Bgl I cleavage by Bam HI-linearized pBR322 DNA photoreacted with HMT. Plasmid DNA samples that contained covalently bound HMT to the extent of a)0; b)4; c)8; d)10; e)24; f)42; g)83; h) and i)250 molecules per 1000 bp of DNA were digested with Bgl I and electrophoresed on agarose gels as described in Materials and Methods. Typically, 1μg of DNA was digested at 37°C for 4 hrs with 20 U of Bgl I per 25 μl reaction mixture. Two samples of non-psoralen reacted pBR322 DNA digested with one-fourth the enzyme used above are also shown (a,j). Lettered fragments refer to those in Figure 2. Photoreaction conditions are described in Materials and Methods.

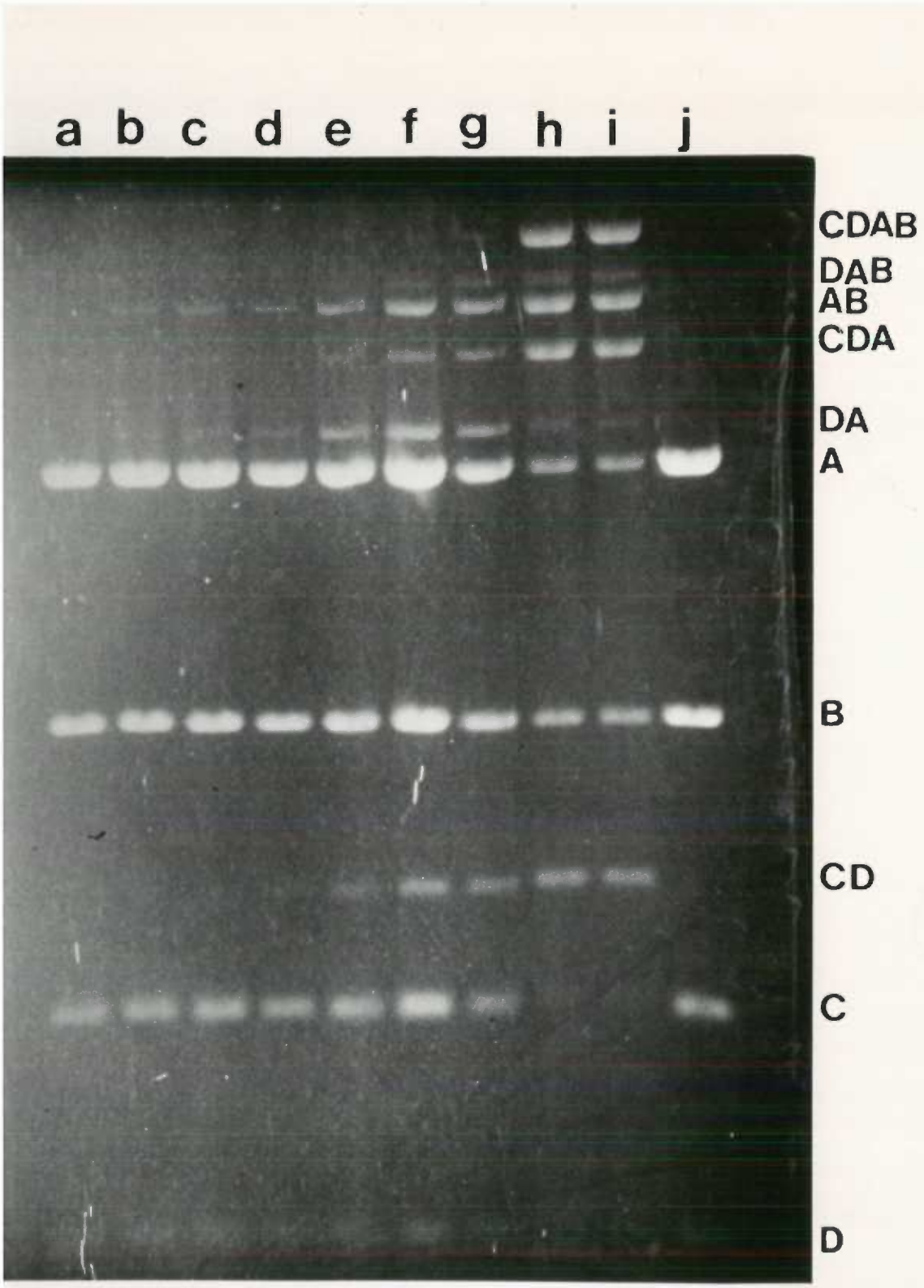

a b c d e f g h i j
CDAB
DAB
AB
CDA
DA
A
B
CD
C
D

Figure 6. Inhibition profiles showing Bgl I cleavage of HMT-photoreacted Bam HI-linearized pBR322 DNA. Data plotted was that obtained from fluorescence analysis of an agarose gel (Figure 5) as described in Results. Bgl I sites I (Δ), II (●), and III(0) are indicated. Points were fit to the lines by linear regression analysis and had r-values of 0.994, 0.994, and 0.990 respectively.

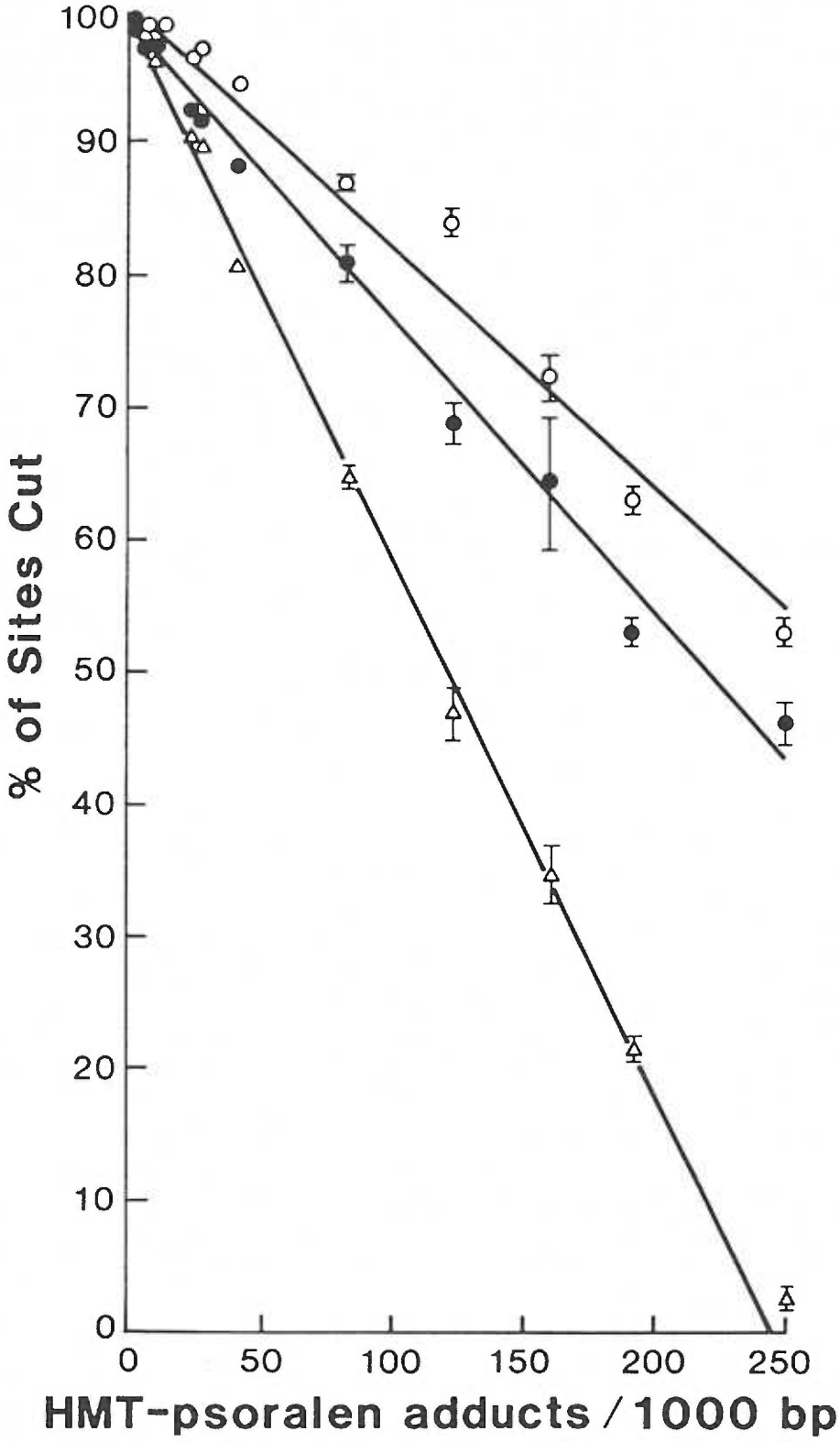
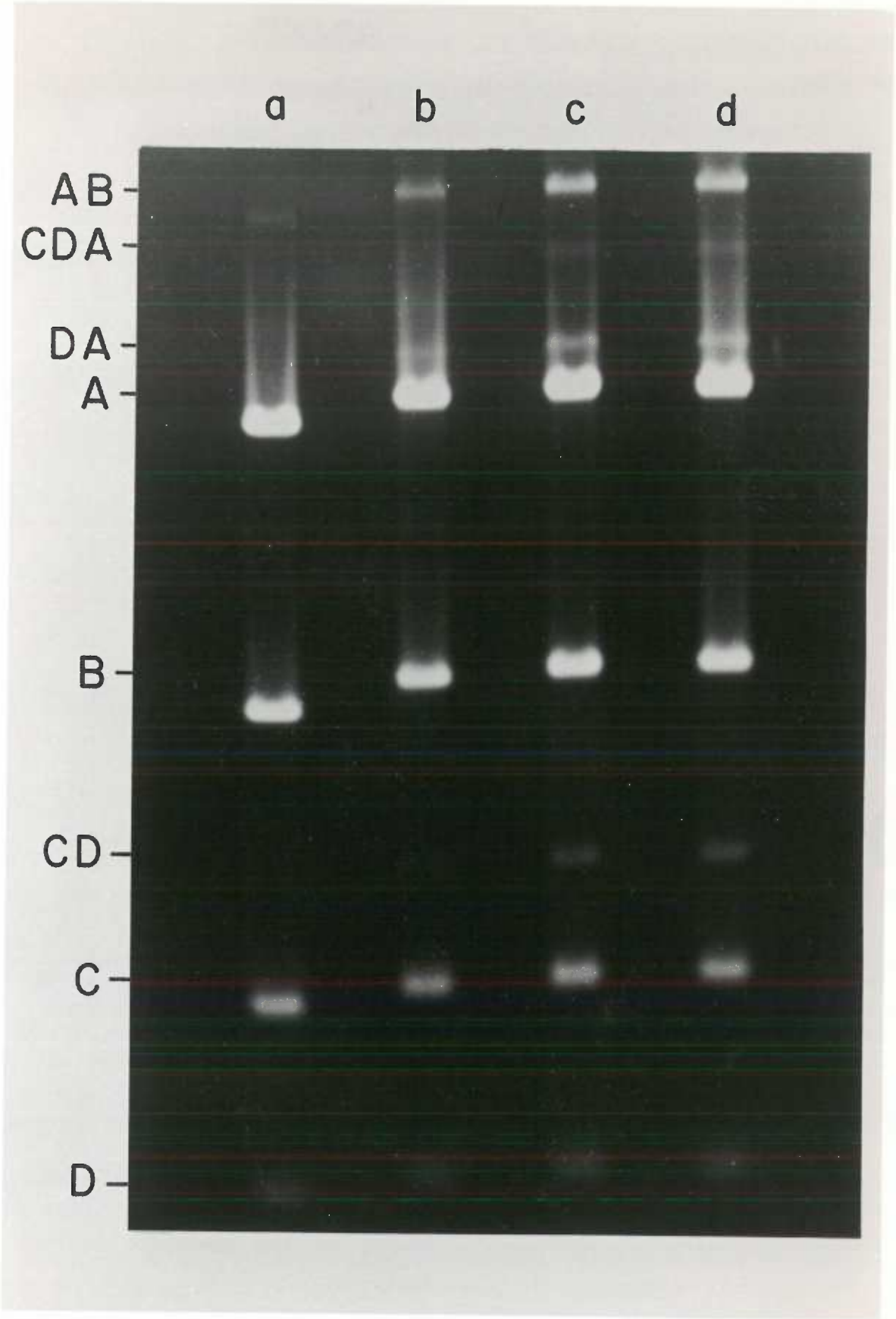

100
90
80
70
60
50
40
30
20
10
0
% of Sites Cut
0
50
100
150
200
250
HMT-psoralen adducts / 1000 bp

Figure 7. Inhibition of Bgl I cleavage by Bam HI-linearized pBR322 DNA photoreacted with 5-MIP. Plasmid DNA samples containing covalently bound 5-MIP to the extent of a)0, b)57, c)85, and d)91 molecules/1000 bp of DNA were digested with Bgl I and electrophoresed on horizontal agarose gels as described in Materials and Methods. Typically, 1μg of DNA was digested for 4 hours at 37°C with 20 U of Bgl I per 25μl reaction mixture. Lettered fragments refer to those in Fig. 2.


a
b
c
d
AB
CDA
DA
A
B
CD
C
D

Manuscript 2:

Fine Structure Mapping of 5-Methylisopsoralen Adducts in the SV40 Origin Region

ABSTRACT

Preferential psoralen binding sites have been mapped *in vitro* on restriction fragments spanning the SV40 origin region and surrounding sequences by a new fine structure analysis technique. Purified DNA fragments were photoreacted with ^{3}H-5-methylisopsoralen (^{3}H-5-MIP), a psoralen derivative which forms only monoadducts. Fragments were then end-labeled and digested with λ exonuclease, a 5' processive enzyme which we have determined pauses at 5-MIP monoadducts. When the psoralen binding sites were mapped on denaturing sequencing gels, it was observed that 5-MIP binds preferentially to 5'-TA sites, and to a lesser degree to 5'-AT sites. Utilizing this approach, we have identified a psoralen hypersensitive region in which the binding sites were much stronger than those in the surrounding sequences. This region extends from 150 base pairs (bp) to the late side of the enhancers to their early boundary. We suggest that a sequence-directed structural alteration of the DNA helix exists within and to the late side of the SV40 regulatory region which can be detected by this unique psoralen mapping analysis.

INTRODUCTION

Psoralens are three ring heterocyclic compounds that have become powerful tools for the in vivo analysis of DNA nucleoprotein structure and RNA secondary structure. These compounds have the ability to penetrate living cells and viruses without disruption of membranes and then intercalate between the stacked bases of the helix (1-4). Upon irradiation with long wavelength UV light, the intercalated molecule undergoes cyclophotoaddition to the 5,6 double bond of a pyrimidine to form a monoadduct. With additional UV irradiation, some psoralen derivatives may undergo a second photoreaction and form a crosslink between pyrimidines located in adjacent and opposite positions on the helix (reference 5 for review). In the case of 5-methylisopsoralen (5-MIP), the angular nature of the molecule prevents correct alignment for crosslink formation so that only monoadducts are formed.

Since nucleosomal core DNA is preferentially protected from psoralen photoaddition, psoralen derivatives are valuable probes for examining the structure of cellular and viral chromatin (2,3,6,7). We have previously used psoralen photoaddition to examine the structure of SV40 minichromosomes during a lytic infection and in assembled virus (8,9). In these experiments, a restriction fragment of about 360 bp encompassing the SV40 regulatory region was found to be preferentially accessible to psoralen in minichromosomes purified from infected cells, as well as in intracellular and extracellular virus. The spacing of nucleosomes along the SV40 genome has been shown to be nonrandom in this region on a

subpopulation of minichromosomes, in which a nucleosome free region of approximately 300 bp has been detected by both nuclease sensitivity (10-14), and by electron microscopy (15,16). This region spans the viral origin of replication, early and late promoters, the 72 bp repeats, which function as enhancers of viral transcription (17,18), and the T antigen binding sites. To date, the ability to characterize the chromatin structure of SV40 by psoralen photoaddition has been limited by an inability to localize psoralen monoadducts and crosslinks precisely at the nucleotide level from both in vivo and in vitro samples.

Piette and Hearst have mapped psoralen adducts by sizing the DNA products resulting from nick translation of circular ϕX174 templates previously photoreacted with 4,'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT) (19). Their results showed that E. coli Polymerase I was blocked by psoralen crosslinks with termination sites of equal intensity appearing before each 5'-AT and 5'-TA sequence. Unfortunately, it is impossible to map monoadducts with this approach since they do not block DNA Pol I activity. In addition, these experiments require synthesis of a complementary strand followed by hybridization to a single-stranded template, so this approach cannot readily be applied to the analysis of samples generated in vivo. Becker and Wang have developed a chemical method for mapping protein-DNA interactions utilizing UV light and have suggested its application for mapping psoralen photoadducts (20). Finally, Zhen et al. have very recently described a method for detecting psoralen crosslinks based on the inhibition of Bal 31 exonuclease activity. Like DNA Pol I, this enzyme is only inhibited by intrastrand

crosslinks and apparently cannot detect psoralen monoadducts (21).

In this report we have taken an enzymatic approach to mapping psoralen monoadducts generated in vitro on SV40 DNA. This approach, which readily lends itself to in vivo applications, takes advantage of several features of λ exonuclease, an enzyme required for general recombination in the bacteriophage λ (reference 22 for review). This enzyme cleaves 5'-mononucleotides from the 5' phosphoryl termini of double-stranded DNA, and preferentially initiates at 5'-recessed ends but not at nicks or gaps. We have determined that this enzyme pauses at 5-MIP monoadducts on a series of restriction fragments encompassing the SV40 origin region and regulatory sequences. When the locations of these pauses were mapped, it was determined that 5-MIP monoadducts were present at each 5'-TA in the sequence. Within these sequences, a small subregion was found to be hypersensitive to psoralen photoaddition in vitro. This segment is of particular biological interest and has been shown to include the repeated 72 bp enhancers, the late promoters, and the major late DNase I hypersensitive sites.

MATERIALS AND METHODS

Purification of DNA. Supercoiled SV40 DNA was purified from virus as described by Kondoleon et al. or from infected cells by a modification of the Hirt procedure (23). The maintenance of cell cultures, preparation of SV40 stocks and infection of cells have been described elsewhere (8).

Electrophoresis and Restriction Fragment Isolation. Restriction enzymes were purchased from BRL (Gaithersburg, MD) and were used according to the manufacturer's specifications. Restriction fragments of interest were separated by agarose gel electrophoresis using Bio-Rad ultra pure agarose (Richmond, CA) and isolated by electroelution into dialysis bags (24). Each sample was extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1), once with ether, and then ethanol precipitated. The DNA was resuspended in 1 ml TE (10 mM Tris, pH 8.0 and 1 mM EDTA) and further purified by passage over Elutip-d^{TM} minicolumns (Schleicher and Schuell, Keene, NH).

