Title

Photochemical crosslinking and mass spectrometric characterization of DNA-binding proteins and oligonucleotides

Author

Connor, Dallas Allan

Publication Date

1997

Peer reviewed|Thesis/dissertation

Photochemical crosslinking and mass spectrometric characterization of DNA-binding proteins and oligonucleotides

by

Dallas Allan Connor

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

To my parents

ACKNOWLEDGEMENTS

Special thanks to...

Dr. Martin Shetlar, who over the years has shown infinite patience with me. He has been a motivator and a mentor; an advisor and advocate; worldly and wise.

Dr. Arnold Falick, co-advisor and silent partner, who showed the same infinite patience. He always kept me on my toes. His dedication to my project was amazing.

Those who served on both my orals and thesis committee, Dr. James Cleaver for his irrepressible enthusiasm about my work, and Dr. Bradford Gibson for extending his advice and lab keys so that I could run all the mass spectrometry I wanted--late at night when no one could hear me scream in joy or agony.

To my labmates, both past and present, including Dr. Dennis Benjamin, Dr. Robert Cerpa; Dr. Lech Celewicz, whose curiosity was in abundance and advice was always welcomed; Yoko Haga, a friend, lab partner, fellow singer and disaster survivor, and co-founder of Vocal Chords Headquarters in room S-1014.

Drs. Ole Jensen and Douglas Barofsky for their work in this field and wonderful advice; Drs. Mark Young and Peter von Hippel who were instrumental in my laser crosslinking studies; Drs. Rajendra Prasad and Samuel Wilson, with their generous gifts of pol β protein.

Dr. Henry Schreiber, who fueled my interest in research when I was in high school; Dr. William Herndon, my undergraduate advisor who paved the way for my future. He would not let me off the hook with just a Bachelor's degree.

To the National Science Foundation for their financial support.

To the UCSF School of Pharmacy class of 1996, for making my second year so enjoyable and so memorable. To the UCSF School of Medicine class of 2000 for welcoming me in.

Margaret Valdez, my great-aunt who gave me a stack of science books when I was seven years old. Mr. Leroy Fulgenzi, my Spanish teacher.

To my friends: Stephen Miller--roommate for almost six (yes, six) years, Melinda Shockley, Marci Schaner, Ruihua Yuan. To Carol Lim, who for over half a decade patiently listened to every wild story I had to tell.

To my father's parents, Velvert and Helen Connor, who passed away eleven weeks apart in 1993.

To my mother's parents, Leandro and Frances Campos, who helped my mother out so very much.

To my parents, Allan Connor and Patricia Campos-Bustos.

ABSTRACT

Photochemical crosslinking and mass spectrometric characterization of DNA-binding proteins and oligonucleotides

by Dallas Allan Connor

Protein-nucleic acid interactions lie at the heart of an extraordinary number of essential biological processes. Part of the challenge in the fields of molecular biology and biochemistry is the elucidation of the molecular nature of these transient non-covalent interactions. Such knowledge could provide a foundation for further detailed structural analysis *in vitro* and *in vivo*, and would be an important first step in the development of therapeutic agents. Photochemical crosslinking of these complexes, which creates covalent bonds between proximal amino acids and nucleobases, is a well-established and attractive way of freezing non-covalent protein-nucleic acid interactions. One particular method of photochemical crosslinking, which uses a nanosecond-pulse laser as the ultraviolet (UV) light source, is a promising technique for examining interactions on time scales previously inaccessible by any other spectroscopic means. Mass spectrometry has recently emerged as a sensitive and versatile method for accurately measuring the molecular weights of biomolecules.

Here we present a novel approach toward determining the interface of protein-nucleic acid binding complexes. We have successfully used mass spectrometry to determine which amino acid was modified in a number of model peptide-nucleoside, peptide-nucleotide and peptide-dinucleotide systems which had been irradiated with UV light. We have also used mass spectrometry to determine the binding interface of the single-stranded DNA-binding domain of recombinant rat DNA polymerase β, using both low-intensity and nanosecond-pulsed laser UV crosslinking. Having established the

methodology for examining the binding sites at the peptide level, as well as the protocols for digestions of unreacted amino acids and nucleobases, we are now in a position to pinpoint protein-nucleic acid interactions at the amino acid level, with the view of applying our protocols to other protein-nucleic acid systems of interest.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

Protein-nucleic acid interactions lie at the heart of many essential biological processes. Among these are nucleic acid packaging, replication, rearrangement and repair, transcription and translation, protein scaffolding for ribosomal formation, as well as cellular maintenance and regulation. For instance, transcription factors can interact with promoter regions along the DNA (deoxyribonucleic acid) template, signaling where transcription should begin. In turn, polymerase complexes, sometimes consisting of multiple proteins and polypeptides with total molecular weights often exceeding one million, initiate transcription of the DNA template, and create an RNA (ribonucleic acid) sequence that would have one of a number of fates in the cell.

Another example of the interaction of proteins and nucleic acids on a massive scale is in replication of the genome in mitotic cell division. The fidelity and the integrity of the chromosomal DNA must be maintained in order to preserve the essential genetic information. As a result, DNA is folded and refolded into an extremely small volume *via* efficient packaging, and the billions of base pairs in the human genome are replicated, usually with minimal error, in a matter of hours, by protein complexes that can replicate thousands of base pairs *per second*. Such wonders of nature have been passed down for hundreds of millions--perhaps billions of years, and are essential processes for every living organism.

Genomic DNA can undergo a number of types of mutations arising from chemical cleavage and modifications, radiation-induced reactions, and the rare instance where proteins do create a base-pair mismatch. As a damaged residue or base-pair mismatch could potentially have catastrophic consequences for the cell or organism, elaborate

molecular machinery has been developed to recognize, bind specifically to and repair damaged DNA.

Maintenance of the genome is essential for survival of every living organism. Failure of any one of these types of protein-nucleic acid interactions in the cell could prove to be fatal for the cell, or even the organism itself. Such loss of control, whether intrinsic or induced, can result in such afflictions as genetic disorders, diseases such as acquired immunodeficiency syndrome (AIDS), and cancers in humans.

Of primary interest in the field of biochemistry and molecular biology is in the elucidation of these structures and the dynamic nature of such interactions at the molecular level. Such knowledge is essential for researchers to begin to understand the mysteries of life, and would provide vital information for therapeutic development.

1.1 PROTEIN-NUCLEIC ACID BINDING

1.1.1 GENERAL CONSIDERATIONS

It is known that protein-nucleic acid complexes involve a multitude of specific non—covalent interactions. These include van der Waals, hydrogen-bonding, and ionic interactions. The heterogeneity and chemical properties of amino acid side chains, as well as the nucleic acid residues, provide an amazing array of possibilities for non-covalent interaction; rules for determining the exact nature of given interactions *a priori* within a given protein-nucleic acid complex are essentially non-existent[1, 2]. However, some general statements can be made. The negatively-charged phosphate backbone of DNA, for instance, can interact with positively-charged amino acid side chains, including arginine and lysine, producing stable complexes without regard to the specific nucleic acid sequence. These kinds of interactions could be important for nucleic acid packaging, such as the formation of nucleosomes, or in case of the need to expose a single strand of DNA, which can be accomplished by any one of a number of single-stranded DNA

binding proteins. Aromatic amino acids, such as tyrosine and tryptophan could participate in binding by intercalation of their planar armoatic side chains into the DNA double helix. Glutamine and arginine have been widely implicated in specific DNA recognition: for example, crystal structures of *Eco*R I[1], the λ cro repressor[2], and the catabolite gene activator protein (CAP)[3] show arginines making a specific contact with guanine; structures of bacteriophage 434 repressor[4] and the λ repressor[5] show glutamine making specific contacts with adenine.

1.1.2 DOMAINS FOR BINDING

There is substantial experimental evidence that certain kinds of peptide motifs are also responsible for binding to nucleic acids. These include tandem repeats of basic and proline-rich peptide segments (*e.g.* SPKK and SPRK), zinc-fingers and helix-turn-helix motifs.

1.1.2.1 Basic and Proline-rich motifs

Many proteins contain tandem repeats of SPKK and SPRK peptide segments, along with variations of this motif, including GRP and KPK. These sequences are found in a number of chromosomal proteins, and appear to assume a special type of α-helical structure that could preferentially bind to minor grooves of A-T rich sites. The abundant lysines might serve as phosphate, hence nucleic acid, localizers, while the prolines can introduce kinks into the α-helix. Some specific examples include the termini in histone H1 and sea urchin sperm H2B[6, 7], and the repressor from bacteriophage 434[8, 9].