^{3}H-5-MIP Photoaddition. ^{3}H-5-MIP was obtained from HRI, Inc. (Emeryville, CA) in ethanol stock solutions of approximately 90 µg/ml. The specific activity was determined as previously described, and was approximately 9.54 x 10^{6} cpm/ug (12). DNA samples were suspended in 15-50 µl of H_2O in 1.5ml microcentrifuge tubes for irradiation. The irradiations were for 5-20 minutes and were performed by suspending the tubes between two GE F15T8 B1B fluorescent light bulbs (3.0-4.4 mW/cm^{2}/tube) spaced 7 cm apart. Controls were irradiated under identical conditions in the absence of psoralen.

End-labeling and Lambda Exonuclease Reaction. DNA was end-labeled with the Klenow Fragment (Boehringer Mannheim Biochemicals, Indianapolis, IN) as described by Maniatis et al. (24). Restriction enzymes were chosen such that only one end of a given restriction fragment would be labeled.

Following the Klenow reaction, 1 µg of sheared salmon sperm DNA and

0.5 volumes of 7.5M ammonium acetate were added, and the reaction mixture was heated 15 minutes at 70-80°C. The sample was then extracted twice with phenol/chloroform/isoamyl alcohol, and ethanol precipitated. This procedure was repeated a second time to insure that all protein and unincorporated nucleotides were removed.

Prior to digestion with λ exonuclease, samples were normalized so that the same number of counts went into each reaction. Reactions were carried out at 37°C in 67 mM glycine-KOH (pH 9.4), 3 mM MgCl, and 3 mM β-mercaptoethanol for 3 to 40 minutes. Twenty to 65 units of λ exonuclease (New England Biolabs, Beverly, MA) were used for each digestion. Following the reaction, samples were phenol extracted and ethanol precipitated as described previously, and resuspended in electrophoresis dye containing 80% deionized formamide, 50mM tris-borate (pH 8.3), 1 mM EDTA, and 0.1% (w/v) xylene cyanol and bromphenol blue.

Sequencing Gels and Autoradiography. DNA sequencing reactions and preparation of denaturing gels were carried out as described by Maxam and Gilbert (25). Following electrophoresis, gels were fixed in a 5% methanol/5% acetic acid solution, vacuum dried onto 3 mm Whatman chromatography paper for 1 hour at 80°C, and exposed to Kodak X-omat AR film for 18-36 hours.

RESULTS

The locations of 5-MIP monoadducts on restriction fragments spanning the SV40 regulatory region and surrounding sequences were mapped

following in vitro photoaddition. Since the angular nature of this particular psoralen derivative prohibits formation of crosslinks, the exact position of the monoadducts are determined by analysis on single-stranded sequencing gels following λ exonuclease digestion. The experimental approach is summarized in Figure 1 for a 362 bp Hind III-Kpn I (C_2) restriction fragment which extends on the labeled strand from nucleotides 294 to 5176 (see reference 26 for sequence). The fragment was purified as described and then photoreacted with ^{3}H-5-MIP. All controls were irradiated under parallel conditions in the absence of ^{3}H-5-MIP. Since initial reports characterizing λ exonuclease had concluded that 5'-recessed ends were important for initiation of digestion (reference 22 for review), the asymmetry of the fragment was maintained by end-labeling with [^{32}P]dATP alone. The enzyme should then preferentially initiate digestion on the ^{32}P labeled strand rather than the cold strand, although the degree to which this occurs is unknown.

The experiment shown in Figure 2, was conducted in order to determine whether λ exonuclease paused at 5-MIP monoadducts, and to ascertain the relative effect of different levels of photoaddition. Lane 3 contains the undigested experimental sample photoreacted with ^{3}H-5-MIP to a level of about one adduct per 100 bp. Digestion of that sample with an excess of λ exonuclease shows a unique pattern with specific bands of varying intensities (lane 4). Digestion of an irradiated control under the same conditions shows that the fragment is completely degraded in the absence of psoralen (lane 2). When the photoadduct level was decreased to approximately one adduct per 300 bp (lanes 7 and 8), the same discrete

pattern of bands is observed (lane 4).

The positions of the psoralen binding sites were mapped by comparison to Maxam-Gilbert sequencing of the C_2 fragment. Additional experiments under different gel conditions allowed mapping of all psoralen binding sites within the fragment. These data are summarized in Figure 3. As the arrows indicate, a psoralen-associated band is found 5' to every 5'-TA dinucleotide. Very faint bands were mapped 5' to most 5'-AT sequences. Because the psoralen-associated bands spanned a variable distance of 1-3 bp that could be attributed to the retardation of electrophoretic mobility by psoralen intercalation, the results are only accurate to within ±1 bp. They clearly indicate however, that 5'-TA conformations are the primary sites of 5-MIP photoaddition.

The kinetics of the reaction were examined to determine whether 5-MIP adducts totally block or, alternatively, inhibit the forward movement of λ exonuclease (Figure 4). The ten experimental samples were standardized for radioactivity. The purified, end-labeled DNA fragments with and without ^{3}H-5-MIP are shown in lanes 1 and 7, respectively. The samples in lanes 8 through 11 are irradiated controls and were digested for 2, 20, 40, and 80 minutes respectively. The control in lane 12 was digested for 80 minutes with 120 U of enzyme. The psoralen-irradiated counterparts in lanes 2 through 6 were photoreacted to a level of about one adduct per 100 bp. The gradual disappearance of bands in the ^{3}H-5-MIP reacted samples with increasing time of digestion indicates that psoralen photoadducts inhibit the processive nature of the enzyme but do not completely block the reaction.

Interestingly, in both experiments described above the strongest bands were mapped to the two 72 bp direct repeats which function as bidirectional enhancers of transcription. The fainter bands correspond to the direct 21 bp repeats and T antigen binding sites. Since the enzyme digests from both 5' ends, but may slow down considerably when the fragment has been reduced to two half-length single strands, it is possible that the enzyme did not always reach the 3' end of the 362 bp molecule, thus accounting for the weak psoralen bands in the promoter region. In order to eliminate this possibility, the C_2 fragment was digested with Sph I, which cleaves at nucleotides 200 and 128. When the remaining end-labeled 196 bp fragment was digested with λ exonuclease, the pauses associated with the two remaining 5'-TA sites in the enhancer region were as strong as those seen in the samples which were not cleaved with Sph I, and the pauses associated with 5'-TA sites in the rest of the molecule were as faint as those seen in the full-length fragment (compare samples I and III, Figure 5). This suggests that the transition from strong to weak bands is probably sequence specific and unrelated to the distance of these sequences from the λ exonuclease start site at the 5' end of the fragment.

To determine whether sequences at the 3' end of the molecule were unusually inaccessible to psoralen, or those associated with the enhancer were particularly accessible, Ava II-Sph I fragments which extended both 5' (nucleotides 201-560; termed E), and 3' (nucleotides 128-5122; termed F) from the origin region were analyzed. The additional psoralen stops found at the 3' end of the regulatory regions were as faint as those seen

previously (data not shown). However, a second transition was noted 5' to the enhancers at approximately nucleotide 398 (see Figure 6, lanes 5 and 6). When this fragment was digested with Hpa II, which cleaves at nucleotide 348, and the remaining end-labeled fragment was digested, the conclusions were unchanged, indicating that these results were also independent of the relative position of sequences within the restriction fragment (Figure 6, lane 2 and 3).

DISCUSSION

In these experiments 5-methylisopsoralen adducts on restriction fragments spanning the SV40 regulatory region and surrounding sequences have been mapped by a novel fine structure analysis. At multiple levels of photoaddition, 5-MIP monoadducts inhibit the forward movement of λ exonuclease and a series of specific bands of varying intensity are observed. The absence of bands in the irradiated control DNA indicates that the observed pattern is not related to an inability of the enzyme to digest this region.

When the psoralen binding sites were mapped by comparing band positions to Maxam-Gilbert DNA sequences, 5-MIP monoadducts were found 5' to each TA dinucleotide in the fragment. This work is in agreement with previous reports showing that psoralens establish covalent bonds preferentially at 5'-TA sites. Initially, crystallographic and solution studies (summarized by Sobell et al., 27) indicated that simple intercalating agents like psoralen show a preference for binding 5'-CG

and 5'-TA sequences in short chain oligomers. Gamper et al. subsequently showed that a double-stranded Kpn I linker is readily crosslinked by HMT while under identical conditions a corresponding Bam HI linker is resistant to modification (28). The linkers are identical except that a centrally located 5'-TA in the Kpn I linker is replaced by a 5'-AT in the Bam HI linker. They suggest that this preference reflects the ease of unstacking pyrimidine-purine sequences relative to purine-pyrimidine dimers. While this manuscript was in preparation, Zhen et al. reported that 5'-TA sites were preferred crosslinking sites for 4,5'8-trimethylpsoralen and 8-methoxypsoralen in restriction fragments isolated from pUC19 and pBR322 plasmids (21).

The psoralen bands within the enhancer region and immediately upstream were much more accessible than sequences elsewhere in the genome. A clear transition was noted immediately 3' to the enhancer region: by densitometer analysis the psoralen pause at nucleotide 106 is at least 100 fold stronger than that at nucleotide 94 and those downstream (data not shown). Cleavage of the C_2 fragment with Sph I to remove most of the enhancer sequences and subsequent digestion with λ exonuclease gave the same result, verifying that this transition point was independent of the distance of the sequences from the 5' start site of exonuclease digestion. Approximately 150 bp upstream from the enhancers at bp 398, a second region of transition was observed. Cleavage of the Ava II-Sph I E fragment with Hpa II, which cleaves at bp 348, and subsequent treatment with λ exonuclease gave the same result indicating, again, that this phenomenon is independent of the positioning

of the transition relative to the 5' end of the end-labeled strand.

The psoralen hypersensitive region spans sequences of particular biological significance. Interestingly, the 3' transition point (bp 106) occurs precisely at the 3' boundary of the enhancers. Jongstra et al. have shown that sequences contained mostly within the 72 bp repeat region are sufficient to generate a visible nucleosome gap (30). The psoralen hypersensitive sites overlap the major late promoter which initiates late transcription at bp 325. Sequences within this region are also critical for the formation of DNase I hypersensitive sites. Initially, Cremisi reported the formation of three hypersensitive sites on the late side of the Bgl I site (29). The first of these occurred around nucleotide 190 when the early genes are expressed. The other two are centered around nucleotides 270 and 370 and occur after or during viral replication when the late genes are expressed. More recently, Jongstra et al. have delineated two major regions of DNase I hypersensitivity in this region (30). One of these extends from approximately nucleotide 110, which coincides with the early border of the enhancers and the psoralen hypersensitive site, to approximately nucleotide 323 ± 20. This is within the upstream transition point of the psoralen hypersensitive region.

While the structural and molecular mechanisms underlying the generation of both the psoralen hypersensitive region and the nucleosome free region in this same portion of the genome are unknown, a number of explanations might account for both observations. First it is possible that sequence directed alterations of the helix in the enhancers may make

this region and surrounding sequences more susceptible to modification by psoralen. Nordheim and Rich have reported that the SV40 enhancer region contains three 8-bp segments of alternating purine-pyrimidine tracts which have the potential to form Z-DNA (31), and Nickol et al. have shown that nucleosomes cannot be formed over Z-DNA (32). However, in the experiments of Nordheim and Rich, Z-DNA was only observed in molecules with greater than 31.5 superhelical turns, whereas the experiments described in this paper were done with purified small restriction fragments. It should be noted that an identical pattern of band intensities was observed with linear and supercoiled SV40 DNA (Ostrander, unpublished data). In addition, this explanation could not account for the high accessibility of sequences 150 bp to the late side of the enhancer where no putative Z-DNA has been identified. Secondly, it is possible that sequences within the three 21 bp repeats and the T antigen binding sites might exclude psoralen from binding. Ryder et al. have established that a sequence-associated bend is located in T antigen binding site I, and a similar structure exits in site II (33,34). No evidence published to date, however, suggests that bent DNA is less sensitive to intercalating agents. Also, this explanation would not account for the pattern of reduced psoralen accessibility which occurs in the vicinity of nucleotide 398. Finally, it is theoretically possible that this region of the genome is uniformly accessible to 5-MIP, but that structural changes such as kinks are more readily induced by adduct formation in and near the enhancer region, and that these structural changes cause λ exonuclease to pause more readily.

We suggest that sequences in this region contain a structural alteration that differs from those previously described. Both Wasylyk et al. and Hiwasa et al. have shown by in vitro reconstitution experiments that nucleosomes form less readily on the SV40 origin region than on the rest of the SV40 genome, suggesting that some DNA sequences within the origin region cannot easily wrap around the histone octamer core (35,36). Structural factors which affect the twist or tilt of the helix might account for this phenomenon as well as the unusual psoralen hypersensitivity of this region.

Previous reports have shown that neighboring sequences can play a role in the site selection of intercalating agents, particularly psoralen. We have demonstrated that both HMT and 5-MIP display different binding preferences for the three unique Bgl I restriction sites in pBR322 (37). This preference cannot be tied to the distribution of 5-TA sequences within the restriction site. In addition, Zhen et al. found that only about 70% of a total of 20 5'-TA sequences within the DNA fragments they analyzed appeared to constitute efficient photo-crosslinking sites for HMT (21). Finally, in the Kpn I and Bam HI restriction fragment studies mentioned previously, Gamper et al. showed that the 100-fold preference of HMT for Kpn I linkers over Bam HI linkers drops only 10-fold when the same restriction enzymes are examined on full length SV40 DNA (28). This suggests that neighboring sequences can have a restricting effect on psoralen binding. While the described reports suggest that the DNA helix is not entirely uniform with respect to its accessibility to intercalating agents, this paper is the first report of

a long continuous sequence with an accessibility pattern that differs significantly from that of the surrounding region.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Nancy Kurkinen for excellent technical assistance and graphic arts; Gary Wiesehahn, Alex Elbrecht, Gordon Robinson, Sophia Kondoleon and Stephen Isaacs for helpful discussions; and Cheryl Maslen and John Hearst for critical reading of this manuscript. The authors would particularly like to thank James Shen and his group for their assistance in establishing Maxam-Gilbert sequencing technology in our laboratory. This work was supported by Public Health Service Grant 24799 to L.M.H. and a Tartar Trust Fellowship to E.A.O.

REFERENCES

1. Hanson, C.V., Shen, C.J., and Hearst, J.E. (1976) Science 193, 62-64.
2. Wiesehahn, G.P., and Hearst, J.E. (1976) In Nierlich, D.P., Rutter, W.J., and Fox, C.F. (eds), Molecular Mechanisms in the Control of Gene Exression ICN, UCLA Symposium, Academic Press, New York, Vol. 5, pp. 27-32.
3. Cech, T., and Pardue, M.L. (1977) Cell 11, 631-640.
4. Hallick, L.M., Yokota, H.A., Bartholomew, J.C., and Hearst, J.E. (1978) J. Virol. 27, 127-135.
5. Cimino, G.C., Gamper, H., Isaacs, S.T., and Hearst, J.E., (1985) Ann. Rev. Biochem. 54, 1151-1193.
6. Wiesehahn G.P., Hyde, J.E., and Hearst, J.E. (1977) Biochemistry 16, 925-932.
7. Sinden, R.R., Carlson, J.O., and Pettijohn, D.E. (1980) Cell 21, 773-783.
8. Robinson, G.W., and Hallick, L.M. (1982) J. Virol. 41, 78-87.
9. Kondoleon, S.K., Kurkinen, N.A., and Hallick, L.M. (Submitted, 1986).
10. Scott, W.A., and Wigmore, D.J. (1978), Cell 15, 1511-1518.
11. Sundin, O., and Varshavsky, A. (1979), J. Mol. Biol. 132, 535-546.
12. Varshavsky, A.J., Sundin, O.H., and Bohn, M.J. (1978), Nucl. Acids Res. 5, 3469-3477.

13. Varshavsky, A.J., Sundin, O., and Bohn, M.J. (1979), Cell 16, 453-466.

14. Waldeck, W., Fohring, B., Chowdhury, K, Gruss, P., and Sauer, G. (1978) Proc. Natl. Acad. Sci. USA 75, 5964-5968.

15. Jakobovits, E.B., Bratosin, E., and Aloni, Y. (1980) Nature (London) 285, 263-265.

16. Saragosti, S., Moyne, G., and Yaniv, M. (1980) Cell 20, 65-73.

17. Benoist, C., and Chambon, P. (1981) Nature 290, 304-310.

18. Gruss, P., Dhar, R., and Khoury, G. (1981) Proc. Natl. Acad. Sci. USA 78, 943-947.

19. Piette, J.G. and Hearst, J.E. (1983) Proc. Natl. Acad. Sci. USA 80, 5540-5544.

20. Becker, M.M. and Wang, J.C. (1984) Nature 309, 682-687.

21. Zhen, W., Buchardt, O., Nielsen, H., and Nielsen, P.E. (1986) Biochem. 25(21), 6598-6603.

22. Little, J.W., (1981), in Gene Amplification and Analysis. (J.G. Chirikjian and T.S. Papas, Eds.) Vol. 2, pp. 136-147. Elsevier/North Holland, New York.

23. Kondoleon, S.K., Robinson, G.W., and Hallick, L.M. (1983) Virology 129, 261-273.

24. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York.

25. Maxam, A.M., and Gilbert, W. (1980) In Grossman, L. and Moldave, K. (eds), Methods in Enzymology, Academic Press, New York, Vol. 65, pp.

499-560.

26. Tooze, J. (ed) (1981) Molecular Biology of DNA Tumor Viruses, 2nd edn.pp. 799-842. Cold Spring Harbor Laboratory, New York.

27. Sobell, H.M., Sakore, T.D., Jain, S.C., Banerjee, A., Bhandary, K.K., Reddy, B.S., and Lozansky, E.D. (1982) Cold Spring Harbor Symposium on Quatitative Biology 47, 293-314.

28. Gamper, H.J., Piette, J.G., and Hearst, J.E. (1984) Photochem. Photobiol. 40, 29-34.

29. Cremisi, C. (1981) Nucl. Acids. Res. 9, 5949-5964.

30. Jongstra, J., Reudelhuber, T.L., Oudet, P., Benoist, C., Chae, C.-B., Jeltsch, J.-M., Mathis, D.J., and Chambon, P. (1984) Nature 307, 708-714.

31. Nordheim, A. and Rich, A. (1983) Nature 303, 674-679.

32. Nickol, J., Behe, M., and Felsenfeld, G. (1982) Proc. Natl. Acad. Sci. USA 79, 1771-1775.

33. Ryder, K., Silver, S., Delucia, A., Fanning, E., and Tegtmeyer, P. (1986) 44, 719-725.

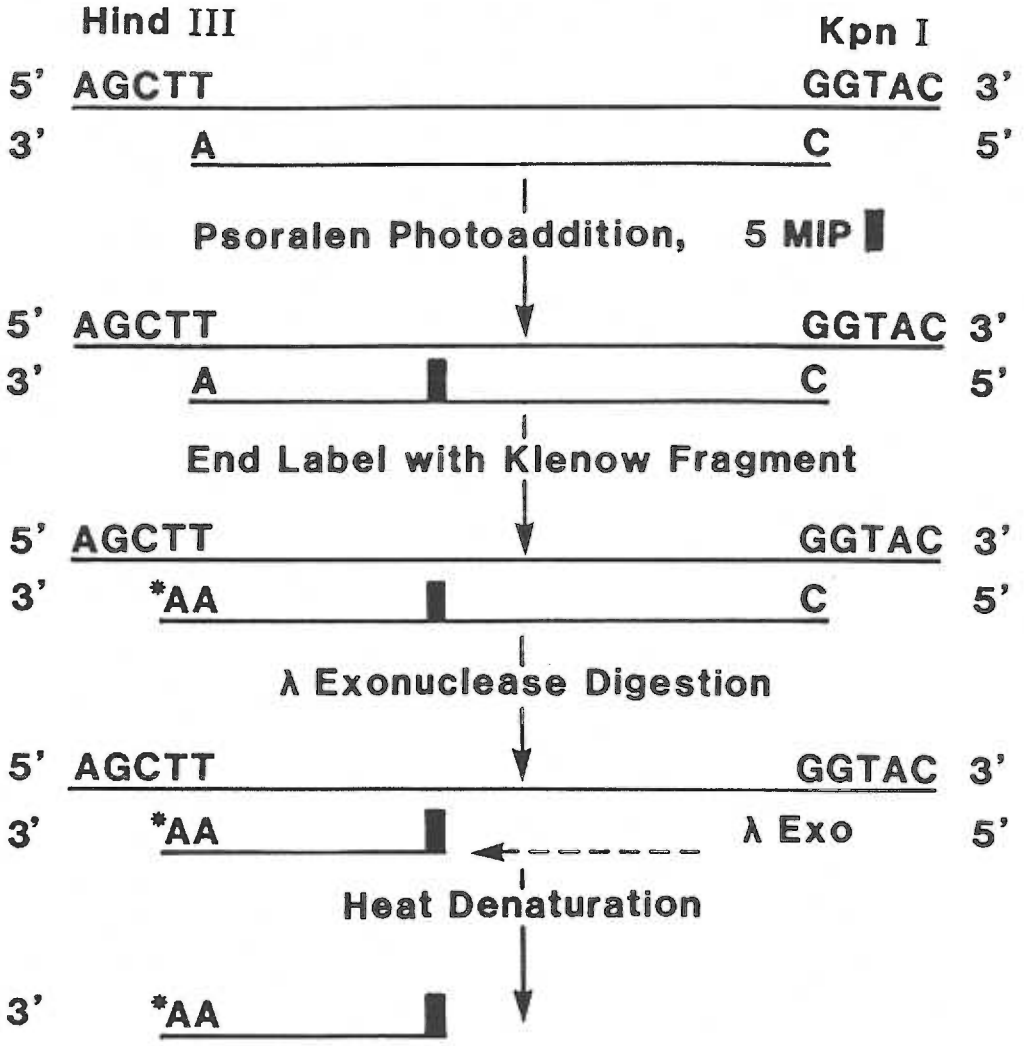
34. Deb, S., DeLucia, A., Koff, A., Tsui, S., and Tegtmeyer, P. (1986) Mol. and Cell Biol. 6(12), 4578-4584.

35. Wasylyk, B., Oudet, P., and Chambon, P. (1980) Nucl. Acids. Res. 7, 705-713.

36. Hiwasa, T., Segawa, M., Yamaguchi, N., and Oda, K.I. (1981) J. Biochem. 89, 1375-1389.

37. Ostrander, E.A., Robinson, G.W., Isaacs, S.T., Tessman, J., and Hallick, L.M. (1986) Photochem. Photobiol. 44(1), 21-29.

Figure 1. Experimental approach for mapping psoralen photoadducts using the enzyme λ exonuclease. Purified DNA is photoreacted with the isopsoralen derivative 5-MIP. The DNA fragment is end-labeled at the 3'-recessed end with the Klenow fragment of DNA polymerase I and ^{32}P-dATP. The labeled fragment is then digested with λ exonuclease, which initiates primarily at 5'-recessed ends, and fragments are analyzed on denaturing sequencing gels. Bands corresponding to λ exonuclease pauses, and hence the positions of 5-MIP adducts, are visualized by autoradiography.



SEQUENCING GEL

Figure 2. Fine structure mapping of 5-MIP monoadducts on a 362 bp Hind III-Kpn I restriction fragment (termed C_2) spanning the SV40 origin. Lanes 1,3 and 8 contain equal aliquots of the restriction fragment irradiated to levels of 0, 1/100 and 1/300 adduct per bp, but not treated with λ exonuclease. Lanes 2, 4, and 7 are the respective counterparts to 1, 3, and 8 but have been treated with an excess of λ exonuclease. Lanes 5 and 6 show C and C+T Maxam-Gilbert sequencing reactions of the C_2 fragment. The number 362 indicates fragment size; the remaining vertical numbers indicate nucleotide map positions of the 5-MIP-associated bands. A nonrandom background nicking pattern is noted in a subpopulation of the DNA fragments in both the experimental and control samples (see lanes 1,3,8). The extent and pattern of nicking was found to vary considerably from experiment to experiment; for example, see Figures 4 and 6 which show very low and moderately high levels of nicking in the undigested samples, respectively. The experiments have been repeated with restriction fragments purified both from agarose and polyacrylamide gels with various levels of nicking in the starting material, and the results were identical in all cases.

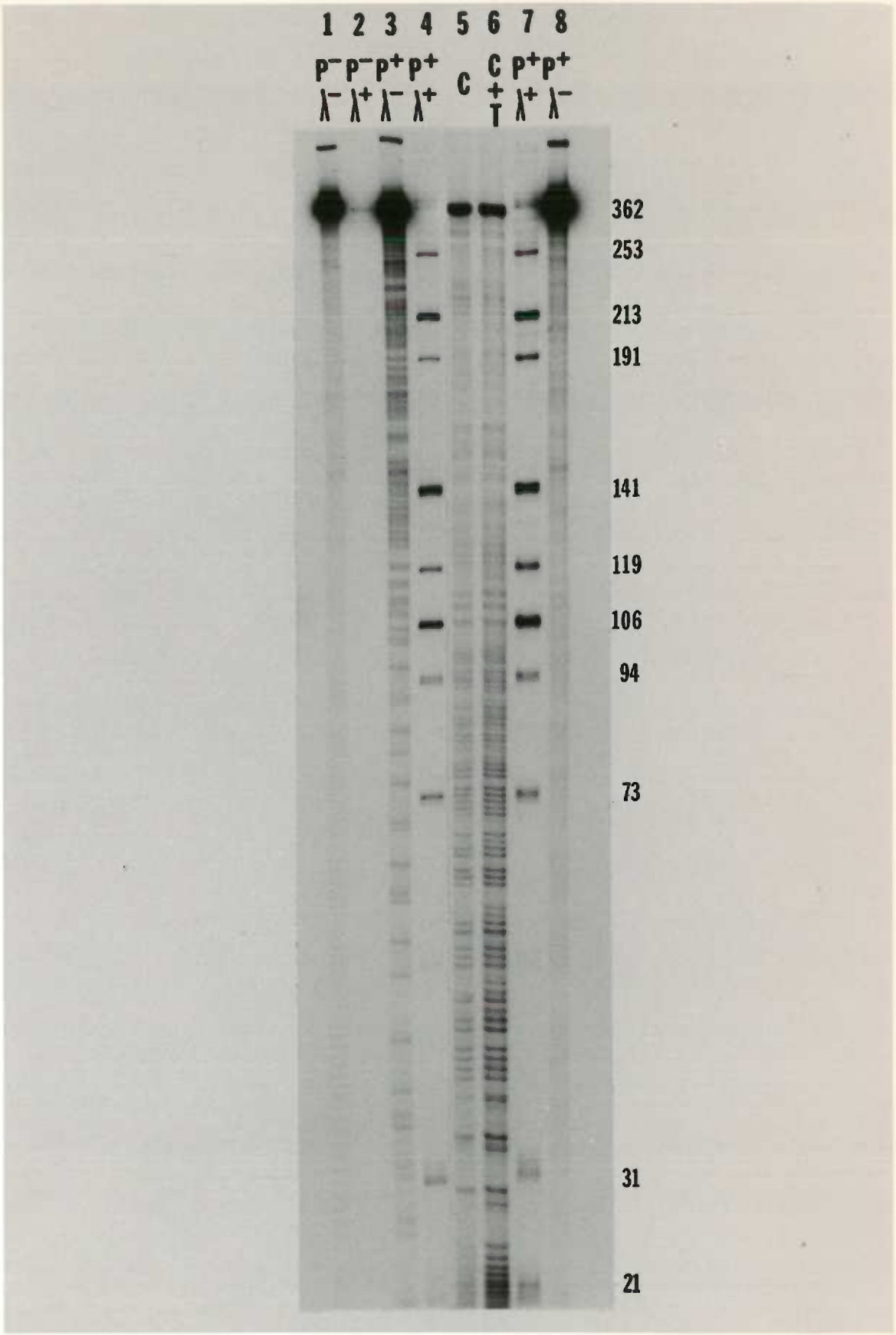

1 2 3 4 5 6 7 8
P⁻ λ⁻
P⁻ λ⁺
P⁺ λ⁻
P⁺ λ⁺
C
C + T
P⁺ λ⁺
P⁺ λ⁻
362
253
213
191
141
119
106
94
73
31
21

Figure 3. Localization of psoralen hypersensitive sites. The nucleotide sequence of the region mapped is shown. Early transcription is to the left, late to the right. Arrows in the line drawing above the sequence indicate pertinent restriction sites for fragments which were examined and the direction in which λ exonuclease traveled for each fragment. The 72 and 21 bp repeat regions, as well as T antigen binding sites I, II and III are indicated. Arrows in the sequence indicate the position of each psoralen-associated pause. Pauses were found 5' to every 5'-TA in the regions examined, and the intensity of each arrow is indicative of the relative strength of that band on the original gels. The striped rectangle delineates the psoralen hypersensitive region.

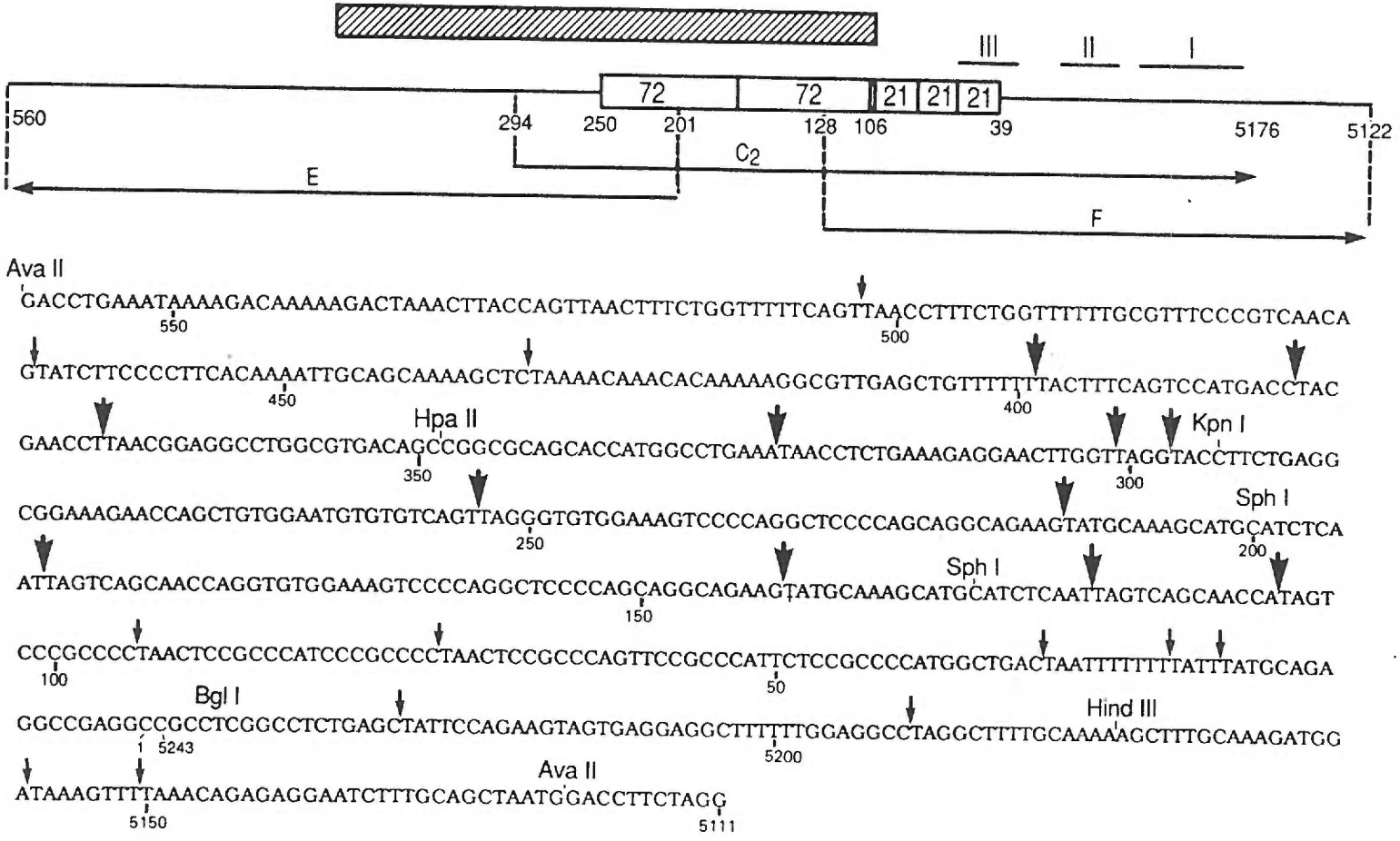

III
II
I
72
72
21
21
21
560
294
250
201
128
106
39
5176
5122
C2
E
F
Ava II
GACCTGAAATAAAAGACAAAAAGACTAAACTTACCAGTTAACTTTCTGGTTTTTCAGTTAACCTTTCTGGTTTTTTGCGTTTCCCGTCAACA
550
500
GTATCTTCCCCTTCACAAAATTGCAGCAAAAGCTCTAAAACAAACACAAAAAGGCGTTGAGCTGTTTTTTTACTTTCAGTCCATGACCTAC
450
400
Hpa II
Kpn I
GAACCTTAACGGAGGCCTGGCGTGACAGCCGGCGCAGCACCATGGCCTGAAATAACCTCTGAAAGAGGAACTTGGTTAGGTACCTTCTGAGG
350
300
Sph I
CGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCA
250
200
Sph I
ATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGT
150
CCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGA
100
50
Bgl I
Hind III
GGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTTTGCAAAGATGG
1
5243
5200
Ava II
ATAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGACCTTCTAGG
5150
5111

Figure 4. Kinetic analysis of λ exonuclease digestion on psoralen-labeled DNA. The SV40 C_2 fragment was end-labeled as described in the text, and equal amounts of radioactivity were used for the 10 experimental samples (lanes 2-6, 8-12). In lanes 1-5 the samples were irradiated with 5-MIP and digested with λ exonuclease for 0,2,20,40, and 80 minutes respectively. The sample in lane 6 was digested for 80 minutes with a double aliquot of enzyme. Lanes 7 through 12 correspond to lanes 1 through 6 but were irradiated in the absence of 5-MIP. The photoaddition level was approximately 1 adduct/200 bp for this and subsequent experiments. The number 362 indicates fragment size, while the remaining vertical numbers indicate nucleotide positions of 5-MIP-associated bands.

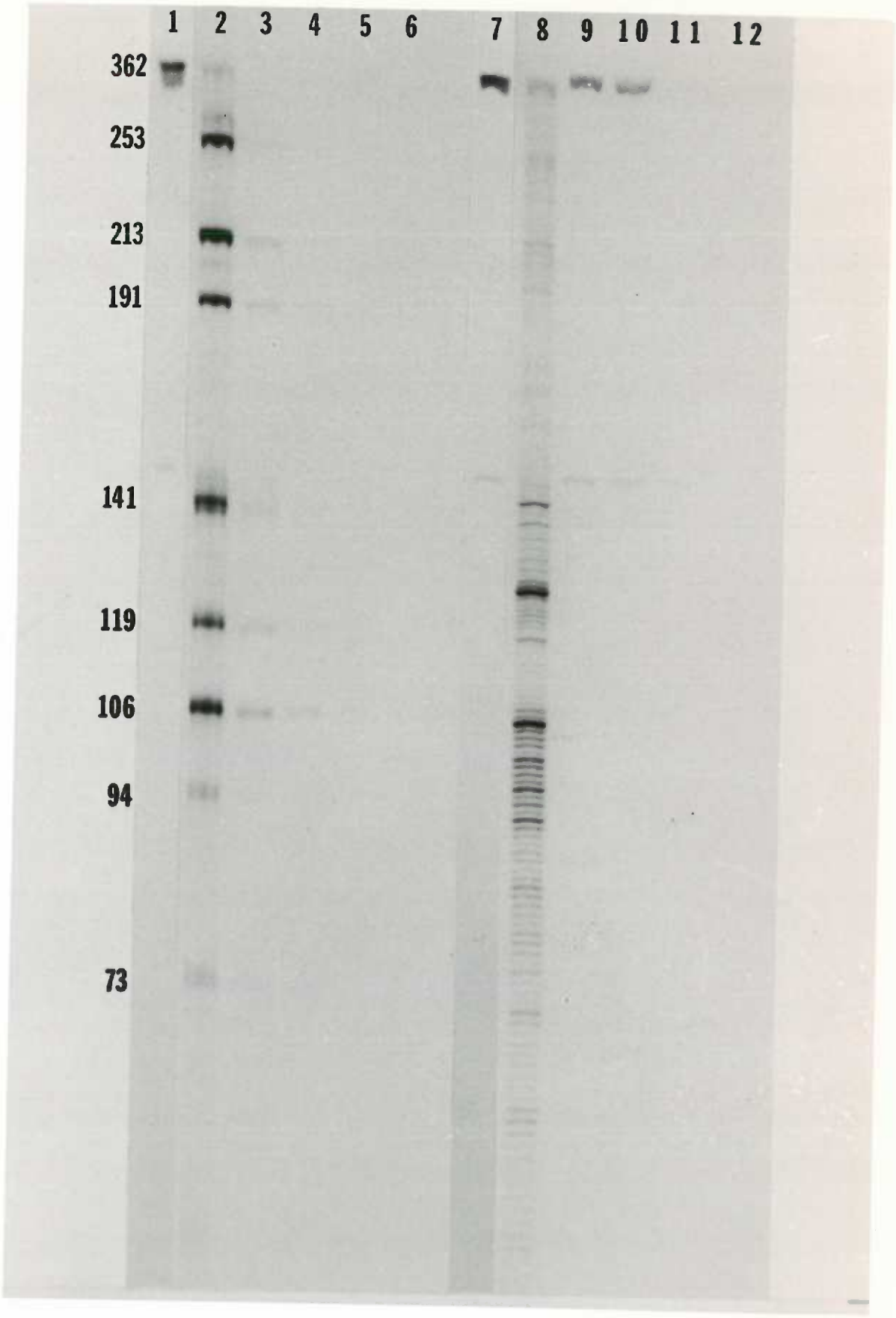

1
2
3
4
5
6
7
8
9
10
11
12
362
253
213
191
141
119
106
94
73

Figure 5. Effect of partial deletion of the enhancer region on psoralen photoaddition in the C_2 fragment. The purified fragment was digested with Sph I which cuts at bp 200 and 128 on the labeled strand to remove four of the six 5'-TA sites in the enhancer region (see Figure 3). This leaves a 196 bp fragment which was treated with λ exonuclease following irradiation in the presence and absence of 5-MIP (Sets III and IV). Sets I and II show the full length C_2 III-Kpn I fragment irradiated in the presence and absence of 5-MIP, respectively, before λ exonuclease treatment. Arabic numerals at the top of each lane refer to the length of time in minutes of exonuclease digestion. The number 362 indicates fragment size, while the remaining vertical numbers indicate nucleotide map positions of 5-MIP-associated bands.