1.1.2.2 Zinc-fingers

Zinc-fingers have been shown to be an integral substructure that allows for specific, sequential binding to nucleic acids. In general, a zinc-finger consists of a zinc atom tetrahedrally-coordinated by two closely-spaced cysteine and two closely-spaced histidine residues. There are variations, however. A subclass of zinc-fingers replaces the histidines with cysteines, while a third class uses six cysteine residues to coordinate *two* zinc atoms. The region in between the coordinating residues can then "protrude" into the

major groove of DNA. A number of crystal and solution structures have been solved recently of complexes containing this type of motif, including the PML RING finger[10], the erythroid transcription factor GATA-1[11], the Myb DNA-binding domain[12], the p53 tumor suppressor-DNA complex[13], the Tramtrack DNA-binding domain[14], the glucocorticoid receptor[15], the five-finger GLI-DNA complex[16], the DNA binding domain of GATA-1[17], and the yeast transcription factor SWI5[18].

1.1.2.3 Helix-turn-helix motifs

Another substructure widely implicated in nucleic acid binding is the helix-turn-helix. This motif usually consist of two short α-helices separated by a glycine residue, which acts as a hinge. This allows key hydrophobic residues on each α-helix to come into contact with each other such that the system forms a compact tertiary substructure. While this structure is highly conserved, its relative orientation in the major groove of the DNA double-helix is quite variable. Some examples of proteins containing this motif include the DNA-binding domain from the *Kluyveromyces lactis* heat shock factor[19], the DNA-binding domain of the human ETS1 oncoprotein[20], the dimeric DtxR holo-repressor[21], the Oct-1 POU domain[22], Drosophila paired protein[23], *Thermus thermophilus* aspartyl tRNA-synthetase[24], the purine repressor with its corepressor, hypoxanthine[25], *Escherichia coli* trp repressor[26], the lac repressor[27], the DNA-binding domain of Myb[28], and the λ repressor[29].

1.1.2.4 Other motifs

Some proteins have been shown to bind to DNA *via* heterodimer formation. The "leucine zipper" and the "helix-loop-helix" motifs contain similar features: usually leucine (or sometimes valine) is repeated every third or fourth residue, thus allowing this relatively hydrophobic amino acid to lie on the same face of the α-helix. This helix then forms hydrophobic interactions with a similar (though not necessarily identical) polypeptide. This class of binding motifs allows for greater flexibility in recognizing and binding to specific non-palindromic sequences of DNA. Some examples include the

purine repressor[25], the MyoD basic-helix-loop-helix domain[30], the yeast transcriptional activator GCN4[31, 32] and RNase H[33].

1.2 X-RAY CRYSTALLOGRAPHY AND NMR

Structural studies of protein-nucleic acid complexes can be performed using two very powerful spectroscopic techniques: NMR and X-ray crystallography. A few of the complexes that have been recently solved by X-ray crystallography are the murine leukemia inhibitory factor[34], human uracil-DNA glycosylase[35], bacteriophage T4 gene 32 protein-DNA complex[36], the serum response factor[37], the Klenow fragment of *E. Coli* DNA polymerase I[38-40], EcoR V restriction endonuclease complexed with double-stranded DNA[41], DNase I[42], TFIID-τ[43], and DNA polymerase β complexed with a template primer[44-46].

Some recent complexes that have been solved by NMR spectroscopy are the N-terminal fragment of rat DNA polymerase β complexed with p(dT)8[47], *E. Coli* lac repressor[27], the POU binding domain[22], the LexA repressor DNA binding domain[48], the high mobility group (HMG) domain of the SRY gene[49], the Sso7d protein[50], the Myb protein[12, 28, 51] and the RNP domain[52].

Both NMR and X-ray crystallography have intrinsic limitations, however. In the case of NMR, which determines the average structure over a relatively long time, the proteins are required to be in a homogeneous environment. Structural studies are normally limited to proteins having a molecular weight of no larger than about 30 kDa, because the tumbling rates of these systems in solution cause substantial peak broadening, resulting in an inability to assign specific residues. In addition, isotopically-enriched proteins are required, (either ^{13}C, or ^{15}N, or both) which increase the signal-to-noise ratios of carbon and nitrogen atoms in NMR experiments, and allows for multi-dimensional heteronuclear experiments. In addition, many DNA-protein systems are

multimeric. And some of these systems are so labile and transient that it becomes increasingly difficult--if not impossible--to analyze over the time course of an NMR experiment.

X-ray crystallography has enjoyed much success recently in contributing to our knowledge of the molecular basis for protein-nucleic acid interactions. The "resolution" of the structure, *i.e.* the RMS deviations of positions of the atoms in the structure, can be high enough to allow for assignment of individual amino acid side chains. One of the limitations of X-ray crystallography is that the "structure" of the protein *in vivo* cannot *necessarily* be deduced from the crystal structure of the protein, because there are often physical constraints imposed upon the protein when crystallized. Proteins can adopt an extraordinary number of conformations *in vivo*, and it would not be prudent to surmise that any one structure could account for the distribution of structures that the system can adopt.

Another problem is posed by looped structures. Looped structures are regions of the protein which appear not to be involved with any kind of interaction with other regions of the protein, and thus can adopt a number of conformations both in solution and in a crystal lattice. This ensemble of structures makes assignment of individual residues very difficult. One example of a looped structure can be found in recombinant rat DNA polymerase β. The protein is composed of two subunits joined by a flexible strand. One of the subunits contains the catalytic core domain. The second subunit, located at the N-terminus, contains the region that binds single-stranded DNA. The crystal structure of the protein[44, 45] reveals that the hinged region is "floating" far above the rest of the catalytic domain. What is not known about this structure is how the entire complex behaves during the course of the addition of a nucleotide. This process, which could take place on the order of nanoseconds to microseconds, is essentially impossible to analyze directly by either X-ray or NMR.

Given these limitations, one could wish for a method that would analyze protein-nucleic acid complexes indirectly by "freezing" these transient non-covalent interactions in place. By creating covalent bonds between proximal amino acid and nucleic acid residues, we will have removed a degree of freedom in the interaction and effectively created a "snapshot" of this complex. This complex can then, in turn, be subjected to a host of analytical procedures that will elucidate the location of these covalent crosslinks. Having determined these contact regions, we will have overcome one of the major obstacles that often forbids detailed structural analysis. We describe such a method below.

1.3 PHOTOCHEMICAL CROSSLINKING

1.3.1 GENERAL CONSIDERATIONS

One method for determining protein-nucleic acid contact points is by inducing a covalent bond between two residues that are in close proximity. Photochemical crosslinking [53-55], which uses photons to form chemical bonds, is a promising method for providing insight into specific amino acid-nucleobase contact points, even if the structure of the complex is not known *a priori.*

In general, photochemical crosslinking of proteins and nucleic acids can be thought of as a photochemical reaction involving a nucleobase (usually a pyrimidine) and an amino acid. A photon strikes an organic molecule (*e.g.* thymine) which elevates an electron to a higher energy level. There are a number of competing fates for this excited system, which include radiative (fluorsecence, phosphorescence) and non-radiative (intersystem and intrasystem crossing) processes. However, in some cases an electron in the excited system (A*) can be transferred to the ground state of another species (B, Figure 1.1), which would, in essence, produce two radical species--one positively and one negatively charged. These radical species could then combine, thus forming a covalent

bond between the two. A number of photoreactions of nucleobases with protic species (*e.g.* alcohols) are thought to involve an electron transfer, and would thus be considered a model for a number of potential photochemical reactions that could occur between proteins and nucleic acids.

$$A^* + B \rightarrow [A^*B] \rightarrow A^{\bullet+} + B^{\bullet-} \rightarrow \text{radical combination} (\rightarrow \text{products})$$

Figure 1.1 Electron transfer

1.3.2 UV-CROSSLINKING OF AMINO ACIDS AND RELATED SUBSTRATES TO NUCLEIC ACID BASES

In order to elucidate the possible molecular nature of amino acid-nucleic acid crosslinks, studies have been done with model compounds that have similar structures to those of some amino acid side chains. In the case of serine and threonine residues, photoreactions of purines and pyrimidines with alcohols have been studied[56]. In general, adducts are formed by the radical addition of the alcohol at either the C5- or C6-position of the pyrimidine, with a saturation of the 5,6 double bond (Figure 1.2). In the case of purines, one observes additions at the C6- or C8-position[57-59] (Figure 1.3). As a typical example, consider the photochemical reactions of a protected uridine in methanol[60]:

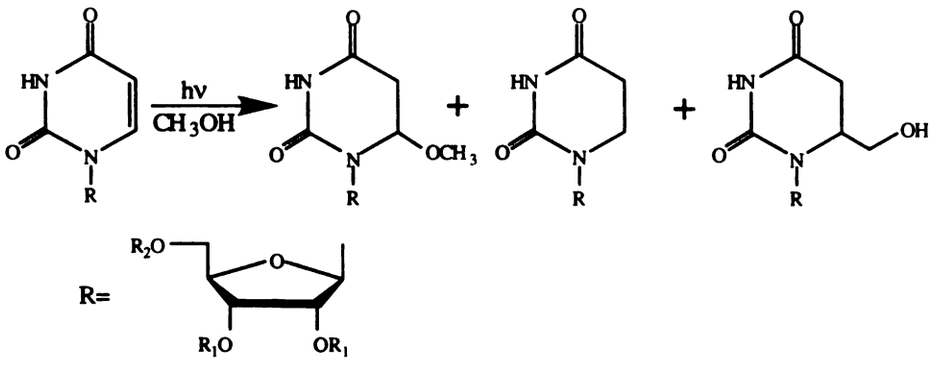


Figure 1.2 Photoreaction of a protected uridine in methanol

In the case of protected thymines, there are reports of a number of photoproducts formed when 1,3-dimethylthymine is irradiated in the presence of ethanol and isopropanol. In all cases involving the isopropanol irradiation, the photoproducts formed involved saturation of the 5,6 double bond[61]. In a subsequent report, however, irradiation of 1,3-dimethylthymine in the presence of ethanol, two of the four products that formed did not involve such saturation[62].

Both adenine and guanine (and their corresponding nucleosides) have been irradiated in the presence of isopropanol. In all cases reported, photoaddition occurs at the 8-position of the purine ring[57, 58] (Figure 1.3).

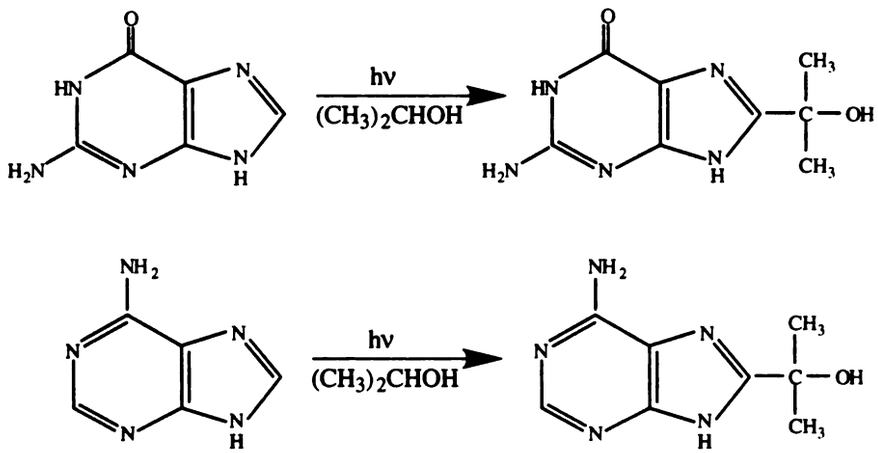
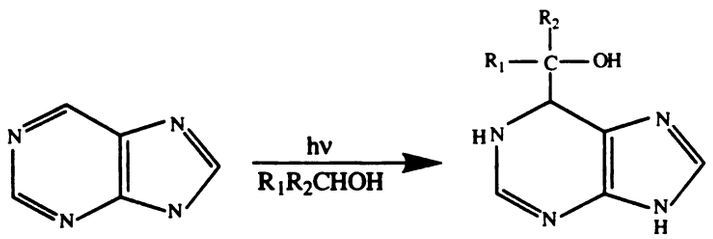


Figure 1.3 Photoreaction of adenine and guanine in isopropanol

Purine itself, however, reacts with alcohols by photoaddition at the 6-position[59]:



$R_1 = R_2 = H$
$R_1 = CH_3, R_2 = H$
$R_1 = R_2 = CH_3$

Figure 1.4 Photoreaction of purine in methanol

In order to evaluate the potential reactivity of amino acids for involvement in protein-nucleic acid crosslinking, studies were carried out with all 20 of the naturally occurring amino acids and DNA[63-65]. Shetlar and co-workers have shown that 15 of the 20 amino acids, and all glycyl- containing dipeptides can potentially react photochemically with DNA, suggesting that a majority of proximal residues can be photochemically crosslinked.

In the case of cysteines participating in photochemical reactions with nucleic acid bases, Shetlar and Hom[66] have shown that the irradiation of thymine in the presence of cysteine gives rise to one major product, 5-S-L-cysteinyl-5,6-dihydrothymidine, which is readily separable into two diastereomers, and two other products.

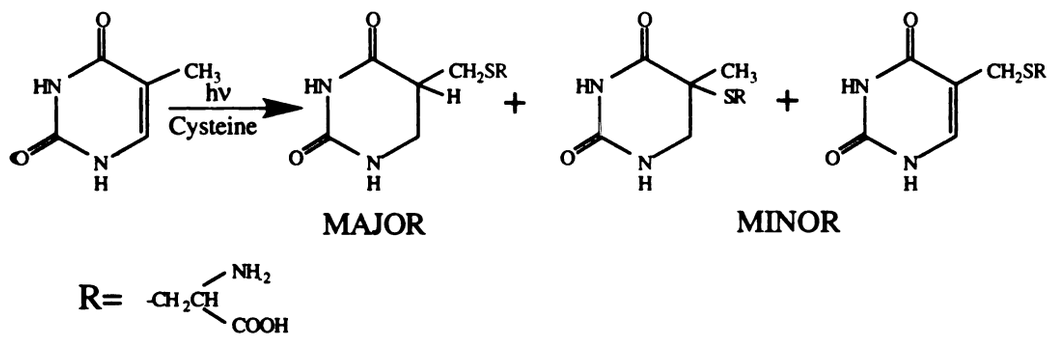


Figure 1.5 Photoreaction of thymine with cysteine

Another example of a photoaddition between a specific amino acid and a nucleic acid base that has been extensively characterized is that of tyrosine and thymidine. It is thought that tyrosine, as well as phenylalanine and tryptophan, by virtue of their planar side chains, can participate in binding to double-stranded DNA by intercalation[20, 67, 68]. The hydroxyl substituent could also participate in binding by forming hydrogen bonds with the nucleic acid bases. Work done by Shaw and co-workers demonstrated that N-acetyltyrosine could be photochemically crosslinked to thymine and thymidine[69]. Even more interesting, as it became the basis for work by the author of

this thesis, is that the tyrosine residue, when incorporated into a small peptide (Angiotensin I, amino acid sequence NRVYIHPFHL) and irradiated in the presence of thymidine, was shown *via* LSIMS(+) and tandem mass spectrometry (discussed later in the Introduction) to be covalently bound to thymidine[69].

A plausible mechanism for the formation of the covalent bond is shown in Figure 1.6. The excited tyrosine residue is thought to transfer an electron and a proton to the thymidine, and by radical recombination, a covalent bond is formed between the *ortho* carbon of the tyrosine phenolic ring and the 5-position of thymidine[69].