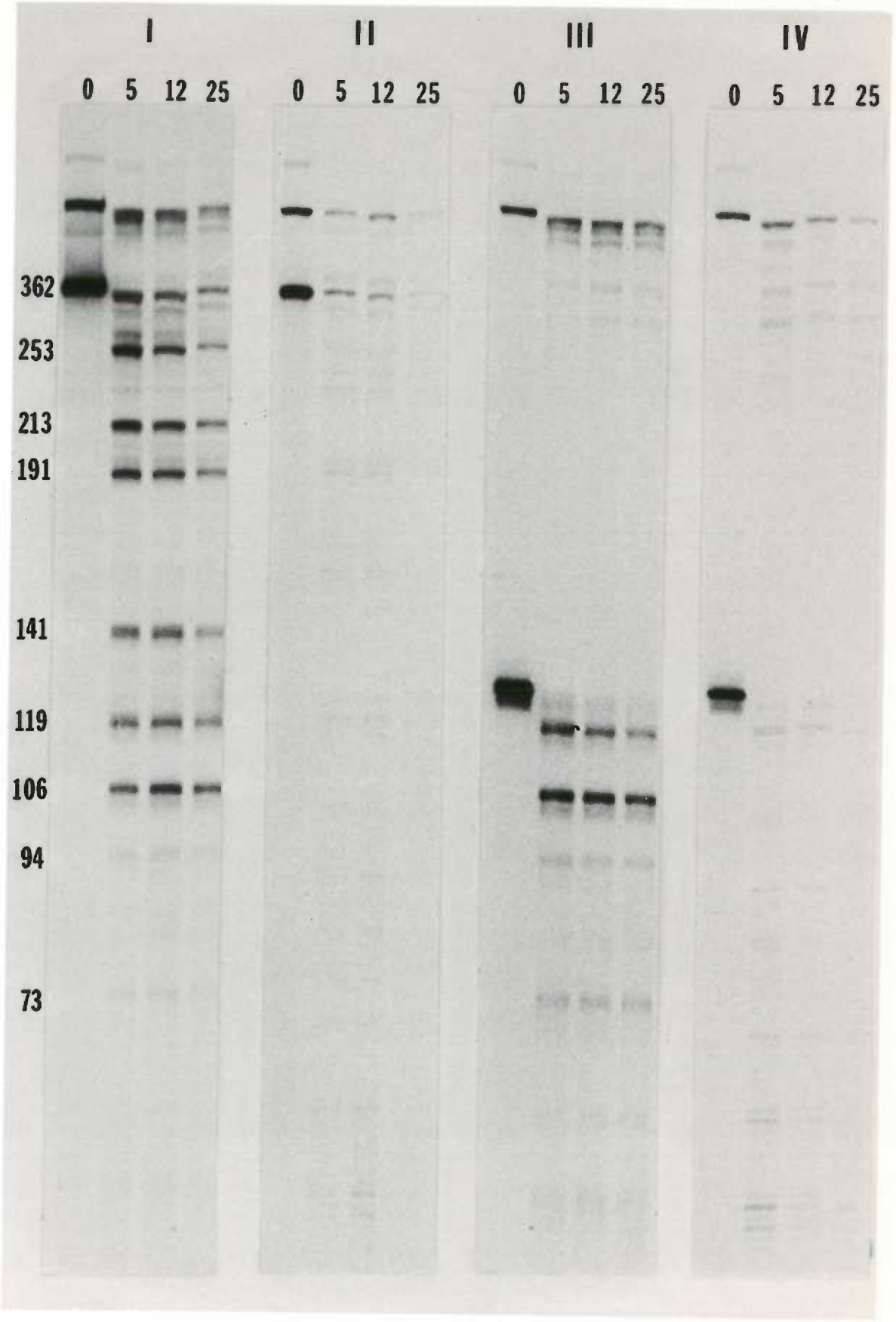

I
II
III
IV
0 5 12 25
0 5 12 25
0 5 12 25
0 5 12 25
362
253
213
191
141
119
106
94
73

Figure 6. Psoralen accessibility pattern on the late side of the origin region. The Ava II-Sph I sample in lanes 4-6 was psoralen photoreacted and then digested for 0, 5, and 20 minutes respectively with 65 U of λ exonuclease. In lanes 1-3 a portion of the sample represented in lane 4 was cleaved with Hpa II at bp 348 and then digested for 0, 5, and 20 minutes with 65 U of λ exonuclease. Lanes 7 and 8 show the Ava II-Sph I samples irradiated in the absence of 5-MIP and digested for 0 and 20 minutes with exonuclease. Lanes 9 and 10 show the corresponding Hpa II-Sph I samples treated for 0 and 20 minutes with λ exonuclease. Lanes 11-14 are the G, A+G, C and C+T Maxam-Gilbert sequencing lanes of the E fragment. A small region 5' to nucleotide 500 was too close to the end of the fragment to be sequenced.


1 2 3 4 5 6 7 8 9 10 11 12 13 14
212
252
296
300
324
370
379
397
432
466

Manuscript 3: Localization of 5-Methylisopsoralen Monoadducts in SV40 Viral DNA In Vivo

ABSTRACT

We have examined the psoralen accessibility of the SV40 origin region in minichromosomes irradiated at late times postinfection and in extracellular virus using a fine structure analysis technique recently described in manuscript 2. This technique takes advantage of the inhibition of λ exonuclease by 5-methylisopsoralen (5-MIP) monoadducts. Previous analysis of purified restriction fragments from this region irradiated in vitro delineated a region of psoralen hypersensitivity which encompassed the 72 bp repeat enhancers and sequences 150 bp upstream. Current studies demonstrate first, that changes in the structure of the DNA helix due to supercoiling do not affect the pattern of psoralen accessibility in this region. Second, no significant difference in the psoralen accessibility of this region was observed between late nucleoprotein complexes and extracellular virus. Finally, the 5-MIP preference for sequences within and to the late side of the enhancers observed in vitro is maintained for all samples analyzed.

The SV40 minichromosome isolated from infected cells is organized into a nucleosome-like "beads on a string" structure that is very similar to that of cellular chromatin (McGhee and Felsenfeld, 1980). Each double-stranded circular DNA molecule contains 21-23 nucleosomes and a highly variable linker region of 58±40 base pairs (bp) of DNA (Griffith, 1975; Shelton et al., 1980). Previously, nuclease digestion of isolated minichromosomes detected a 400 bp region that was particularly sensitive at both early and late times postinfection (Scott and Wigmore, 1978; Varshavsky et al., 1978; Waldeck et al., 1978). Electron microscopy of minichromosomes at late times postinfection revealed a nucleosome gap at the same position in about 25% of the molecules (Jakobovits et al., 1980; Saragosti et al., 1980). While the function of this "open region" during the virus life cycle remains unclear, sequences contained within are particularly important and include the origin of replication, the region encoding the 5' ends of early and late SV40 mRNA, and the preferred binding sites for T antigen.

This unusual chromatin structure was detected in this laboratory by examining the accessibility of intracellular nucleoprotein complexes to the psoralen derivative 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT) (Robinson and Hallick, 1982). Psoralens can penetrate intact cells and viral capsids without perturbing them. Irradiation with long-wavelength UV light then covalently binds psoralen to the pyrimidines of nucleic acids. Since nucleosomal core DNA is preferentially protected from psoralen photoaddition, these compounds are particularly valuable for examining the structure of both viral and cellular chromatin (for review

see Cimino et al., 1985). In the experiments of Robinson and Hallick, SV40-infected cells were irradiated with ^{3}H-HMT. The viral DNA was purified and digested with multiple restriction enzymes, and the relative level of radioactivity per restriction fragment assessed. Results demonstrated that a 400 bp region spanning the origin of replication was preferentially tritium labeled. This indicated that the nucleosome-free region detected by others in isolated viral chromatin also exists *in vivo*.

There is great interest in the SV40 nucleoprotein structure and how it may change at various stages in the viral life cycle. Recently, this laboratory detected an open origin in chromatin isolated from intracellular and extracellular virus particles (Kondoleon et al., submitted). This finding is interesting in light of structural and biochemical evidence that the SV40 genome in virus particles is significantly different from SV40 minichromosomes in infected cells. First, previous reports from this laboratory demonstrated that both intracellular and extracellular virion lack a psoralen accessible origin region when virion is processed and isolated by standard techniques (Kondoleon et al., 1983). Second, it was shown that when virions are dissociated by gentle treatment, the entire nucleoprotein complex shows greater sensitivity to nucleases as compared to minichromosomes, implying that the usual distinction between nucleosomal core DNA and linker DNA is absent. In addition, the preferential nuclease sensitivity found in minichromosome at the origin is absent (Brady et al., 1981; Hartman and Scott, 1981). Further examination of nucleoprotein complexes obtained by

mild dissociation at neutral pH indicates that the four core histones, although present, are not arranged in typical nucleosomes (Moyne et al., 1982). Most recently, Milavetz (1986) determined that the nucleosome-free region can be detected by nuclease sensitivity in nonencapsidated and partially encapsidated SV40 chromosomes and previrions, but not in mature virions.

These conclusions must be reevaluated in light of the recent study of Kondoleon et al. which showed that the routine freezing of virus particles prior to analysis causes a rearrangement of the chromatin structure and concomitant loss of ability to detect the nucleosome-free origin. In reviewing the studies of Milavetz et al. (1986) and Moyne et al. (1981), the virus was frozen prior to analysis, while it was stored at 4° C in the studies of Brady et al. (1981) and Hartman and Scott (1981).

Since the changes that occur in the maturation of the viral nucleoprotein complex are not well understood, we have examined the psoralen accessibility of the SV40 origin region in minichromosomes irradiated at late times postinfection and in extracellular virus using a fine structure analysis technique recently developed by Ostrander et al. (see previous manuscript). This approach takes advantage of the fact that λ exonuclease, an enzyme required for general recombination in the bacteriophage λ, (Shulman et al., 1975) is inhibited by 5-methylisopsoralen (5-MIP) monoadducts, thus allowing precise determination of nucleotide position for each DNA-psoralen adduct. If differences in the nucleoprotein structure of the minichromosome and the

mature virus were either caused by, or resulted in, changes in the overall secondary structure of the helix, they should be detected by this approach. Previous studies demonstrated the particular sensitivity of this method to factors which influence the structure of the helix (see Manuscript 2 and Discussion, this thesis). It should be noted that these data are preliminary and for that reason are included in report form.

MATERIALS ND METHODS

The growth and maintenance of MA134 cells, the preparation of SV40 776 stocks, the infection of cell cultures and the purification of virus were described previously, with the single exception that virus was stored at 4° C (Kondoleon et al., 1983).

^{3}H-5-MIP was purchased from HRI, Inc. (Emeryville, CA) in ethanol stock solutions of approximately 90 µg/ml and had a specific activity of 9.54×10^{6}. Stocks were diluted with unlabeled 5-MIP prior to use. Extracellular virus in tris buffered saline (pH 7.4) was irradiated for 20 min. with 10µg of 5-MIP in a total volume of 4 ml to a level of 1 adduct/400 bp. Irradiation conditions have been described elsewhere (Kondoleon et al., 1983). Supercoiled DNA was irradiated in 25 µg aliquots with 0.1 µg of ^{3}H-5-MIP for 16 min. These irradiations were done in 1.5 ml Eppendorf tubes suspended between two General Electric F15T8BLB fluorescent tubes (2.2-3.0 mW/cm^{2}/tube) spaced 7 cm apart, to a level of 1 adduct/150 bp. The conditions for the purification and irradiation of the Hind III-Kpn I C_2 restriction fragment were similar

and are described elsewhere (Ostrander et al., manuscript 2).

Psoralen irradiation of infected cells and purification of SV40 DNA occurred 60 hours postinfection, and was performed as described by Robinson et al. (1982). A single modification was that each 150 mm plate was irradiated with 10 μg aliquots of ^{3}H-5-MIP for a total of four 15 min. irradiations, since both the solubility and reactivity of ^{3}H-5-MIP is much lower than that of ^{3}H-HMT. Each aliquot of ^{3}H-5-MIP was solubilized in 2 ml of optically transparent albino media (described by Hallick et al., 1978) and removed from the culture dish prior to the addition of the next aliquot. The final adduct level for these samples was 1 adduct/150 bp.

End-labeling and λ exonuclease reactions were performed as described (Ostrander et al., Manuscript 2).

RESULTS

We have used a fine structure mapping analysis to compare psoralen accessibility patterns in the SV40 regulatory region from samples irradiated in vivo and in vitro. In this analysis restriction fragments were isolated by agarose gel electrophoresis, purified extensively, and end-labeled using the Klenow fragment and ^{32}P-dATP. Samples were then digested with λ exonuclease, a 5' processive enzyme which digests double-stranded DNA faithfully until it encounters a monoadduct, at which time the enzyme pauses. By denaturing such samples and running them on standard 8M urea sequencing gels, the exact positions of 5-MIP

monoadducts in restriction fragments of interest may be localized. We previously used this technique to analyze the SV40 origin region and surrounding sequences in vitro. This study demonstrated the presence of a structural perturbation of the DNA helix which results in an increased psoralen accessibility of sequences within the 72 bp repeat enhancer region extending about 150 base pairs upstream.

Analysis of in vivo generated samples for the experiments summarized below was initially difficult for technical reasons. Following purification and digestion of in vivo irradiated DNA, the Hind III-Kpn I C_2 fragment was electroeluted from ethidium-bromide stained agarose or polyacrylamide gels. Initial results were consistent with an absence of adducts in these experiments, even though the samples were sufficiently psoralen-labeled. In addition, these samples usually displayed a high nonspecific background. After several trials we determined that handling of the preparative gels on a short wave light box caused photoreversal of 5-MIP adducts. This phenomenon occurs quite easily with exposure to 240-330 UV light, particularly on photoreacted samples with a low adduct level (Cimino et al., 1986). Although typically it is desirable to avoid ethidium bromide-UV exposure of preparative gels, in this case it was done to allow efficient and precise isolation of the C_2 fragment from bands of similar size. Unfortunately, elimination of photoreversal does not alleviate the usually high background that is observed with in vivo samples.

Initially we compared the psoralen accessibility of DNA samples irradiated as supercoils and as linearized molecules (see figure 1,

samples IV and V). In all cases the pattern of strong bands associated with TA dinucleotides in the enhancer region (nucleotides 106-253), and weak bands associated with TA dinucleotides in the TAg binding region (nucleotides 94-39) was maintained. Interestingly, supercoiled DNA irradiated under the same conditions as Kpn I-linearized DNA had an adduct level of 1/154 bp while the linearized sample was 1/218. This indicates that psoralens bind to supercoiled DNA more readily than to linearized DNA, and is in agreement with the work of Hyde and Hearst (1978) who compared the intercalation of the hydrophobic psoralen derivative 4,5',8-trimethylpsoralen (TMP), and the hydrophilic derivative 4'-aminomethyl-4,5',8-trimethylpsoralen (AMT) to supercoiled and linearized ColE1 plasmid. They determined that the association constant for both derivatives to supercoiled DNA was much higher than to linear DNA. Intercalation by psoralen locally unwinds the supercoiled DNA, thus overwinding occurs somewhere else in the molecule. Because the supercoiled DNA is underwound to begin with, this action reduces torsional stress and hence there is a favorable free energy state associated with binding to supercoiled molecules relative to linear molecules. Nucleotide level data presented in this paper suggests that this change in free energy does not facilitate psoralen binding to less favorable sequences.

In the next set of experiments, the accessibility pattern of the C_2 fragment isolated from purified virus as well as from intracellular nucleoprotein complexes was compared to the pattern observed on naked DNA (See figure 1). There is no significant difference in the pattern of

psoralen accessibility in virion and intracellular nucleoprotein complexes. In particular, the 5-MIP preference for sequences within and to the late side of the enhancers which was also observed *in vitro* is maintained. We consider these data preliminary, however, because of the high background level in the *in vivo* samples. The poor signal to noise ratio of virion DNA in particular probably represents the low psoralen photoaddition level of this sample (1 adduct/400 bp) compared to the intracellular and the *in vitro*-labeled supercoiled samples (1 adduct/150 bp).

DISCUSSION

There are significant structural differences between purified DNA and DNA which is arranged in a nucleosome. Levitt (1978) has argued that linear DNA free in solution is most stable with 10.6 base pairs per turn, whereas supercoiled DNA in chromatin is most stable with 10 base pairs per turn. These parameters were also measured experimentally by Germond et al. (1975) and Keller (1975) who showed that the linking number (number of times one strand of the helix winds around the second) changes by about -1.25 when a single nucleosome is added. Since 10.6 and 10.0 base pairs per turn are equivalent to linking numbers of 17.5 and 18.9 for 200 base pairs respectively, these studies are in good agreement.

The preference of linear DNA for a helix of 10.6 bp per turn and of superhelical DNA for 10 bp/turn can be explained by differences in the tilting of the stacked bases. When there are fewer bases per turn, as in

superhelical DNA, the bases on a single stand overlap less because they are more extended. Keeping in mind that in B-form DNA the typical hydrogen-bonded base pair does not lie on a plane, but is assembled with a propeller twist (Callidine, 1982; and references therein), in chromatin individual bases may tilt to create more favorable stacking interactions. (See the Discussion of this thesis for a more complete explanation.) When there are 10 rather than 10.6 base pairs per turn, the bases on the inside of the curve of the DNA wrapped around the nucleosome tilt more to avoid close intra-strand base contact, and thus have more favorable stacking interactions (Levitt, 1978). Since the psoralen accessibility pattern for chromatin as well as supercoiled and linear SV40 DNA are the same, it may be argued that the psoralen hypersensitive region which characterizes the SV40 enhancers and sequences upstream has a helical perturbation other than a change in linking number or base pair tilt. Although psoralen binds preferentially to linker DNA, the fact that the 21 bp repeats and TAg binding sites are contained within the nucleosome free region but not in the psoralen hypersensitive subregion, supports the claim that both linker and nucleosomal DNA are altered in terms of linking number and base pair tilt. This allows us to speculate that helical variation in parameters such as roll or slide account for the hypersensitivity of the enhancers and upstream sequences. The precise nature of this structural variation and its biological role are yet to be determined.

BIBLIOGRAPHY

Brady, J., Radonovich, M., Lavialle, C., and Salzman, N.P. (1981). Simian virus 40 maturation: Chromatin modifications increase the accessibility of viral DNA to nuclease and RNA polymerase. J. Virol. 39, 603-611.

Callidine, C.R. (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. J. Mol. Biol. 161, 343-352.

Germond, J.E., Hirst, B., Oudet P., Gross-Bellard, M., and Chambon, P. Folding of the DNA double helix in chromatin-like structures from simian virus 40. Proc. Natl. Acad. Sci. USA 72, 1843-1847.

Cimino, G., Gamper, H.B., Isaacs, S.T., and Hearst, J.E. (1985). Psoralens as photoactive probes of nucleic acid structure and function: Organic chemistry, photochemistry and biochemistry. Ann. Rev. Biochem. 54, 1151-1193.

Griffith, J. (1975). Chromatin structure: Deduced from a minichromosome. Science 187, 1201-1203.

Hyde, J.E., and Hearst, J.E. (1978). Binding of psoralen derivatives to DNA and chromatin: Influence of the ionic environment on dark binding and photoreactivity. Biochem. 17, 1251-1257.

Hallick, L.M., Yokota, H.A., Bartholomew, J.C., and Hearst, J.E. (1978). Photochemical addition of the crosslinking reagent 4,5',8-trimethylpsoralen to intracellular and viral simian virus 40-histone complexes. J. Virol. 27, 127-135.

Hartmann, J.P., and Scott, W.A. (1981). Distribution of DNase

I-sensitive sites in simian virus 40 nucleoprotein complexes from disrupted virus particles. J. Virol. 37, 908-915.

Jakobovits, E. B., Bratosin, S., and Y. Aloni. (1980). A nucleosome-free region in SV40 minichromosomes. Nature 285, 263-265.

Keller, W. (1975). Determination of the number of superhelical turns in simian virus 40 DNA by gel electrophoresis. Proc. Natl. Acad. Sci. USA 72, 4876-4880.

Kondoleon, S.K., Robinson, G.W., and Hallick, L.M. (1983). SV40 virus particles lack a psoralen-accessible origin and contain an altered nucleoprotein structure. Virology 129, 261-273.

Kondoleon, S.K., Kurkinen, N.A., and Hallick, L.M. (Submitted). The SV40 nucleosome free region is detected throughout the virus life cycle.

Levitt, M. (1978). How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. Proc. Natl. Acad. Sci. USA 75, 640-644.

McGhee, J.D., and Felsenfeld, G. (1980). Nucleosome structure. Ann. Rev. Biochem. 49, 1115-1156.

Milavetz, B. Analysis of the origin-specific nucleosome-free region in SV40 encapsidation intermediates. Virology 153, 310-313.

Moyne, G., Harper. F., Saragosti, S., and Yaniv, M. (1982). Absence of nucleosomes in a histone-containing nucleoprotein complex obtained by dissociation of purified SV40 virions. Cell 30, 123-130.

Ostrander, E.A., Karty, R.K., and Hallick, L.M. (submitted). Fine

structure mapping of 5-methylisopsoralen adducts in the SV40 origin region.

Robinson, G.W., and Hallick, L.M. (1982). Mapping the in vivo arrangement of nucleosomes on simian virus 40 chromatin by the photoaddition of radioactive hydroxymethyltrimethylpsoralen. J. Virol. 41, 78-87.

Saragosti, S., Moyne, G., and Yaniv, M. (1980). Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. Cell 65-73.

Scott, W.A., and Wigmore, D.J. (1978). Sites in simian virus 40 chromatin which are preferentially cleaved by endonucleases. Cell 15, 1511-1518.

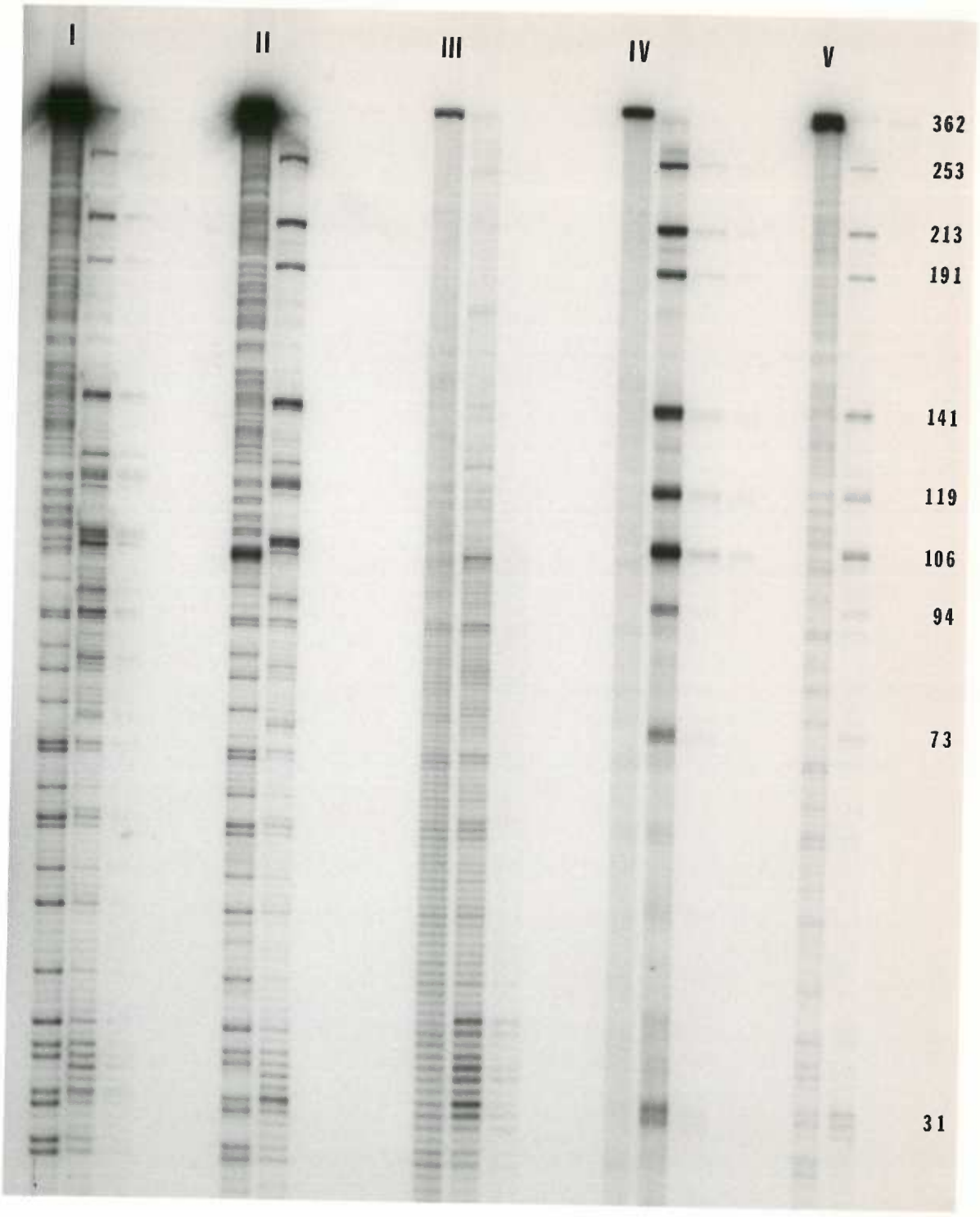
Shelton, E.R., Wassarman, P.M., and DePamphilis, M.L. (1980). Structure, spacing and phasing of nucleosomes on isolated forms of mature SV40 chromosomes. J. Biol. Chem. 255, 771-782.

Shulman, M.J., Hallick, L.M., Echols, H., and Signer, E.R. (1970). J. Mol. Biol. 52, 501-520.

Varshavsky, A.J., Sundin, O.H., and Bohn, M.J. (1978). SV40 viral minichromosomes: preferential exposure of the origin of replication as probed by restriction endonucleases. Nucl. Acids Res. 5, 3469-3477.

Waldeck, W., Fohring, B., Chowdhury, K., Gruss, P., and Sauer, G. (1978). Origin of DNA replication of papovavirus chromatin is recognized by endogenous endonuclease. Proc. Natl. Acad. Sci. USA 75, 5964-5968.

Figure 1. Fine structure mapping of psoralen accessible regions within the SV40 regulatory region. Samples were irradiated and prepared as described in the text. After end-labeling with ^{32}P-dATP, samples were normalized so that the same number of counts went into each reaction. Reactions were carried out at 37° C in 67 mM glycine-KOH (pH9.4), 3 mM MgCl, and 3 mM β-mercapthoethanol for 0, 5, and 20 minutes. 60 units of λ exonuclease (New England Biolabs, Beverly, MA) were used for each digestion. Samples were suspended in 80% deionized formamide, 50 mM tris-borate (pH9.3), 1 mM EDTA, and 0.1% (w/v) xylene cyanol and bromphenol blue for loading. They were then heated at 90°C for 10 min., and analyzed on standard 8% denaturing gels. Sample I is the Hind III-Kpn I C_2 fragment isolated from 5-MIP-irradiated extracellular virus. In sample II the C_2 fragment was isolated from virus infected cells irradiated 60 hours postinfection. In samples III and IV the C_2 fragment was purified first and then irradiated in the absence and presence of 5-MIP respectively. In sample V supercoiled SV40 DNA was irradiated and then the C_2 fragment purified and analyzed. The number 362 indicates fragment size, while the remaining vertical numbers indicate nucleotide positions of the 5-MIP-assoicated bands.


I
II
III
IV
V
362
253
213
191
141
119
106
94
73
31

DISCUSSION

"...In the space of one hundred and seventy-six years the Lower Mississippi has shortened itself two hundred and forty-two miles. That is an average of a trifle over one mile and a third per year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oolitic Silurian Period, just a million years ago next November, the Lower Mississippi River was upwards of one million three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing-rod. And by the same token any person can see that seven hundred and forty-two years from now the Lower Mississippi will be only a mile and three quarters long, and Cairo, Illinois and New Orleans will have joined their streets together and be plodding comfortably along under a single mayor with a mutual board of aldermen. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact." (Mark Twain, 1875)

In recent years, the structural diversity of helical DNA and its relationship to gene expression have been the subject of numerous physical, chemical, and theoretical studies. Out of these studies has come an elementary understanding of the polymorphic and dynamic nature of duplex DNA, as well as wide speculation about the role of structural diversities in processes such as transcription. In this thesis I have used psoralen photoaddition to study variation in DNA sequences

encompassing the SV40 regulatory region. Psoralen-DNA interactions in random-sequence DNA have also been studied by examining the inhibition of Bgl I cleavage of pBR322 DNA with different populations of adducts. In both cases we have determined that accessibility to psoralen is not random, but rather that there are regions of preferential binding as well as regions of relative inaccessibility. Our findings regarding SV40 are particularly intriguing in view of both the biological importance of the sub-region studied, and because of previous findings regarding its chromatin structure. In this discussion I speculate on the possible contribution of structural diversity to the findings and significance of the observations reported in this thesis. It should be noted that in this discussion the term "accessibility" is operationally rather than physically defined, and indicates DNA sequences of particular psoralen sensitivity described by the techniques noted.

A. The Basis of Structural Discontinuity in the DNA Helix.

At its simplest level DNA molecules fall into one of three primary structural families: A, Z, and the familiar B form. The most striking difference between the families, aside from the left handedness and lack of a major groove in Z DNA, is the difference in physical proportions of the respective molecules. For the same number of base-pairs, A-DNA is short and thick, B-DNA is long and slim, and Z-DNA is elongated and even thinner (Dickerson et al., 1982). Since the attachment of the base to the deoxyribose sugar occurs at the nitrogen position of the base and the C_1 position of the sugar (see figure 1), and because this glycosyl bond

is branched off one side of the sugar, the base pairs are regularly displaced from the helix axis. This phenomenon accounts for the existence of a major and minor groove once the double strands are twisted. The primary distinction between A, B, and Z forms of DNA is the major and minor groove dimensions. These differences arise from variations in the tilt of the base pairs with respect to the vertical helix axis, and from the overall positioning of the axis (in a horizontal sense) relative to the base pairs. In B-DNA the helix axis passes through the hydrogen bonds which connect the bases within the helix (see figure 1). For reasons the importance of which will become apparent shortly, this means that the major and minor grooves of B-DNA are of comparable depth. In A-DNA, the helix axis passes through the major groove side of each base pair, so the major groove is very deep and the minor groove very shallow. In Z-DNA the situation is reversed and the helix axis passes down the minor groove, so the minor groove is deep, and the major groove is flattened out on the helix surface to the point of nonexistence.

It is important to realize that the A, B, and Z forms are concepts which represent idealized molecules, and that the actual geometry of a given stretch of DNA can reflect all three conformations to varying extents (Wang et al., 1982). The relative proportions of A versus B-DNA in a given polymer (in solution) depend on both base sequence and hydration (Calladine and Drew, 1984, and references therein). Sequences such as poly(dA).poly(dT) remain in the B form regardless of hydration, while poly(dA-dT) and poly(dG-dC) switch from B to A upon dehydration,

and others such as poly(dG).poly(dC) favor A form even when well hydrated.

The idea that sequence is one of the primary determinants of overall helical geometry has been extended by Calladine (1982) who proposed that all observed deviations from an "ideal" B form DNA conformation (and he argues that such deviations occur frequently) are due to the steric hindrance between nearest-neighbor purines on opposite phosphate backbones. To understand this argument consider the geometry of two successive base pairs (See figure 2A and 2B). For a given dinucleotide, steric hindrance results from two combined facts. The first is the larger size of purines compared to their complementary pyrimidines; consequently nearest-neighbor purines on opposite DNA strands overlap each other inside the helix and their hydrogen bonds are eccentrically positioned closer to the pyrimidine backbone than to the purine backbone. The second is that base pairs do not lie flat on a plane within the helix, and instead are assembled with what is termed propeller twist. This means that when a base pair is viewed along its long axis, the nearer base will be rotated clockwise with respect to the distant one (Levitt, 1978; Hogan et al., 1978). A positive propeller twist of 10°-20° enables each base to overlap and stack more efficiently against its neighbors on the same strand. But the phenomenon of propeller twist combined with the difference in relative size for purines and pyrimidines also creates a problem of "steric clash" (see figure 2C). As the base pairs slide across one another and a base on one strand overlaps with a base from the opposite strand, the introduction of propeller twist brings

the atoms involved in the overlaps closer together (from 3.4 Å to 2.8 Å), resulting in clashes between base pairs with propeller rotations in opposite directions (Dickerson and Drew, 1981). How does the helix deal with such steric irritations?

B. Structural Solutions of Steric Clash.

To address this question, it is important to understand that the geometry of any isolated dinucleotide can be described by three translational and three rotational degrees of freedom (Callidine and Drew 1984; see figures 2A and 2B). The first of these is twist (d), which is simply the rotation of the bases about the vertical helix axis. Tilt is the up and down rotation of the axis passing through the base pair plane (f). Finally, roll (e) is the rotation of the bases about an axis which is perpendicular to a line bisecting the region of hydrogen bonding. It is distinguished from propeller twist in that it defines a base pair rather than a single base. (More simply, consider two people at opposite ends of a canoe [representing the bases], with the space between representing the region of hydrogen bonding. There are three ways the boat may be rotated; it can <u>roll</u> and go over on its side, it can <u>tilt</u> either forward or backward to such a degree the boat flips over end to end, or it can simply <u>twist</u> around, left or right, in the plane of the water). The three translational degrees of freedom are shown in figure 2A. The most important of these for purposes of this discussion is slide (b) which is the translation of the base-pairs relative to one another perpendicular to the helix axis, in a manner that is similar to pulling a

single quarter out of a vertical stack of coins. The important implication of this is that the torsion angles defining the attachment of the base to the sugar-backbone are changed when slide is adjusted.

With these definitions in mind, the original question of how DNA accommodates the steric hindrance imposed by base sequence can be addressed. Callidine (1982) has proposed a set of four maneuvers which DNA may undertake to deal with steric clash. It is from these maneuvers that much of the structural discontinuity in the helix is thought to derive. Clash may be relieved by:

1) decreasing the local twist angle between base pairs,
2) opening up the roll angle between base-pairs on the side on which the clash occurs (change in roll),
3) separating purines by sliding one or both of the base-pairs along their long axes so that the purines are partially pulled out of the base-pair stack (change in torsion angle, the angle between the sugar-base covalent bond and the sugar phosphate backbone), or
4) suppressing or flattening the propeller twist in one or both of the base pairs.

Dickerson (1983) has defined simple sum functions Σ_1 through Σ_4 for calculating the expected variation in helix twist, roll angle, torsion angle, and propeller twist for any given base sequence, with twist and roll being properties of base-pair steps from one base pair to the next, while torsion angle and propeller twist are properties of individual base pairs.

C. Support for Theoretical Proposals Using Restriction Enzyme Digestion Patterns.

The considerations of this discussion so far have been strictly theoretical. To date there has been little biological justification for postulates which explain the nature of structural discontinuity in the DNA helix. The earliest biological studies which can be used to support the claim that structural discontinuities exist frequently within the DNA helix were simple restriction enzyme studies. The first such study, by Thomas and Davis (1975) mapped the five Eco RI sites in bacteriophage λ. They demonstrated that sites 1 and 2 which were located in the middle of the linear λ molecule were cleaved 1/10th as frequently as site 5 located near the end. The reverse, however, was true with P4 DNA, where the site in the middle was cleaved more rapidly than sites at the end, suggesting that the distance from the end of a given restriction site played no role in its accessibility to a restriction enzyme. This implies that sequences adjacent to the restriction sites are important for enzyme recognition and play a key role in aligning or "presenting" such sites to enzymes.

A more complete study is that of Armstrong and Bauer (1982) who examined the order of appearance of partial digestion products from four Pst I sites in pSM1 DNA. They found that cleavage occurred at markedly different rates for the four sites and suggested that either variations in the DNA sequences adjacent to the recognition site, or the presence of unusual duplex structures such as hairpins or cruciforms, play a role in either the recognition or efficiency of restriction enzyme cutting. In

regards to this, Courey and Wang (1983) have argued convincingly that there is little compelling evidence for the existence of cruciforms in vivo, and even less for their involvement in gene expression. Since the nucleotide sequence adjacent to the most highly inhibited site is G-C rich, both 5' and 3' to the recognition site, it is more likely that this sequence conformation distorts the duplex in some way that is inhibitory to restriction enzymes. The following year the same group drew similar conclusions from studies with Hinf I (Armstrong and Bauer, 1983), as did Alves et al. (1983), who found preferential cleavage of Eco RI in dA rich strands versus dG rich strands.