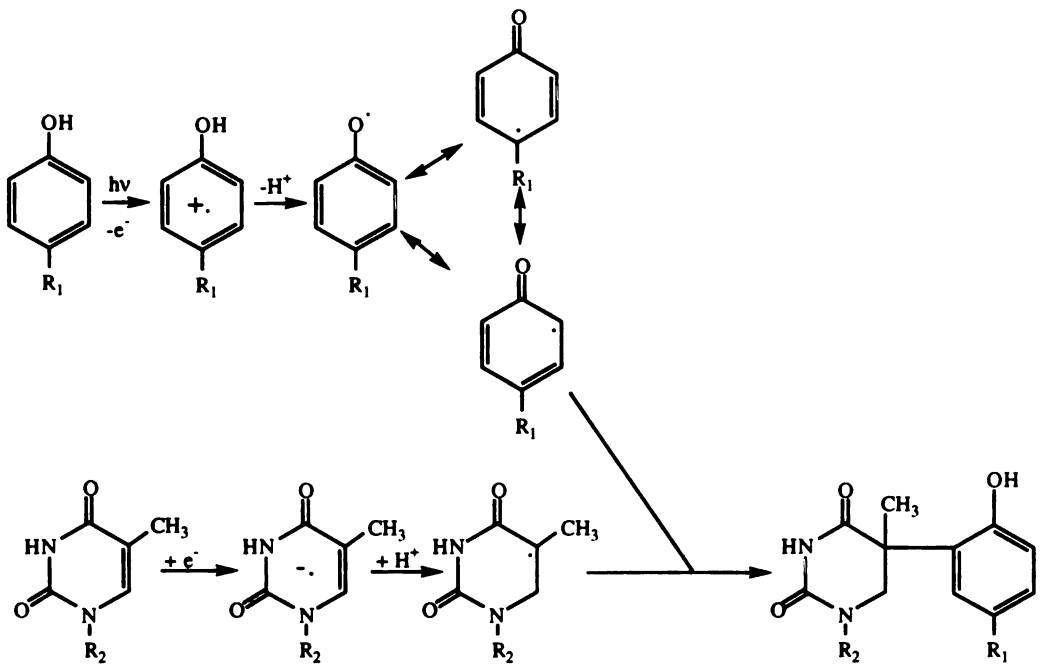


Figure 1.6 Proposed mechanism of the formation of tyrosine-thymidine crosslinks

Another amino acid thought to play a critical role in protein-nucleic acid binding is lysine. Shetlar and co-workers first noted the formation of a thymidine-lysine cross-adduct (after acid hydrolysis) from the UV-irradiation of DNA and L-lysine[70]. Subsequently, Saito and co-workers investigated photochemical reactions of alkylamines[71, 72] and L-lysine[73] with thymidine. The photoreaction appears to

involve a ring-opening reaction of the thymidine residue *via* a nucleophilic attack of the C-2 carbon by the ε-nitrogen from the alkylamine followed (upon heating or acidification) by incorporation of the nitrogen into the N1-position of the thymine, resulting in a displacement of the ribose[73] (Figure 1.7). Shetlar and co-workers reported similar results with photochemical reactions of alkylamines with cytosine and 5-methylcytosine[74].

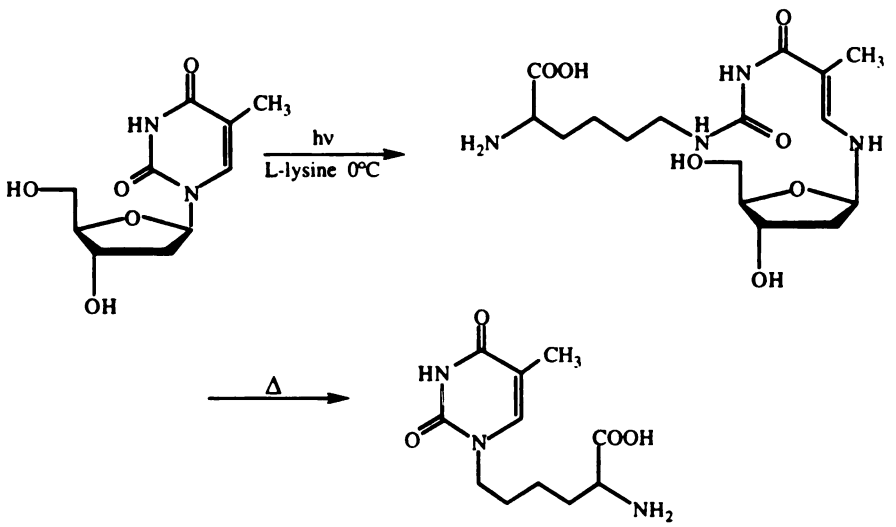


Figure 1.7 Photoreaction of thymidine with lysine

Another kind of crosslinking technique that could be useful is the irradiation of peptides (or proteins) in the presence of 5-bromo-2'-deoxyuridine. The mechanism for formation of a covalent bond with tyrosine was initially thought to be quite different from that of thymidine photochemical reaction[75, 76]. Upon selective irradiation (around 310 ± 10nm) there is an n to π* transition, followed by an abstraction of an electron. The radical anion is then protonated and the bromine is subsequently lost. Figure 1.8 shows a plausible mechanism for the formation of a uracil-N-acetyltyrosine ethyl amide conjugate *via* photoexcitation of a halogenated uracil. While this mechanism may still hold true for iodouracils, a recent report reveals that the mechanism for formation of covalent bonds

between bromouracil and tyrosine[77] might be very much like that proposed for thymidine and tyrosine[69].

There are many examples in which 5-bromouridine-substituted oligonucleotides were used in crosslinking experiments, including crosslinking telomeric proteins[67], the rat uterine estrogen receptor[78], the lac repressor-DNA complex[79, 80], the Myc DNA complex[81], EcoP15I DNA methyltransferase[82], adenosine cyclic 3',5'-phosphate receptor protein[83], the bZIP DNA-binding domain in GCN4[84], bacteriophage R17 coat protein[85], archaebacterial chromosomal protein MC1[86], the B-block region of adenovirus VA1 DNA complexed with human RNA polymerase III transcription factor (TFIII) C2[87] and as a method for detecting protein-DNA interactions at sites of chromatin replication[88].

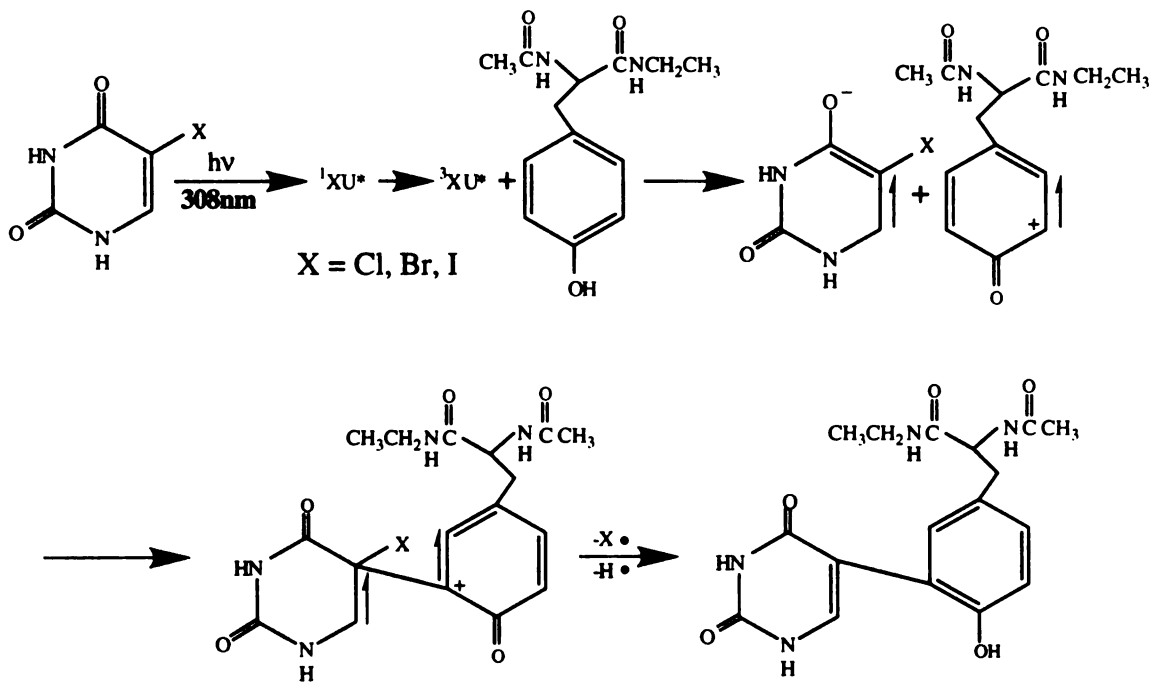


Figure 1.8 Proposed mechanism for the formation of 5-halouracil-tyrosine conjugates. The vertical arrows indicate the spin of the electron

In theory, there would be several advantages for substituting halogenated uridines for thymidines. Probably the biggest advantage is the lower energy (longer wavelength)

UV light needed to induce photochemical crosslinking which can be preferentially absorbed by the halogenated pyrimidine. There is less of a chance of photodestruction of the aromatic amino acids and the oligonucleotide. Another advantage is that the van der Waals volume swept out by a bromine atom approximates that of a methyl group (such as would be found on thymine), so it would be reasonable to assume that the modes of binding to a protein between a halogenated uracil incorporated into an oligonucleotide would be the similar to that for thymidine. One disadvantage is that 5-bromo-substituted DNA does not occur naturally in living systems, and so any results obtained from crosslinking could be subject to the criticism that this system is not an exact analog of a natural system. In addition, the photochemistry of halogenated uracils has not been thoroughly investigated, though they have been shown also to be crosslinkable by UV-light to cysteine derivatives[89] and alkylamines[90].