The pBR322 studies summarized in Manuscript 1 of this thesis were a natural outgrowth of the studies mentioned above. The restriction enzyme Bgl I has an 11 bp recognition sequence with a 5 bp internal neutral region. Conveniently, pBR322 contains three such sites with different conformations of thymines, and informative sequences surround each recognition site. In these studies, preferential inhibition of the site which contained a 5'-AT sequence was observed. Interestingly though, cleavage at all three sites was inhibited equally by HMT-samples with the same loading level but very different ratios of monoadducts to crosslinks. Parallel samples irradiated with 5-MIP displayed significantly less inhibition. It can be argued that the greater inhibition by HMT monadduct samples is due to a structural difference between HMT monoadducts and 5-MIP monoadducts. In fact, 5-MIP unwinds the helix by 18° (Isaacs et al., 1984) while a psoralen derivative similar to HMT, namely TMP, unwinds the helix by 28° (Wiesehahn, 1978).

This unwinding would be expected to distort the phosphate backbone to some degree, and since restriction enzymes appear to recognize the DNA backbone in addition to the base sequence (Drew and Travers, 1985), it is conceivable that HMT monoadducts could be more inhibitory than 5-MIP monoadducts. A second line of thinking is that the few crosslinks present in the samples irradiated with high intensity UV are responsible for the majority of Bgl I inhibition. From this it follows that there is a "gradient of accessibility" for various DNA sequences to psoralen, with a few particular "hotspots" filling first. This is supported by findings from Manuscript 2 of this thesis which show a 100 fold difference in psoralen accessibility by sites only a few bases apart with no known structural perturbation (Z-DNA, bends, etc.). The data of Pearlman et al. (1984) also support this interpretation in that it provides a model for the extreme disfiguration of the DNA helix by the presence of even a single psoralen crosslink. This data suggests that a crosslink generates a kink angle in the helix of 46.5°.

Finally, it is formally possible that 5-MIP binds with a different sequence specificity than HMT, accounting for the difference in inhibition. This possibility seems remote, however, in view of the combined data of Gamper et al. (1984) and information presented in Manuscript 2 of this thesis, both of which show that psoralens bind preferentially to 5'-TA dinucleotides.

The most likely explanation is a combination of two of the three lines of thinking suggested. There are probably hot spots for both psoralen binding and crosslink formation. However, the high intensity

samples with 86 adducts/1000 bp, but containing only 2% crosslinks (1.7 crosslinks/1000 bp) were significantly more inhibited than samples irradiated in broad band light to the same crosslink level. Since the latter samples must contain only 5 total HMT adducts/1000bp, the difference in the level of HMT monoadducts probably plays a significant role in the higher Bgl I inhibition on the high intensity samples. It seems likely that there are hotspots for psoralen crosslinking, and that the presence of monoadducts in addition to well placed crosslinks allows for a synergistic effect; i.e. HMT monoadducts in an already distorted environment inhibit Bgl I digestion more than monoadducts alone.

D. DNAase Studies of Variation in the Helical Backbone.

Two critical points are unresolved in the Bgl I studies. The first is the type and frequency of the structural discontinuity which plays a role in the initial psoralen adduct site selection; and second, the distance effect of such discontinuities. To answer these questions it is necessary to realize that the discontinuities in the helix detected by restriction enzyme studies, and the maneuvers suggested by Callidine et al. (1982) to deal with the problem of steric clash, all ultimately affect the DNA phosphate backbone, and hence the dimensions of the major and minor grooves. Importantly, most of our knowledge about the frequency and effect of structural discontinuities in the helix has come from the studies of Drew and Travers who examined enzyme digestion patterns of DNAases which recognize variation in the helical backbone.

It had been shown prior to the initiation of their studies that two

different types of overall global phosphate backbone motion (observed from X-ray crystallography and fiber data) existed (summarized in Drew and Travers et al., 1984). In the first type, as the plane of the bases tilts from the horizontal in a positive or negative sense, the attached sugar-phosphate chains follow suit with one strand moving up as the other moves down. The result of this is a change in the diameter of the minor groove. A +20° tilt, for instance, opens the minor groove from 12 Å to 17 Å, while a -10° tilt closes it to 9 Å. The spacing across the major groove changes in reverse fashion. At the time the studies of Drew and Travers were initiated, it had been predicted that A/T rich regions would exhibit a very narrow minor groove because of their large propeller twist (Shakked et al., 1983; Fratini et al., 1982), while G/C runs would exhibit a wide minor groove (Wang et al., 1982).

The second type of global helix motion was first proposed by Arnott et al. (1983) to describe anomalous features of the poly(dA).poly(dT) fiber X-ray pattern, and is a model for describing sequence-induced radial asymmetry. This model states that if one strand of the helix moves inward toward the helix axis, such as a run of purines might do for better stacking interactions, the partner strand must then move outward. When this happens the distance between adjacent phosphate groups on the strand which has moved outward increases from 5.8 to 7.0 Å, and there is a concomitant increase in the helical radius from 8 to 10 Å.

Drew and Travers (1984) have examined at single-base pair resolution the interactions of two commonly used nucleases (DNAase I and DNAase II), and copper-phenanthroline which also cleaves the DNA backbone, in an

attempt to better understand sequence-induced variability in the helical backbone. DNAase I and DNAase II perform nucleophilic attack on DNA phosphates, while copper-phenanthroline, an intercalating agent, performs free-radical attack on DNA sugars. All three reagents produce sequence-dependent patterns of cleavage in duplex DNA without any obvious preference for a particular base or base pair step (such as the psoralen preference for 5'-TA dinucleotides). More importantly for the purposes of this discussion, all three of these reagents display a sensitivity to DNA backbone geometry.

When they studied the digestion patterns of all three reagents on a DNA of natural origin, the 160 bp tyrT promoter, they found sequence-dependent patterns of cleavage which provided evidence for structural polymorphism in both groove width and helical radius as well as local variation in phosphate accessibility. DNAase I and copper-phenanthroline digestion patterns were found to provide good evidence for sequence-induced changes in groove width. Both reagents cut at equal rates if phosphates face one another across the minor groove, but the rate of attack falls off when the bases inside the minor groove are either all A/T or G/C. Drew and Travers suggest that these reagents cut rapidly only when the width of the minor groove is at an intermediate value near 12Å, and falls off slowly as the spacing becomes either larger or smaller.

DNAase II digestion patterns from this same study provide good evidence for the model of sequence-induced radial asymmetry. In these studies the rate of attack was negatively correlated across any base pair

step, with maximum digestion at runs of purines, and low levels at the complementary pyrimidine sequences. In view of the prediction of Arnott et al. (1983) that purines will move in toward the helix axis to allow for better stacking interactions while the partner pyrimidine strand must move outward, the simplest interpretation of these data is that the enzyme cuts well when the adjacent phosphate spacing is near the 5.8 Å optimum (the most symmetric configuration), and more slowly as the spacing increases to 7.0 Å.

The two previous variations in helix geometry have been described as global. DNAase I and DNAase II also detect sequence-induced local helix conformational changes. A final observation from this study is that the local rates of attack for DNAase I and II are strongly out of phase with one another at any given phosphate; i.e. for a given sugar on one strand, if DNAase I cuts the 3'-P bond well, then DNAase II cuts the 5'-P bond poorly. The rationale for this is that phosphate orientation depends to some degree on adjacent base stacking interactions. Whenever the 3'P points outward so that DNAase I has easy access, the 5'-P points upward so that the accessibility for nucleophilic attack by DNAase II is difficult.

In a separate study, Drew and Travers also measured the DNAase I sensitivity of a 19 bp oligonucleotide of the sequence $(dA)_{11}(dG)_8$ (1985). Sequence dictates that the conformation of this molecule will switch from one extreme (narrow minor groove) to the other (wide minor groove) resulting in a structural discontinuity. These experiments addressed the question of whether this structural change was abrupt or

spread over several bases on either side of the junction. This study has implications for the question addressed by Manuscript 1 of this thesis, namely how many base pairs do such structural discontinuities span. It also bears directly on the distortional effect of sequences within the SV40 regulatory region.

DNAase I, as was noted previously, recognizes the spacing between sugar-phosphate chains as measured across the minor groove (Drew and Travers, 1984; Drew, 1984). It cuts poorly when the spacing is especially narrow or wide, but well when the spacing is intermediate. When the oligonucleotide was cloned as a dimer into pUC19 and digested with DNAase I, their results show that the enzyme cut poorly in the AT and GC rich regions, but well at their junctions. When the experiment was repeated with a naturally occurring 260 bp fragment from sea urchin 5S RNA with a sequence of $G_{12}AGA_8$, the results showed that the transition region extended at least 5 base pairs.

E. Biological Significance of Structural Discontinuities.

SV40 has emerged as a biological system which is particularly informative regarding the significance of structural discontinuities in gene regulation. Within the 400 bp which span the regulatory region, there are, first, the obvious perturbations of the helix which were discussed in some detail in Manuscript 2. These include the two sequence directed bends, and a region of potential Z DNA within the enhancers. As noted in Manuscript 3, this same region is devoid of nucleosomes in a subpopulation of minichromosomes isolated at late times postinfection as

well as in packaged virus. Cellular factors do not appear to be required for this phenomenon, since in vitro assembly studies discussed in Manuscript 2 show that DNA sequences in this region resist nucleosome formation (Wasylyk et al., 1979; Hiwasa et al., 1981). Irradiation studies on purified DNA summarized in Manuscript 2 point to the existence of a structural discontinuity that confers on sequences within the enhancers and about 150 bp upstream a higher degree of psoralen accessibility. The studies summarized in Manuscript 3 comparing virus particles and SV40 chromatin samples with supercoiled and linear samples irradiated in vitro argue against changes in linking number or tilt playing singular key roles in the psoralen hypersensitivity of this region.

The recently reported conclusions of Muller et al. (1987) are particularly interesting in light of our findings. They report evidence for altered DNA conformation in SV40 by examining the site specific cleavage of the chiral metal complex Λ-tris(4,7-diphenyl-1,10-phenanthroline)cobalt (III) [abbreviated Λ-$Co(DiP)_3^{3+}$]. Steric constraints prevent intercalation of the Λ isomer of diphenylphenanthroline complexes into right-handed B-DNA. Strictly speaking, any stacked conformation that locally untwists or significantly unwinds the duplex, thereby increasing the dimensions of the minor groove, should yield specific cleavage, with DNA unwound to the point of a left-handed twist being the most susceptible. Overall, their data determined that cleavage sites of this complex appeared to correlate with alternating purine/pyrimidine stretches that border coding regions in

viral DNA. Their closest correlations were at splice sites. They also found an unusual cleavage pattern in the regulatory region: in the 72-bp repeats of the enhancer, two major sites of reaction with Λ-$Co(DiP)_3^{3+}$ are found at approximate positions of 124-133 and 196-205. These two cleavage sites flank two of the three 8-bp alternating purine/pyrimidine sequences suggested by Nordheim and Rich to contain Z-DNA (1983). See figure 3 for data summary. Additionally, these sequences span the region designated by studies in Manuscript 2 as psoralen hypersensitive. Since specific cleavage was not observed in the region of 258-265 which corresponds to the third 8-bp purine/pyrimidine alternating sequence described by Nordheim and Rich (1983), this probe is clearly detecting something other than just regions with a wide minor groove. Also, since the study of Nordheim and Rich only detects Z-DNA on molecules which are highly supercoiled (31.6 supercoils per molecule), it is unlikely that this probe is detecting regions of Z-DNA. Muller et al. (1987) also found strong cutting at the 5' end of each of the six GC boxes within the early promoter, a region which our studies determined was particularly insensitive to psoralen binding. The nature of the helical discontinuity detected by this probe may be related to a number of things, and follow-up studies with chromatin may prove more informative.

On a theoretical note, Nussinov et al. (1984) have applied the Calladine rules and Dickerson summation functions to the study of DNA sequences in the SV40 regulatory region, with particular attention to the regions of DNAase I sensitivity (see figure 3 for data summary). Their calculations indicate that DNAase I sensitive and hypersensitive sites in

chromatin are correlated with regions of successive, large, helical-twist variation from regular B-DNA. Their studies are based on the work of Saragosti et al., (1980) who showed that sequences from position 5210-360 were particularly sensitive to DNAase I. Within this region are several hypersensitive sites, all but one of which is correlated with a region of high helical twist. More to the point, five of these DNAase I hypersensitive/high helical twist regions are between nucleotide positions 115 and 300 and lie within the region we have determined is particularly psoralen accessible. Only the high helical twist site at nucleotide 35 falls outside the psoralen accessible region. The region at nucleotide 35 is A/T rich, has a narrow minor groove and a sequence-directed bend (Deb et al., 1986). In the studies of Nussinov, particular attention was paid to the enhancers which, according to the Dickerson summation functions, have conformational deviations from ideal B-DNA in at least three helical parameters: twist, roll, and torsion. When the entire SV40 molecule was examined, the twist variation at positions 120 and 180 in the enhancers was found to be 3.7 times higher than the average of the rest of the genome. A variation in helical roll or torsion would fit our data quite well. Since roll is the angle by which base pairs open up toward the minor groove (Callidine and Drew, 1984), and since the psoralen hydrophobic edge must enter from the minor groove in order to be properly oriented for photocycloaddition (Zhen et al., 1986), a change in that parameter would be expected to alter psoralen accessibility. A change in torsion would result in the separation of purines by sliding one or both of the base-pairs along

their long axes so that the purines were partially pulled out of the base pairs stack. This would assuredly affect psoralen intercalation, and probably the activity of the chiral metal complex used by Muller et al. (1987).

A tremendous amount of work is yet to be done before a clear picture will emerge regarding the significance of structural variations in the DNA helix. Collectively however, the studies and experiments summarized in this discussion form the beginnings of a framework which support the claim that sequence-associated variations occur frequently in the normal B DNA helix, and are important for gene regulation.

Alves, J., Pingoud, A., Haupt, W., Langowski, J., Peters, F., Maass G., and Wolff, C. (1984). The influence of sequences adjacent to the recognition site on the cleavage of oligodeoxynucleotides by EcoRI endonuclease. Eur. J. Biochem. 140, 83-92.

Armstrong, K. and Bauer, W.R. (1982). Preferential site-dependent cleavage by restriction endonuclease Pst I. Nucl. Acids Res. 10, 993-1007.

Armstrong, K., and Bauer, W.R. (1983). Site-dependent cleavage of pBR322 DNA by restriction endonuclease Hinf I. Nucl. Acids Res. 11, 4109-4126.

Arnott, S., Chandrasekaren, R., Hall, I.H., and Puigjaner, L.C. (1983). Heternonomous DNA. Nucl. Acids Res 11, 4141-4155.

Callidine, C.R. (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. J. Mol. Biol. 161, 343-352.

Callidine, C.R. and Drew, H.R. (1984). A base-centered explanation of the B-to-A transition in DNA. J. Mol. Biol. 178, 773-782.

Clemens, S.L. (Mark Twain) (1875). Old times on the Mississippi. Atlantic Monthly. 36, 190-193.

Courey, A.J. and Wang, J.C. (1983). Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. Cell 33, 817-829.

Deb., S., DeLucia, A., Koff, A., Tsui, S., and Tegtmeyer, P. (1986). The adenine-thymine domain of the simian virus 40 core origin directs DNA bending and coordinately regulates DNA replication. Mol. and Cell. Biol. 6, 4578-4584.

Dickerson, R.E. (1983). Base sequence and helix structure variation in B and A DNA. J. Mol. Biol. 166, 419-441.

Dickerson, R.E., and Drew H.R. (1981). Structure of a B-Dodecamer II. Influence of base sequence on helix structure. J. Mol. Biol. 149, 761-786.

Dickerson, R.E., Drew, H.R., Conner, B.N., Kopka, M.L., and Pjura, P.E. (1982). Helix Geometry and Hydration in A-DNA, B-DNA and Z-DNA. Cold Spring Harbor Symposia on Quantitative Biology 47, 13-24.

Drew, H.R. (1984). Structural specificities of five commonly used DNA nucleases. J. Mol. Biol. 176, 535-557.

Drew H.R., and Travers, A.A. (1985). Structural junctions in DNA: the influence of flanking sequence on nuclease digestion specificities. Nucl. Acids Res. 13, 4445-4467.

Drew H.R., and Travers, A.A. (1984). DNA structural variations in the E. coli tyrT promoter. Cell 37, 491-502.

Fratini, A.V., Kopka, M.L., Drew, H.R., and Dickerson, R.E. (1982). Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. J. Biol. Chem. 257, 14,686-14,707.

Gamper, H., Piette, J., and Hearst, J.E. (1985). Efficient formation of a crosslinkable HMT monoadduct at the Kpn I recognition site. Photochem. Photobiol. 40, 29-34.

Hiwasa, T., Segawa, M., Yamaguchi, N., and Oda, K.I. (1981). Phasing of nucleosomes in SV40 chromatin reconstituted in vitro. J. Biochem. 89, 1375-1389.

Hogan, M., Dattagupta, N., and Crothers, D.M. (1978). Transient electric

dichroism of rod-like DNA molecules. Proc. Natl. Acad. Sci. USA 75, 195-199.

Isaacs, S.T., Wiesehahn, G., and Hallick, L.M. (1984). In vitro characterization of the reaction of four psoralen derivatives with DNA. Natl. Cancer Inst. Monogr. 66, 21-30.

Levitt, M. (1978). How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. Proc. Natl. Acad. Sci. USA 75, 640-644.

Muller, B.C., Raphael, A.L., and Barton, J.K. (1987). Evidence for altered DNA conformations in the simian virus 40 genome: Site-specific DNA cleavage by the chiral complex Λ-tris(4,7-diphenyl-1,10-phenanthroline)cobalt (III). Proc. Natl. Acad. Sci. USA 84, 1764-1768.

Nordheim A., and Rich, A. (1983). Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature 303, 674-679.

Nussinov, R., Shapiro, B., Lipkin, L.E., and Maizel. J.V. (1984). DNAase I hypersensitive sites may be correlated with genomic regions of large structural variation. J. Mol. Biol. 177, 591-607.

Pearlman, D.A., Holbrook, S.R., Pirkle, D.H., and Kim, S.-H. (1985). Molecular models for DNA damaged by photoreaction. Science 227, 1304-1308.

Saragosti, S., Moyne, G., and Yavin, M. (1980). Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. Cell 20, 65-73.

Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W.B.T., Salisbury, S.A., and Viswamitra, M.A. (1983). Sequence dependent conformation of an A-DNA double helix: the crystal structure of the octamer d(GGTATACC). J. Mol. Biol. 166, 183-201.