1.4 MASS SPECTROMETRY

Over the last fifteen years, mass spectrometry has evolved into a highly accurate, sensitive and versatile tool in the fields of analytical biochemistry and molecular biology. In principle, a mass spectrometer is composed of two semi-independent parts: an ionization source and a mass analyzer. Both are discussed in greater detail below. Before 1980, mass spectrometry was mostly used for analyzing relatively volatile and small molecular-weight compounds. One of the major challenges that faced mass spectrometry was that the samples had to be introduced into the mass spectrometer in the gas phase. One can normally heat volatile compounds until their vapor pressure is approximately 10^{-5} torr. An obvious problem arises from the fact that volatility is related to the sample's polarity and molecular weight. Biological macromolecules are so large, polar and labile that it is impossible to get them intact in the gas phase by heating.

1.4.1 THE SOURCE

1.4.1.1 Fast-atom bombardment and liquid secondary ionization mass spectrometry

The invention of secondary ionization mass spectrometric techniques, like FAB-MS (fast-atom bombardment mass spectrometry)[91] and its sister ionization technique, LSIMS (liquid secondary-ionization mass spectrometry)[92] in the early 1980's brought forth a new era for mass spectrometry. As opposed to imparting energy directly onto the sample, these techniques use a primary atom (or ion) current directed at a semivolatile liquid matrix (Figure 1.9). The matrix transmits the energy received from the source which consists of neutral atoms, such as argon or xenon (FAB-MS), or ionized heavy atoms, such as cesium (LSIMS), of very high kinetic energy (on the order of 10 keV). Near the point where the beam strikes the sample the kinetic energy is converted to thermal energy, which "sputters" out the analyte, as well as clusters of glycerol molecules. An advantage of this method is that FAB-MS and LSIMS are *soft*-ionization techniques, which means the energy imparted to the analyte is just sufficient to get it into the gaseous state, so that the predominant peak in the mass spectrum is due to intact analyte molecular ions.

While FAB-MS and LSIMS have been extremely useful for the analysis of peptides, the sensitivity of these techniques falls off rapidly at masses greater than a few thousand. But during the 1980's two other ionization techniques were developed which greatly extended the practical mass range and utility of mass spectrometry, so much so that molecular weights of up to hundreds of thousands of Daltons could be measured directly. These two methods, electrospray ionization (ESI) first developed in the late 1960's[93] and later refined when coupled to a liquid chromatograph and a quadrupole mass analyzer[94, 95], and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry[96, 97] have enjoyed recent success in biological mass spectrometry.

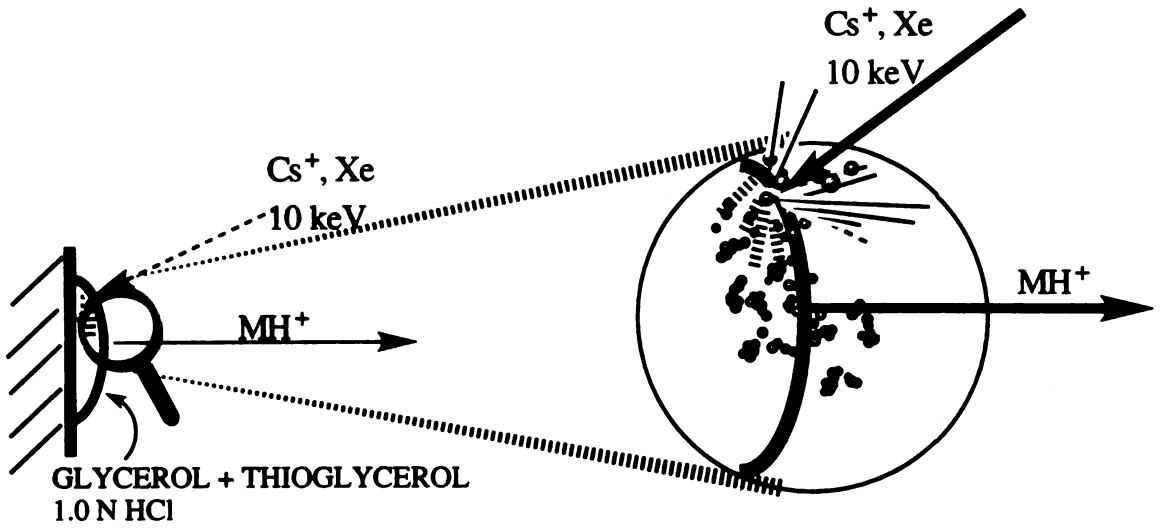


Figure 1.9 Proposed mechanism for SIMS desorption/ionization

1.4.1.2 Electrospray ionization mass spectrometry

The method of ionization for ESI-MS utilizes a simple capillary encased in an electrode (Figure 1.10). The analyte receives its charge from a metal tube held at 2-4kV. The sample, in a volatile liquid eluant (like that from an HPLC) is nebulized as it leaves the electrode. The microscopic particles are then subjected to a potential of 4-6kV, which accelerates the charged species. The stream of charged droplets passes through a drying gas, which speeds evaporation of most of the solvent molecules. As the particles get smaller and smaller, the charge density increases until whan have been described as "coulombic explosions" cause it to break up further, until only a single analyte molecule (with a number of charges) is left. One usually observes a bell-shaped series of peaks arising from a single compound, as there are usually a number of protonation sites on a biomolecule. Each peak represents the average mass (plus a number of charges) divided by that charge.

A typical spectrum of a recombinant protein reveals this distribution of peaks (Fig. 1.11). In order to determine the average mass of a molecule, one needs only to use

two *successive* peaks. There are two unknowns: M_r, the average mass, and n, the charge state of the respective peaks. By a simple algebraic application:

$$\text{observed mass peak 1} = (M_r + n) / n$$

$$\text{observed mass peak 2} = (M_r + n - 1) / (n - 1),$$

$$(\text{obs mass of peak 2} > \text{obs mass peak 1})$$

one can, in principle, calculate the mass of the singly protonated species. If there are more than two peaks, then one could expand the above equation, and calculate the average mass even more accurately. The mass-to-charge (m/z) range typically lies in the region from m/z 200 to around m/z 2500 regardless of the size of the macromolecule being analyzed (but there are exceptions). This mass range is ideally suited for quadrupole mass analyzers. Even though a quadrupole analyzer itself has only modest resolving power, the series of peaks allows the mass accuracy of an ESI-MS to be determined to 0.01 percent, or an error of one mass unit per 10,000.

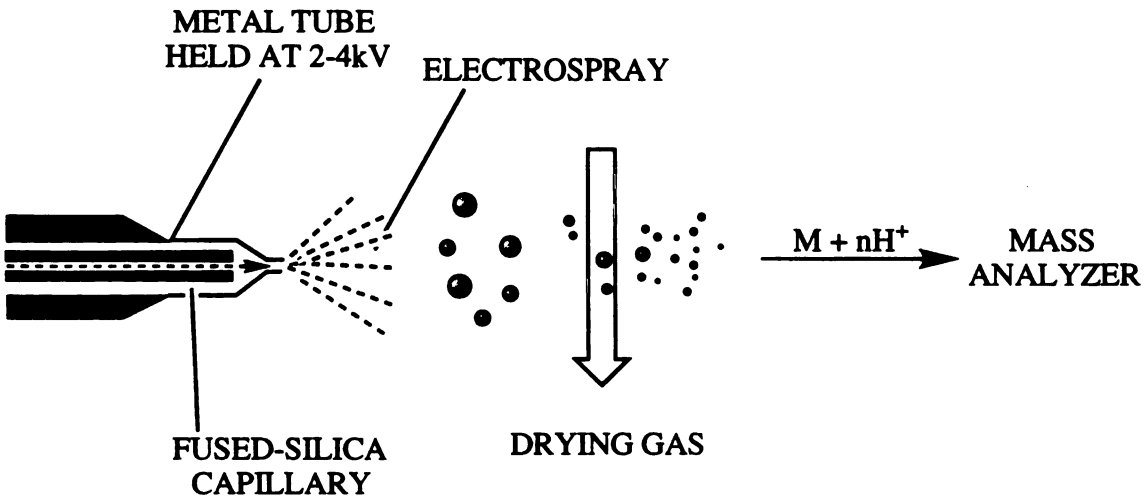


Figure 1.10 Schematic for ESI mass spectrometry

Because this method of ionization can be connected on-line to an HPLC and a quadrupole mass analyzer, both relatively simple instruments, it is little wonder that ESI-MS is a popular method for quickly determining accurate molecular weights of proteins, peptides and oligonucleotides. The practical upper mass limit of ESI-MS is around 100,000.