Thomas, M. and Davis, R. (1975). Studies on cleavage of bacteriophage λ DNA with EcoRI restriction endonuclease. J. Mol. Biol. 91, 315-328.

Wang, A.H.-J., Fujii, S., van Boom, J.H., van der Marel, G., van Boeckel, S.A.A. and Rich A. (1982). Molecular structure of r(GCG)d(TATACGC): a DNA-RNA hybrid helix joined to double helical DNA. Nature 299, 601-604.

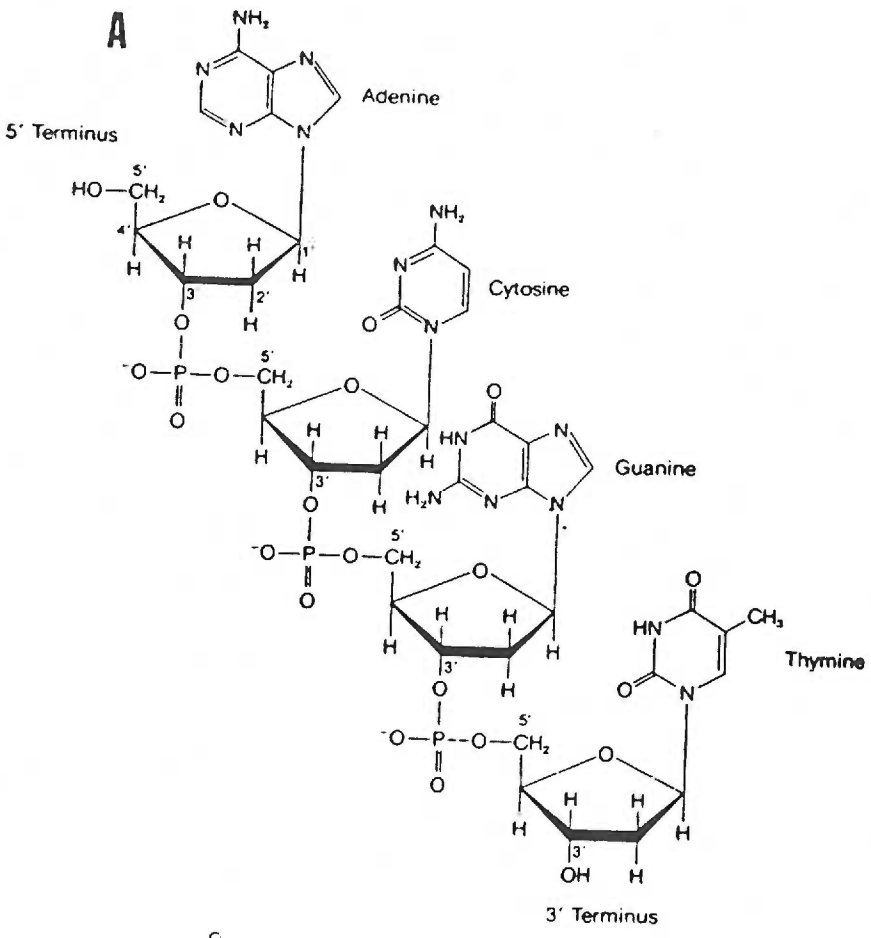
Wang, A.H.-J. Fujii, S., van Boom, J.H., and Rich A. (1982) Right-handed and left-handed double-helical DNA: Structural studies. Cold Spring Harbor Symposia on Quantitative Biology. 47, 33-44.

Wasylyk, B., Oudet, P., and Chambon, P. (1980). Preferential in vitro assembly assembly of nucleosome cores on some A-T rich regions of SV40 DNA. Nucl. Acids Res. 7, 705-713.

Wiesehahn, G.W., and Hearst, J.E. (1978). DNA unwinding induced by photoaddition of psoralen derivatives and determination of dark-binding equilibrium constants by gel electrophoresis. Proc. Natl. Acad. Sci. USA 75, 2703-2707.

Zhen, W.-P., Buchardt, O., Nielsen, H., and Nielsen, P.E. (1986). Site specificity of psoralen-DNA interstrand cross-linking determined by nuclease Bal 31 digestion. Biochemistry 25. 6598-6603.

Figure 1. Basic structure of B-form DNA molecule. Panel A shows a polynucleotide segment consisting of the bases adenine, cytosine, guanine, and thymine. Panel B shows view of the minor groove of a B-DNA dodecamer. Note that the bases within the helix do not lie flat on a plane. Panel B is adapted from the reference of Fratini et al., 1982.


A
Adenine
5′ Terminus
Cytosine
Guanine
Thymine
3′ Terminus

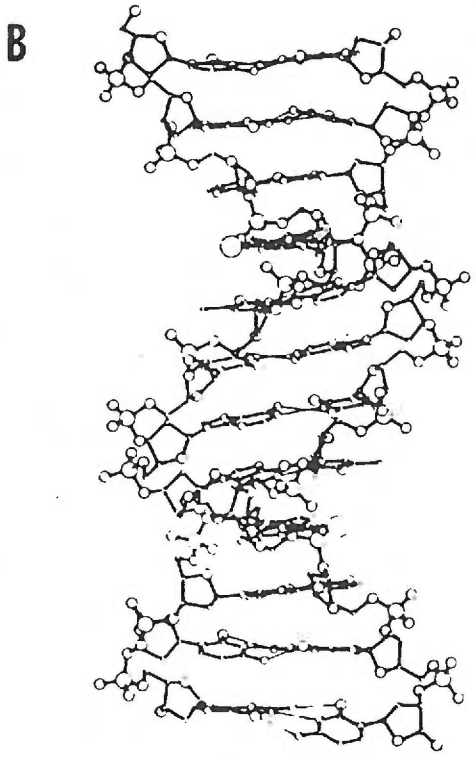
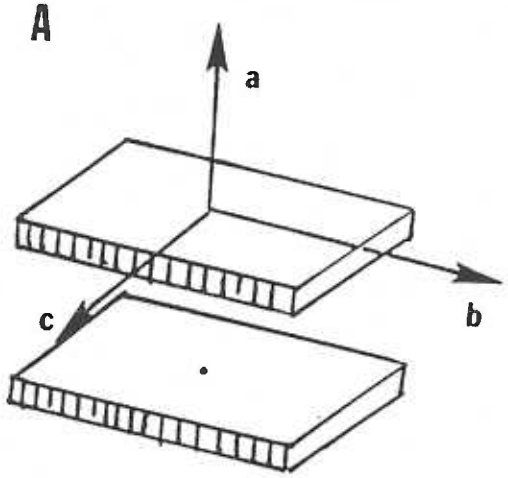
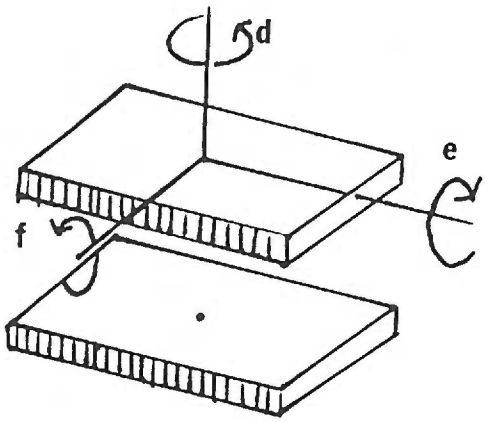

B

Figure 2. Structure of isolated dinucleotide. Panel A shows the three translational degrees of freedom that describes each base pair. Important for purposes of this discussion is b which indicates slide as described in the text. Panel B shows the three rotational degrees of freedom that also describes each base pair. The letters d, e, and f correspond to twist, roll and tilt as they are described in the text. Panel C is a schematic drawing of B DNA indicating the way in which propeller twist induces clash of purine ring edges within the minor groove for the base sequence C-G. The phosphate backbone between the base pairs is symbolized by a broken line. The helix axis runs vertically through each of the guanines. Large arrow indicates region of clash. Figures adapted from schematics in references of Dickerson, 1983 and Callidine and Drew, 1984.

A



B



C

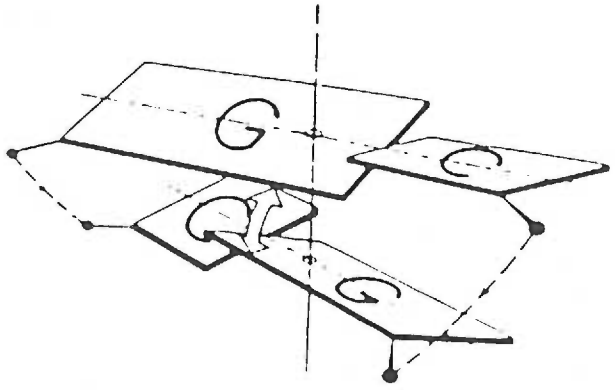
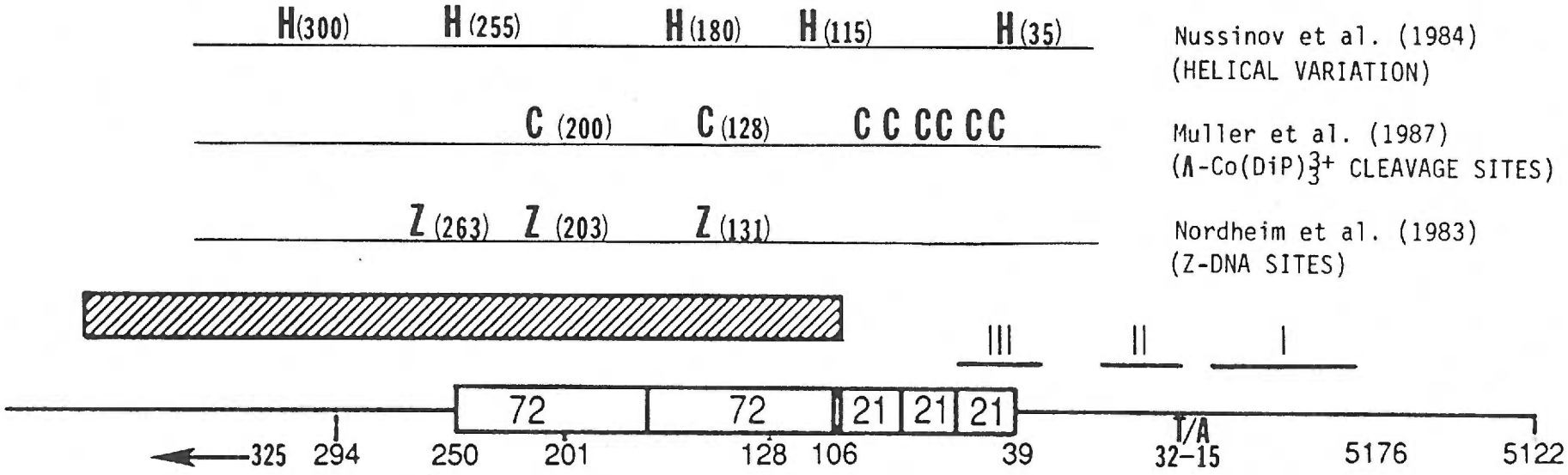


Figure 3. Summary scheme comparing data of Nussinov et al. (1984), Muller et al.(1987), Nordheim and Rich (1983), and this thesis. The psoralen hypersensitive site detected by this thesis is indicated by the striped rectangle. The three putative regions of Z-DNA detected by Nordheim and Rich (1983) are indicated by the letter Z. $Co(DiP)_3^{3+}$ cleavage sites detected by Muller et al. (1987) are indicated by the letter C. Nucleotide numbers are indicated for those within the enhancer region, but not those within the 21 bp repeats. The regions of helical variation detected by Nussinov et al. (1984), which are calculated from the Dickerson summation functions are indicated by the letter H. Numbers indicate nucleotide positions reported.


H(300)
H(255)
H(180)
H(115)
H(35)
Nussinov et al. (1984)
(HELICAL VARIATION)
C (200)
C (128)
C C CC CC
Muller et al. (1987)
(Λ-Co(DiP)$_3^{3+}$ CLEAVAGE SITES)
Z (263)
Z (203)
Z (131)
Nordheim et al. (1983)
(Z-DNA SITES)
III
II
I
72
72
21
21
21
T/A
325
294
250
201
128
106
39
32–15
5176
5122

Appendix A

Mapping Schemes to Localize Psoralen Adducts on DNA Using Exonuclease III and Bal 31

Abstract

The first two schemes which we developed for mapping psoralen adducts on DNA were unsuccessful. Nevertheless, a great deal was learned from these attempts, hence the data is summarized in this appendix. The first approach employed E. coli exonuclease III which is inhibited by pyrimidine dimers as well as dichlorodiammine platinum II adducts. Since this inhibition is thought to be related to the single-strandedness generated in these regions by the DNA modification, and 5-MIP monoadducts unwind DNA by 18°, using this enzyme in a similar strategy seemed feasible. The second approach employed exonuclease Bal 31 which simultaneously degrades both the 5' and 3' termini of duplex DNA. We initially predicted that this enzyme would be inhibited by HMT adducts, and could be used in conjunction with single cut restriction enzymes to map both crosslinks and monoadducts.

The first scheme which we developed for mapping psoralen adducts was based on a successful approach suggested by Royer-Pokora et al. (1981) to map the sites of cis and trans dichlorodiammine platinum II adducts on DNA. This approach took advantage of the processive nature of E. coli exonuclease III which catalyzes the hydrolysis of several types of phosphoester bonds in DNA. This enzyme has a 3' to 5' exonuclease activity which releases 5'-mononucleotides from the 3' ends of DNA strands, and a 3' phosphatase activity which hydrolyzes 3'-terminal phosphomonoesters.

These experiments were done using a 311 bp SV40 Eco RII G restriction fragment purified from pSV07 which contains three copies inserted in tandem via Eco RI linkers (see figure 1). This fragment includes most of the SV40 regulatory region, with the exception of a small portion of the late enhancer. Following digestion with Eco RI and separation by agarose gel electrophoresis (BioRad ultra pure agarose, Richmond, CA), the 311 bp fragment was electroeluted into dialysis bags as described by Maniatis et al. (1981). The DNA was purified by extracting twice with phenol:chloroform:isoamyl alcohol (25:24:1), once with ether, and then ethanol precipitated. The sample was then irradiated to a level of 1 adduct/300 bp with ^{3}H-5-MIP which, as discussed in section I of the Introduction, forms primarily monoadducts (Bordin et al., 1975). Unincorporated ^{3}H-5-MIP was removed by extracting three times with chloroform:isoamyl alcohol (24:1) and ethanol precipitation.

The fragment was end-labeled at both Eco RI ends using the Klenow

fragment of E. coli polymerase I and 2'-deoxyadenosine 5'-O-1-^{35}-thiotriphosphate (dATP[α-^{35}S]) as described by Maniatis et al. (1981). Labeling with dATP[α^{35}S] was critical to the procedure since Putney et al. (1981) had previously determined that restriction fragments labeled in this manner are essentially "capped" and thus resistant to exonuclease III digestion. Moreover, this resistance is sufficiently great that if the cap is placed at only one end, just one strand of the double helix will be digested, resulting in production of an intact single strand.

The DNA was next cut with Ava II which removes a small 27 bp piece from one end of the molecule, yielding an asymmetrically labeled 284 bp fragment with two 5' overhangs. This fragment was treated with exonuclease III which will digest in the 3' to 5' direction only on the unlabeled strand (Yajko and Weiss, 1975; Richardson et al., 1964a; Richardon et al., 1964b).

Royer-Pokora et al. (1981) had previously demonstrated that exonuclease III digestion of DNA was impeded at sites of pyrimidine dimers as well as by the cis and trans isomers of dichlorodiamine platinum II. They suggested that since exonuclease III requires a double stranded substrate for activity, the termination may occur because of local denaturation of DNA in the vicinity of the pyrimidine dimers or dichlorodiamine adducts. Since 5-MIP unwinds the DNA by approximately 18° per adduct (Isaacs et al., 1982), we felt that the same rationale could readily be applied to the mapping of psoralen monoadducts.

Following exonuclease III treatment, the single-stranded tail was

removed with S1 nuclease, and the remaining double-stranded molecule denatured with heat and run on an 8M urea sequencing gel. In order to speed exposure time, the gel was fixed with 5% TCA, treated with Enhance (New England Nuclear, Boston MA), and then dried for 2 hours at 60° C before exposure to Kodak Omat XAR film for 2-10 days. If successful, this approach could be extended to the mapping of psoralen crosslinks. Following exonuclease III and S1 treatment, the crosslink could be photoreversed as described by Cimino et al. (1986) and the resultant product analyzed on denaturing sequencing gels.

After multiple attempts this approach was abandoned for the following reasons. First, as Shalloway et al. (1980) had observed while mapping SV40 TAg binding sites by exonuclease III protection, this enzyme does not digest double-stranded DNA in a completely processive manner. Like Shalloway, we observed a series of discrete limit digestion bands implying the existence of preferential sites at which the processive cleavage stops. The reason for this is not clear, but is probably related to the secondary structure of the DNA molecule. Since the fragment we were most interested in contains a great deal of secondary structure (bends, Z-DNA, etc.), this approach was clearly less than optimal.

Additional problems which made data analysis very difficult were related to substantial nicking of the DNA which occurred during the preparation procedure. This nicking was probably caused by low levels of contaminating short wave UV emitted during the photo-irradiation process, and/or by exposure of the ethidium stained gel to short wave UV during

the purification process. This created two unique sets of technical problems. First, S1 nuclease will digest any regions of single-strandedness, even if it is simply a nick (Beck and Sharp, 1971; Vogt, 1973). This gave the appearance of discrete bands where there were no psoralen binding sites. Second, exonuclease III will initiate at nicks in double-stranded DNA. This means that on fragments which exonuclease III had already terminated because it encountered a monoadduct, the enzyme could reinitiate at a nick located further downstream, thus bypassing the monoadduct. Finally, in order to get data in a reasonable amount of time, the ^{35}S-gels needed to be fixed with 5% TCA and treated with Enhance (New England Nuclear, Boston MA). This process is time consuming and costly.

In view of these findings, we opted for a new approach and developed a second scheme using Bal 31 exonuclease. This enzyme possesses a highly specific single-stranded endodeoxyribonuclease activity as well as an exonuclease activity that simultaneously degrades both the 3'and 5' termini of duplex DNA (Gray et al., 1975; Legerski, et al., 1977). It is typically used to shorten DNA restriction fragments simultaneously from both ends, so the resulting fragments containing blunt ends can be ligated with T4 DNA ligase.

Our approach required purifying relatively large pieces of DNA and irradiating with HMT. Since crosslinks perturb the helical structure of DNA more than monoadducts, we reasoned that they would be more likely to inhibit Bal 31 exonuclease activity. The digestion products would then be analyzed on high resolution non-denaturing gels. Since this scenario

did not require S1 nuclease and the digestion products were analyzed by non-denaturing sequencing gels, the level of internal nicking in the fragment was irrelevant.