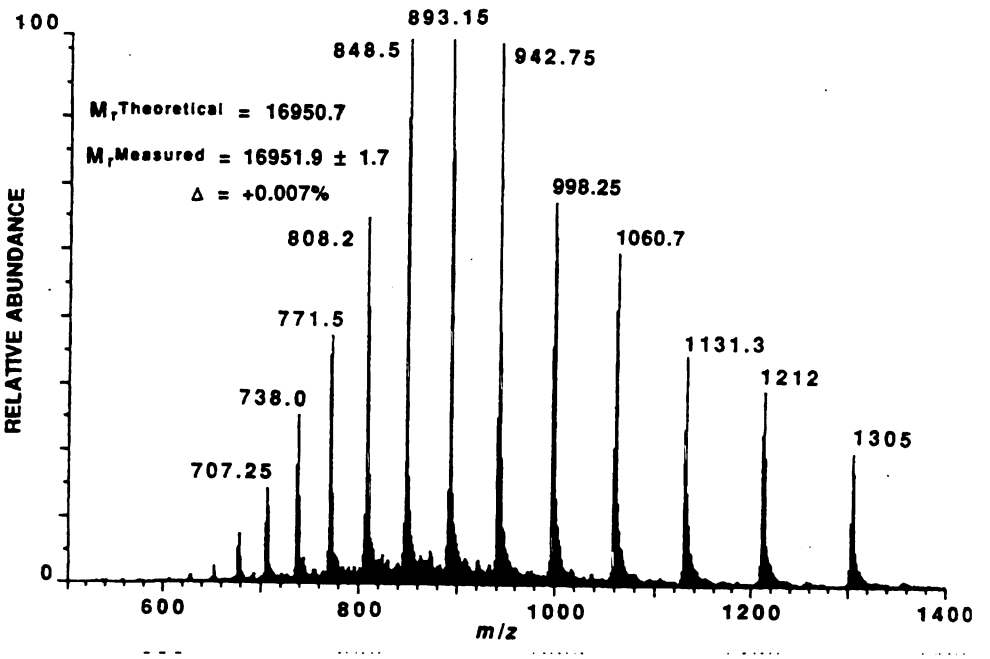


Figure 1.11 ESI of recombinant hemoglobin (from reference [98])

1.4.1.3 Matrix-assisted laser desorption-ionization mass spectrometry

Another mass spectrometric method that has gained widespread acceptance is MALDI-MS. In general, the sample is dissolved in a volatile solution which contains a UV-absorbing matrix. Small aliquots are removed and allowed to dry and co-crystallize with the matrix on a metal target. The target is then inserted into the mass spectrometer. The sample is then irradiated with a simple pulsed nitrogen laser emitting principally at 337nm. While some details of the ionization process still remain unclear, the matrix

appears to absorb the energy from the laser beam, which gets converted to thermal energy, volatilizes, and carries with it the embedded sample being analyzed. It is thought that the sample is ionized by proton transfer to or from the ionized matrix.

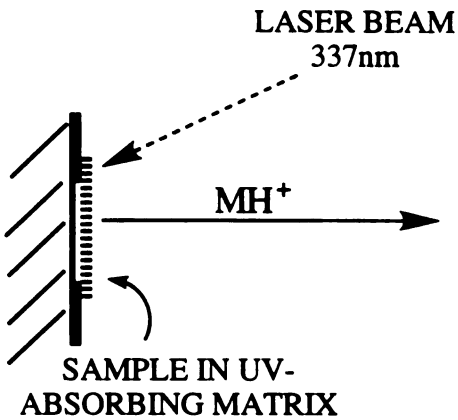


Figure 1.12 Proposed mechanism for MALDI desorption/ionization

The mass range of MALDI-time of flight (MALDI-TOF) is impressive: there are reports that immunoglobulins of molecular weights nearing 1,000,000 can be analyzed directly [99]. Using internal standards, a well-calibrated MALDI-TOF mass spectrometer can have a mass accuracy often exceeding a few ppm for peptides, and the instrument can detect compounds in the low femtomole (10^{-15} mole) to attomole (10^{-18} mole) range.

At this point it is important to briefly discuss the nomenclature used when describing the masses observed. For relatively low masses ($M_r < 3500$) observed on an *e.g.* a double-focusing magnetic sector for modern MALDI-TOF mass spectrometer, where individual isotope peaks can be resolved, one can describe the peak by its *monoisotopic* mass. The monoisotopic mass can be calculated from the summation of all the atoms in the sample based on their most common isotope, where ^{12}C is defined having an atomic mass of 12.0000, ^{1}H has a mass of 1.0078, ^{16}O has a mass of 15.9949, and so on. The additional peaks usually observed ($M + H^+ + 1$, $M + H^+ + 2$, etc. from a positive ion spectrum) arise from an isotopic substitution of one (or more) of the atoms

(^{13}C for ^{12}C, or ^{15}N for ^{14}N, etc). This is the reason why one observes the characteristic clustering of peaks using high-resolution mass spectometry. As the analytes become larger, the probability that there will be at least one isotopic substitution increases. In addition, the probability that there will be species that contain only atoms with their most common isotope decreases, and the monoisotopic peak becomes difficult to identify at higher masses. Another problem arises in that the resolution of the mass spectrometer might not allow for resolution of individual isotopic species (except in fourier-transform ion cyclotron mass spectrometry, or FTICR). At this mass range (usually around M_r = 5000-6000 and above) it becomes more appropriate to describe the *average* mass of the analyte. The average mass can be calculated by adding up the masses of the atoms, where each atomic mass is the weighted sum of all the isotopes for that particular element.

For LSIMS, tandem MS and MALDI-reflectron MS (the latter two will be described below), the molecular weights reported will generally be monoisotopic masses, unless otherwise noted. The ESI experiments molecular weights should be considered as average masses. The advantages of MALDI-TOF include sensitivity and mass ranges far greater than that of magnetic sector instruments, as well as having a lower cost to purchase and maintain, and ease of operation.

1.4.2 THE MASS ANALYZER

The next stage of the mass spectrometer is the mass analyzer. There are at least five types of mass analyzers in common use: (1) single- and double-focusing magnetic deflection, (2) time-of-flight, (3) quadrupole, (4) ion trap and (5) ion cyclotron resonance mass spectrometry. The first three types of analyzers, which were used in this dissertation, will be discussed below.

1.4.2.1 Magnetic deflection

In a magnetic deflection instrument, charged species are separated in the magnetic sector on the basis of their momentum. If an electric sector is coupled with the spectrometer (which acts as an ion kinetic energy focusing device), it is designated as a double-focusing instrument (Figure 1.13). By double-focusing, one means that the analyzer possesses both direction and velocity focusing.

A tandem double-focusing mass spectrometer can be visualized as two double-focusing mass spectrometers separated by a collision cell which, when filled with an inert gas, permits collisions of the ions with the neutral gas molecules, thereby increasing the internal energy and enhancing fragmentation of the ions. By selecting a narrow mass window with the first MS, one can allow a particular mass through (such as a monoisotopic peak of a peptide), fragment the ion, and record the fragment ion spectrum with the second MS. This technique is called collision-induced dissociation, or CID.

At UCSF, the Mass Spectrometry Lab has a Kratos Concept IIHH with a design very similar to the schematic shown in Figure 1.14. The detector used is a diode array, which allows for simultaneous monitoring of 4% of the mass range[100].

The nomenclature used to describe the fragmentation patterns for peptides is found in Figure 1.15. In general, x, y and z notation is used to describe fragmentation when the charged is retained on the C-terminal fragment; a, b and c notation is used if the charge is maintained on the N-terminal fragment[101].

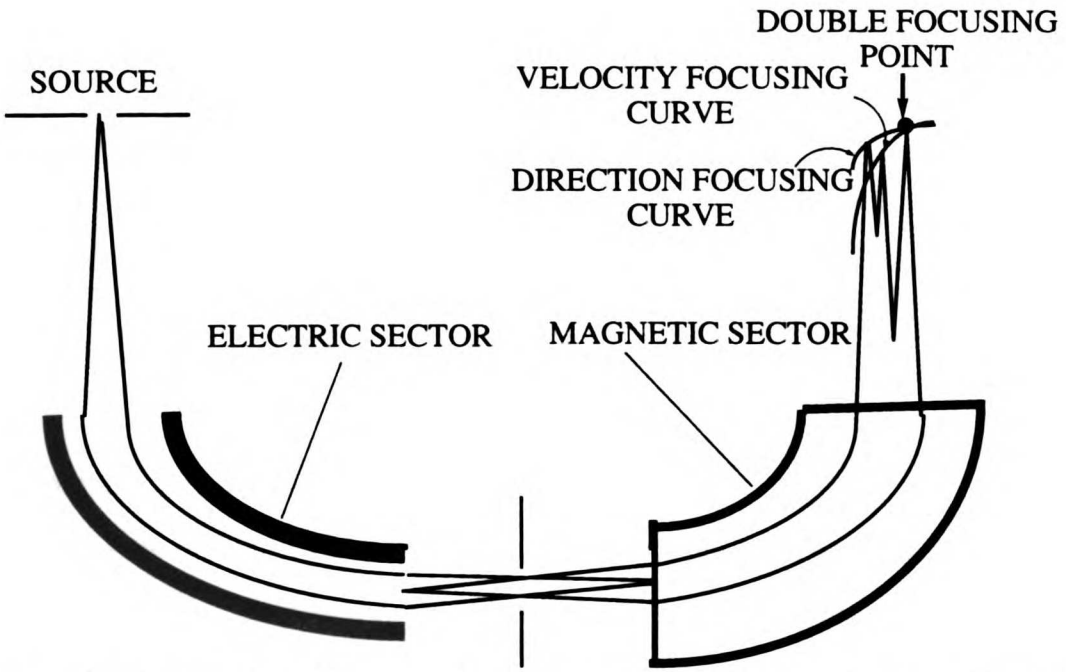


Figure 1.13 Double-focusing magnetic sector mass spectrometer (Redrawn from Ref [102]).

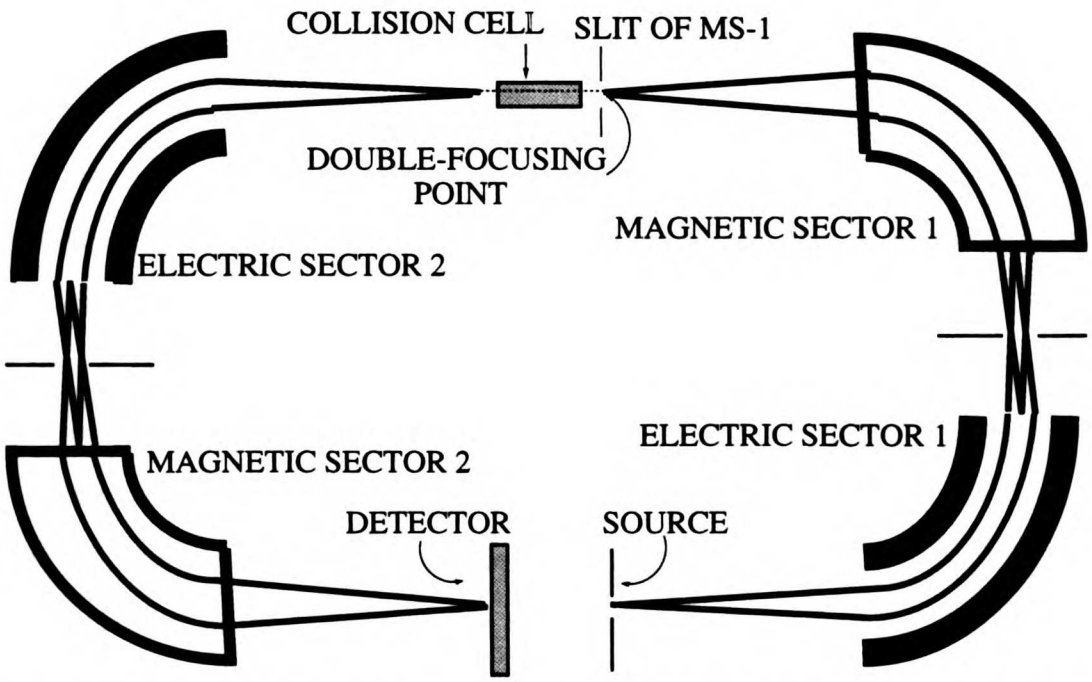


Figure 1.14 Tandem double-focusing magnetic sector instruments. (Redrawn from Ref [102]).

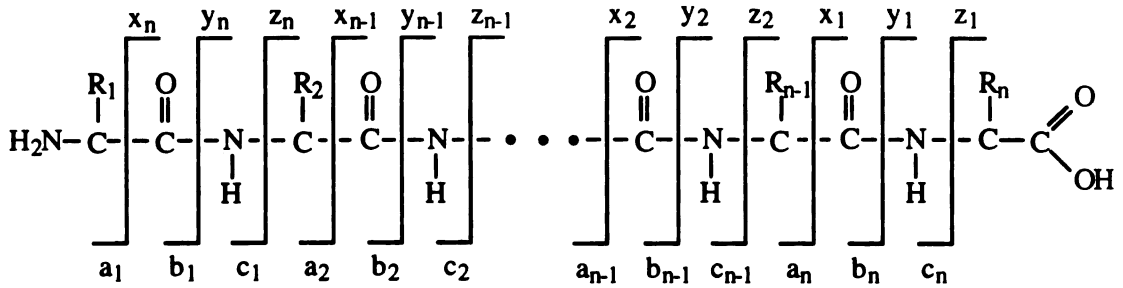


Figure 1.15 CID notation for peptide backbone fragmentation[101]

1.4.2.2 Time-of-flight

In a time-of-flight (TOF) instrument, masses are analyzed on the basis of the time the ions take to reach the detector. Provided all ions are formed in the same plane, and all are given the same amount of energy, an ion of mass m_1 will acquire a kinetic energy of $^1/_2 m_1 v_1^2$. Since all singly-charged species receive the same amount of kinetic energy, (*i.e.* $^1/_2 m_1 v_1^2 = {}^1/_2 m_2 v_2^2$, and so on) the corresponding velocities (v_n) will be proportional to the square roots of their masses. Since all the ions traverse the same distance from the target to the detector, the time it takes for the ion to travel that distance will be dependent on the square root of its mass.

In order to reduce the effect of the kinetic energy spread of the ions (which by the above equations leads to a loss of mass resolution) a second feature has been incorporated into some MALDI-TOF instruments. A reflectron, which is set at a higher potential than the source, retards and reflects the ions. Charged species with excess kinetic energy will penetrate further into this electrostatic field and thus follow a longer path to the detector. Those ions with lower kinetic energy will not penetrate as far, thus following a shorter path. The net result is an electrostatic focusing of all the ions of the same mass. A separate feature of the reflector is that it can be used for post-source decay, or PSD, a method which separates out charged species that have undergone fragmentation in the field-free region of the flight tube. In order to improve the resolution of a PSD spectrum,

the reflectron electrostatic potential is lowered stepwise which allows a narrow (and progressively lower) mass range to penetrate the field. The resultant spectra are summed to give a final spectrum covering the mass range of interest.

1.4.2.3 Quadrupole

A quadrupole mass analyzer is also commonly found on mass spectrometers. The ions are passed in a space between four parallel rods, each with a circular (or hyperbolic) cross section. Opposing pairs of rods are given a dc voltage and an rf voltage. The quadrupole acts as an m/z filter, because as only ions having the appropriate m/z values will follow a stable enough path in between the rods. By varying both the dc and rf voltage (such that the ratio of the two remains constant) ions of different m/z values can pass through.

A quadrupole can be turned into a tandem instrument by attaching another quadrupole. The two are separated by a third quadrupole (hence the name "triple-quad" or Q1qQ2), but the difference is that the middle analyzer (q) has no ion selectivity, due to the fact that it is rf-only. In this second stage, precursor ions (selected by Q1) can undergo collision-induced fragmentation (CID) or photodissociation, and then be analyzed by the third quadrupole (Q2).