Our first task was to determine gel conditions that would allow high resolution of double-stranded molecules of similar size. We found that 6% polyacrylamide gels prepared with 1% glycerol, and run at 500V (80 mAmps) allowed separation of DNA molecules which differed in size by approximately 3 base pairs. Our next task was to develop a method that would facilitate orientation of bands so that the psoralen crosslinking sites could be determined. It would obviously be advantageous if Bal 31 exonuclease were inhibited by a blocking end-group such as dATP[α]^{35}S. Consequently, we made several attempts to define conditions to differentially inhibit the exonuclease activity of Bal 31. These experiments were uniformly unsuccessful.

Alternatively, we proposed that it was not necessary to have such a blocking group if a large enough piece of DNA was used for the starting material. The enzyme should then encounter psoralen adducts from both ends relatively early in the reaction which would arrest its activity. If this step were followed by digestion with an asymmetrically positioned restriction enzyme, the positions of psoralen binding sites at the opposite end of the molecule could easily be determined. Unfortunately, under the conditions which we employed, the HMT irradiated sample was very refractory to digestion with Bal 31. In retrospect, it is possible that the level of adducts/bp was very high, and Bal 31 was inhibited very early in the course of digestion, yielding products which differed so

slightly in size that they were not separable in the double stranded sequencing system.

Zhen et al. (1986) have very recently used this same approach to successfully map 4,5',8-trimethylpsoralen adducts. Interestingly, when they checked their digestion products for renaturation, they observed complete renaturation, indicating that all remaining molecules contained an interstrand crosslink. Since the experiments were done with samples that contained less than one adduct/fragment, their data argues that Bal 31 is arrested only by crosslinks and not by monoadducts.

In retrospect, this approach is not readily applicable to mapping psoralen adducts in the SV40 origin region. While our single-stranded sequencing gels had a resolution level of about 3 bp; this fact was deduced from gels in which the DNA did not contain psoralen adducts. We have since determined that the presence of such adducts, particularly crosslinks, can greatly affect the mobility of the DNA, and under the best of conditions yields bands which span 2-3 bp (see manuscript 2). This fact, compounded with the 3 bp resolution of double stranded polyacrylamide gels, means that mapping of psoralen adducts with this method is only accurate to within 5-7 base pairs. Hence, discernment of individual psoralen binding sites within an A-T rich region such as that which spans nucleotides 14-20 of the SV40 origin of replication is not possible, as it has been using the approach described in manuscript 2.

Bordin, F., Marciani, S., Baccichetti, F., Dall'Acqua, F., and Rodighiero, G. (1975). Studies on the photosensitizing properties of angelicin, an angular furocoumarin forming only monofunctional adducts with the pyrimidine bases of DNA. Ital. J. Biochem. 24, 258-267.

Beck, A.J., and Sharp, P.A. (1971). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1-endonuclease digested hybrids. Cell 12, 721-732.

Cimino, G.D., Shi, Y.B., and Hearst J.E. (1986). Wavelength dependence for the photoreversal of a psoralen-DNA-crosslink. Biochemistry 25(10), 3013-3020.

Gray, H.B., Ostrander, D.A., Hodnett, J.L., Legerski, R.J., and Robberson, D.L. (1975). Nucl. Acids Res. 1459-1492.

Isaacs, S.T., Wiesehahn, G., and Hallick, L.M. (1984). In vitro characterization of the reaction of four psoralen derivatives with DNA. Natl. Cancer Inst. Monogr. 66, 21-30.

Kim, S.-H., Peckler, S., Graves, B., Kanne, D., Rapoport, H., and Hearst, J.E. (1982). Sharp kink of DNA at psoralen-cross-link site deduced from crystal structure of psoralen-thymine monoadduct. Cold Spring Harbor Symposia on Quantitative Biology. 47, 361-365.

Legerski, R.J., Hodnett, J.L., and Gray, H.B. (1977). Extracellular nucleases of Pseudomonas Bal 31. III. Use of the double-stranded deoxyriboexonuclease activity as the basis of a convenient method for mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucl Acids Res. 5, 1445-1463.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Putney, S.D., Benkovic, S.J., and Schimmel, P.R. (1981). A DNA fragment with an α-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated in vivo. Proc. Natl. Acad. Sci. USA 78, 7350-7354.

Richardson, C.C., and Kornberg, A. (1964). A deoxyribonucleic acid phosphatase-exonuclease from Escherichia coli. I. Purification of the enzyme and characterization of the phosphatase activity. J. Biol. Chem. 239, 242-250.

Richardson, C.C., Lehman, I.R., and Kornberg, A. (1964). A deoxyribonucleic acid phosphatase-exonuclease from Escherichia coli. II. Characterization of the exonuclease activity. J. Biol. Chem. 239, 251-258.

Royer-Pokora, B., Gordon L.K., and Haseltine, W. (1981). Use of exonuclease III to determine the site of stable lesions in defined sequences of DNA: the cyclobutane pyrimidine dimer and cis and trans dichlorodiamine platinum II examples. Nucl. Acids Res. 9, 4595-4609.

Shalloway, D., Kleinberger, T., and Livingston D.M. (1980). Mapping the SV40 DNA replication origin region binding sites for the SV40 T antigen by protection against exonuclease III digestion. Cell 20, 411-422.

Vogt, V.M. (1973). Purification and further properties of a

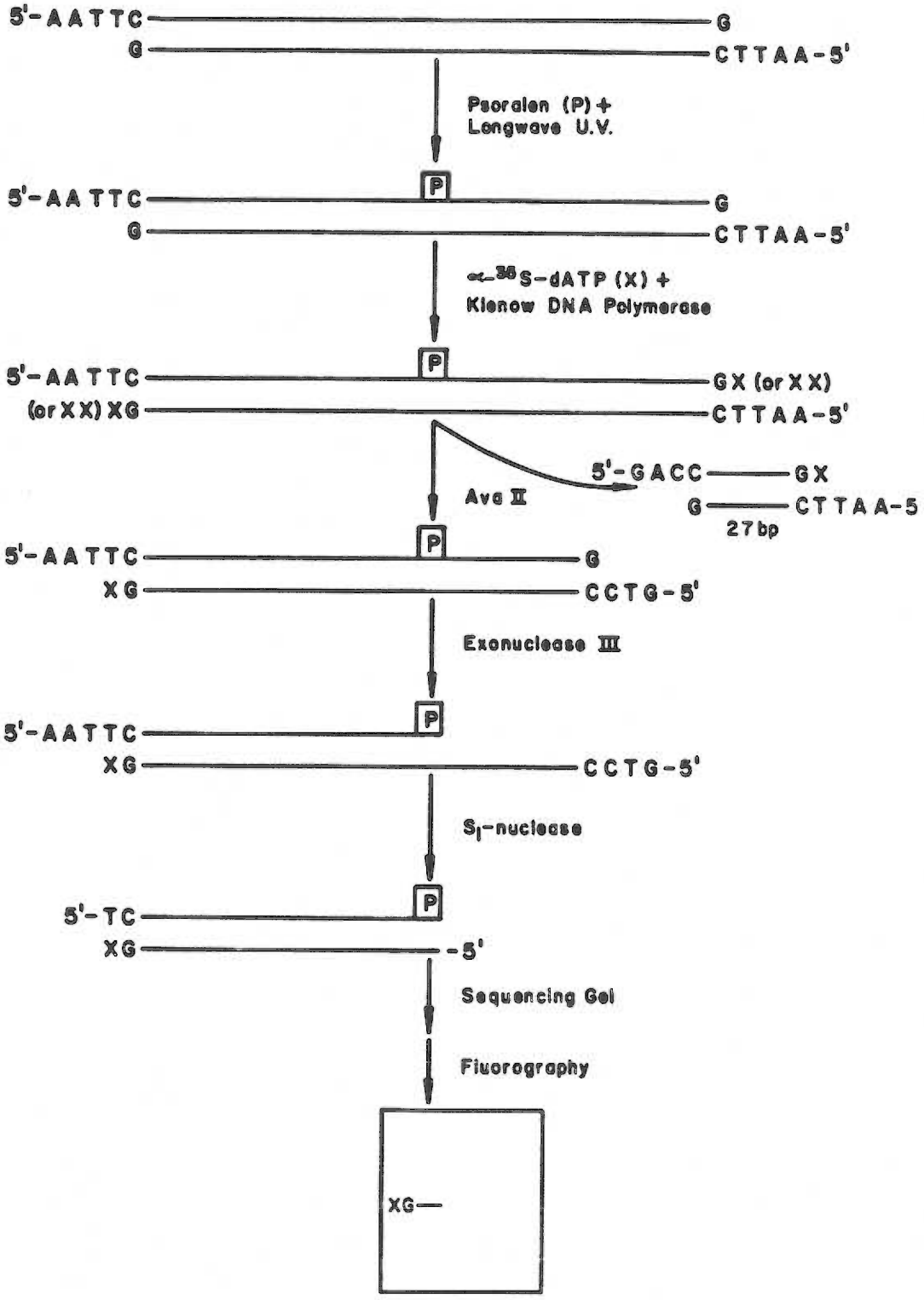
single-stranded specific nuclease from Aspergillus oryzae. Eur. J. Biochem. 33, 192-200.

Yajko, D.M. and Weiss, B. (1975). Mutations simultaneously affecting exonuclease II and exonuclease III in Escherichia coli. Proc. Natl. Acad. Sci. USA 72, 688-692.

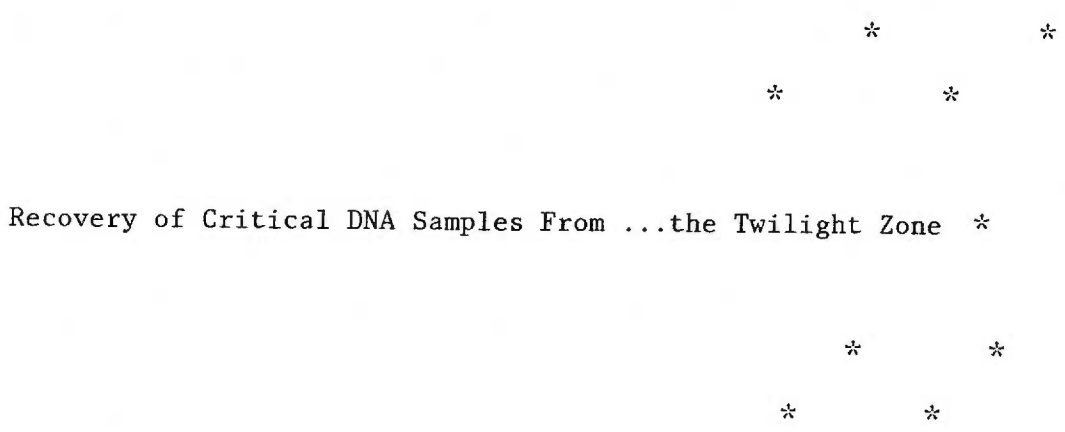
Zhen, W-P., Buchardt, O., Nielsen, H., and Nielsen, P.E. (1986). Site specificity of psoralen-DNA interstrand cross-linking determined by nuclease Bal 31 digestion. Biochemistry 23, 6598-6603.

Figure 1. Psoralen Mapping Scheme Employing Exonuclease III. A 311 bp fragment spanning the SV40 regulatory sequences was purified and irradiated with 5-MIP which only forms monoadducts. The DNA was then end-labeled with dATP[α^{35}P] using the Klenow fragment of DNA polymerase I. Ava II digestion removes a 27 bp piece from one end of the fragment, leaving a 284 bp molecule which is labeled at only one end. This fragment was treated with exonuclease III, and then S1 nuclease, yielding a double-stranded product which was denatured and sized on a standard sequencing gel.

SV40 Origin DNA (311bp)



Appendix B



ABSTRACT

It has been empirically noted by this lab and others that the risk of dropping, spilling, or irreversibly contaminating a DNA sample is directly proportional to its importance, multiplied by the total number of hours spent in preparation. Consequently, we have developed a procedure that allows for efficient recovery of DNA from states previously termed "unsalvagable." This method demonstrates efficient and quantitative retrieval of DNA samples accidentally deposited on lab bench paper.

Generally speaking DNA which is spilled, dumped or pipetted onto bench paper is considered unretrievable, and for that matter might as well have landed in a black hole. We have developed a simple and reproducible method, outlined below, which has assisted us in recovering such d-DNA (displaced DNA).

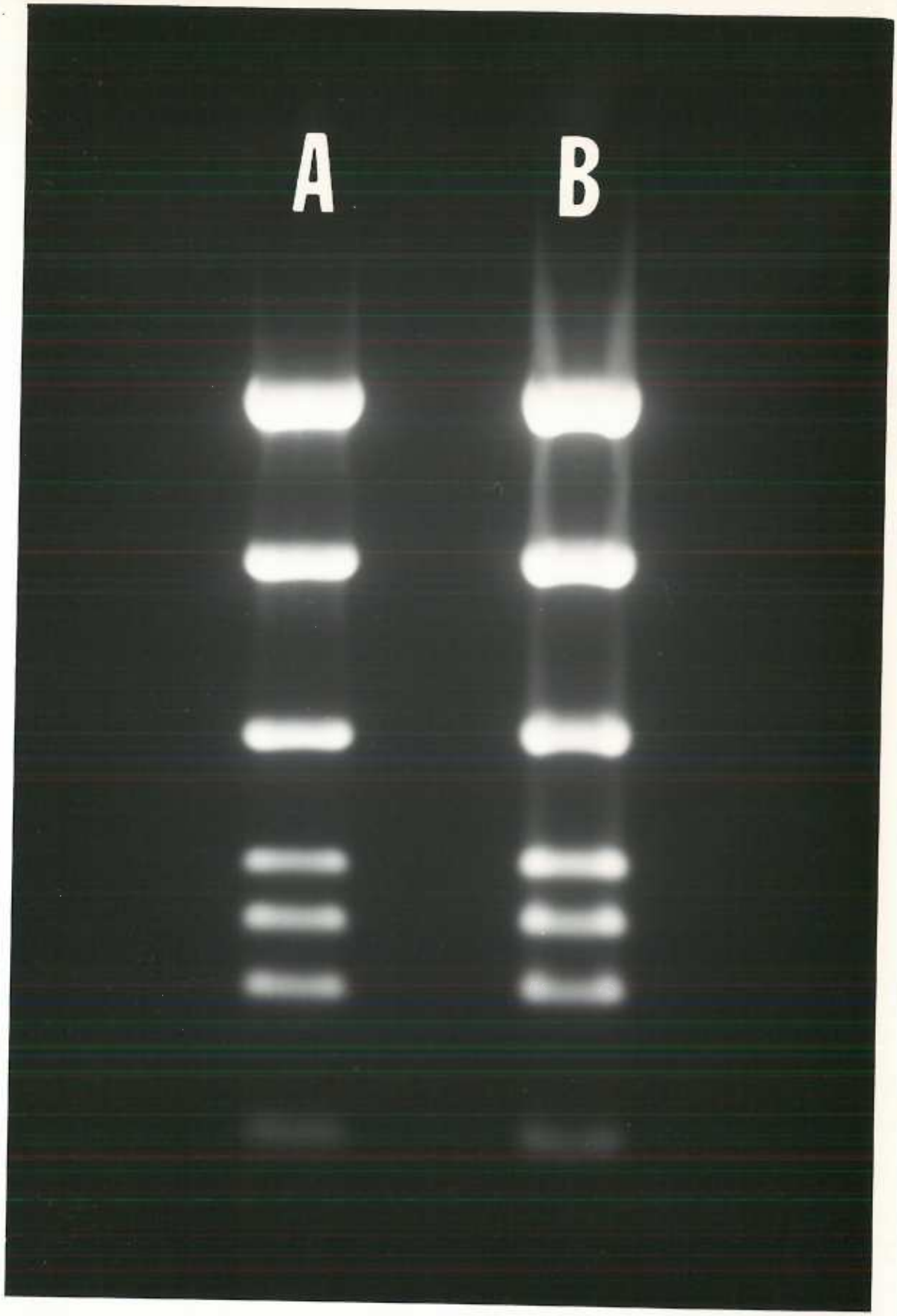
In these experiments, approximately 1.5µg of an extremely valuable SV40 Hind III-Kpn I digest was accidentally deposited on standard plastic-backed bench paper. The spill contained a total volume of 30 µl and the resulting "stain" was 2.5 cm in diameter. After a delay of approximately 2 minutes, the expanding circular spot was hastily cut out and then sub-cut (Latin derivative of subclone) into four equal quarters. Each quarter was folded into a cone and then placed, paper side out, into a siliconized 1.8 ml eppendorf tube containing 300 µl of 0.5X TBE (45mM Tris-borate, pH 8.3, 1.25 mM EDTA). The positioning of the plastic side of the bench paper facing inward appears to be critical for efficient recovery of d-DNA.

The samples were incubated at 37°C in a stationary heat block for one hour, and then with vigorous shaking for an additional 2 hours. We felt that heating above 37°C might partially degrade the paper or melt the plastic and thus jeopardize sample purity. The four tubes were then spun for 2 minutes in an Eppendorf 5414 microcentrifuge, and the aqueous volumes collected. Two aliquots of 150 µl of 0.5X TBE were added to each tube, the tubes vortexed vigorously, and the supernatants combined. The total aqueous sample was spun for 3 min. in the microcentrifuge to remove unidentified bits of solid debris. The total supernatant was then

concentrated to 1 ml using a Savant Speed Vac Concentrator at 42° C. An Amicon centricon^{TM} microconcentrator and three 2ml rinses of deionized water were used to purify and further concentrate the sample to a final volume of 30µl. If this final step was omitted and the aqueous sample taken directly to 30µl with the Speed Vac, the sample retained a viscous quality that made it both difficult and embarrassing to handle.

3 µl of 10X tracking dye was added to the sample which was loaded with the appropriate control on a 1.6% agarose gel and run for 13 hours at 30 volts. As can be seen from figure 1, the total recovery of d-DNA from bench paper was extremely high. Both the sample and control lanes were scanned in the fluorescence mode using a Helena Quick Scan R and D densitometer (Helena Laboratories, Beaumont, TX) equipped with an extended red-range photomultiplier. The relative size of each restriction fragment was not a critical factor in determining overall recovery. 81.9% of band 1 (1.8Kb) and 87.7% of band 6 (376 bp) were recovered, and we obtained an overall average recovery for the six major bands of 77%. Although we have not spectophotometrically analyzed any d-DNA samples, we have used such samples in end-labeling and exonuclease reactions with excellent results.

Figure 1. Quantitative analysis of 1.5 μg of an SV40 Hind III-Kpn I digest of d-DNA. After recovery, the experimental sample (lane A) and 1.5μg of an SV40 Hind III-Kpn I control (lane B) were analyzed on a 1.6% ethidum bromide-agarose gel run at 30 volts for 10 hours. Both lanes were scanned for fluorescence as described, and the percent recovery for each of the six major bands in the experimental sample determined.


A
B