1.4.3 MASS SPECTROMETRY OF PROTEINS AND NUCLEIC ACIDS

Mass spectrometry is obviously a very versatile method for measuring molecular weights of biomolecules. Both ESI-MS and MALDI-TOF MS can readily provide accurate masses for many biological compounds at the picomole level. Each technique has different advantages. As mentioned previously, ESI-MS can be connected on-line to a microbore HPLC, and thus provide simultaneous detection and measurement of molecular weights of peptides, proteins and oligonucleotides as they elute. MALDI-TOF MS enjoys ease of use, and the ability to determine molecular weights of dozens, and even hundreds of samples in a short period of time.

There are a number of matrices available for MALDI-TOF MS depending on the types of biological compounds and molecular weights. In particular, 2,5-dihydroxybenzoic acid[103, 104] and α-cyano-4-hydroxy-cinnamic acid[105] have been the matrices of choice when analyzing smaller molecular weight compounds, such as peptides from a proteolytic digest, while 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid)[99] appears well-suited for large molecular-weight compounds, such as proteins. For oligonucleotides, 3-hydroxypicolinic acid (HPA)[106] and 2,4,6-trihydroxyacetophenone (THAP)[107] appear to be the best matrices. The latter two could be of particular interest because, in general, they are prepared with the addition of a co-matrix consisting of either diammonium hydrogen citrate or diammonium hydrogen tartrate. It is thought that the organic buffer acts as an *in situ* cation exchanger--replacing ammonium ions for alkali (Na^+, K^+) cations, which form adducts with oligonucleotides, thus complicating the mass spectrum and reducing the signal strength of the monoprotonated (MH^+) species. In addition, the matrix works moderately well for proteins, and could conceivably be an excellent candidate to analyze compounds possessing both protein and oligonucleotide characteristics.

1.4.3.1 Peptide Sequencing by Mass Spectrometry

There are a number of methods by which one can sequence peptides using mass spectrometry. Tandem double-focusing mass spectrometry, as mentioned previously, uses high-energy collisionally-induced activation to fragment peptides along the backbone, giving characteristic C-terminal and N-terminal fragments. One can in theory deduce the sequence of the peptide from the fragment masses[101]. A second way of determining peptide sequences arises from the fact that peptides can undergo fragmentation while in the flight tube in a MALDI-TOF instrument. Using a MALDI-TOF in reflectron mode, the fragments can then be analyzed by post-source decay, as mentioned previously.[108] A third method for sequencing peptides uses FAB-MS[109, 110] and more recently, MALDI-TOF MS[111, 112] in which peptides are partially

digested with either carboxypeptidase P or carboxypeptidase Y, which creates a ragged end on the C-terminus of the peptide. In the case of MALDI-MS, the enzymatic reaction can take place right on the sample well.

1.5 NANOSECOND-PULSED UV LASER CROSSLINKING OF PROTEINS AND NUCLEIC ACIDS

One primary interest in this thesis is in the use of nanosecond-pulsed lasers as the source of UV light for crosslinking[113-118]. Our laser crosslinking studies were done using the system in the laboratory of Professor Peter von Hippel at the University of Oregon. A neodymium-yttrium-aluminum-garnet (Nd:YAG) laser (model DCR-3G, Spectra-Physics, Mountain View, CA) was employed[116]. Using a nanosecond-pulsed laser as a UV source overcomes a major limitation of more conventional UV sources: it delivers power extremely rapidly. A five nanosecond "pulse" is equivalent to ten minutes on a conventional germicidal lamp.

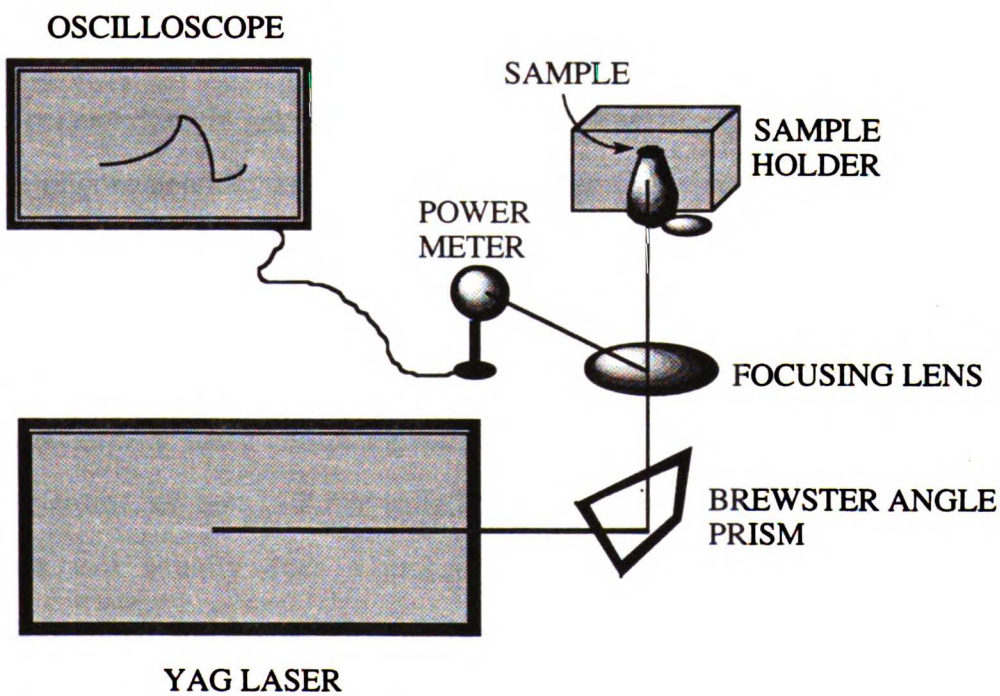


Figure 1.16 Schematic of the Nd:YAG laser

The Nd:YAG laser emits principally at 1064nm, and is frequency-quadrupled (to 266nm) by passing the beam through a Brewster-angle prism. The beam is then passed through a focusing lens, so that it converges at a point just above the sample being irradiated. The beam then diverges from that point and strikes the full surface area of the sample. Normally the sample (in approximately 10μL solution) is held in place (by surface tension) at the bottom of a 1.7mL eppendorf tube (Figure 1.16).

1.6 UV-CROSSLINKING OF PROTEINS AND NUCLEIC ACIDS: PROGRESS TO DATE

Traditional methods for determining contact points in protein-nucleic acid complexes have been described by Williams and Koningsberg[119]. In general, one designs an *in vitro* system that consists of the protein of interest and a ^{32}P-labeled nucleic acid (or oligonucleotide), and incubates the system at room temperature for a short while. Subsequently, the system is then irradiated with UV light emanating from a low-pressure mercury vapor germicidal lamp emitting principally at 254nm or 313nm, depending on the nature of the experiment. The reaction is usually monitored by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). After covalent conjugation with the oligonucleotide, one would expect a second band which migrates more slowly through the gel. These bands can be stained and examined visually or by autoradiography to determine if, and to what extent, crosslinking has occurred.

After determining the optimal dose of irradiation, one usually employs some form of precipitation or some chromatographic procedure to remove unreacted protein (or oligonucleotide, or both). After isolation of the crosslinked protein-oligonucleotide complex, one usually does a proteolytic digest, followed by anion-exchange chromatography. In general, the unmodified peptides will elute fairly early in an anion-exchange run, while peptides that are covalently bound to the oligonucleotide will elute much later, owning to the number of negative charges from the phosphate backbone of

oligonucleotides. Using trypsin as the protease, which is specific for arginine and lysine, insures that there at least one positive charge on the peptide would negate one of the phosphate charges on the oligonucleotide, and hence, the crosslinked peptide-oligonucleotide complex should have a slightly shorter retention time than that of the unreacted oligonucleotide. The eluant can be monitored at 260nm, or in the case when ^{32}P-labeled oligos are used, by scintillation counting of fractions.

The candidate peptide-oligonucleotide complexes can be desalted using a variety of methods, such as reversed-phase HPLC, a Sep-Pak (Waters) or a Nensorb 20 cartridge (DuPont). The samples can then be sequenced using automated Edman degradation. In general, one would look for characteristic "drops" in yields of amino acids as compared to the unmodified peptide, which should indicate that this amino acid has somehow been covalently modified.

There have been a number of complexes that have been analyzed by the methods (or some variation) described below. Some examples are shown in Table 1.1.

Protein-nucleic acid complex	Reference
dNTP binding site in pol β	[120]
Uracil-DNA glycosylase	[121]
Klenow fragment of *E. coli* DNA pol I in its polymerase mode	[122]
Max:DNA complex	[123]
Human immunodeficiency virus type 1 integrase	[124]
E. coli Dam methyltransferase	[125]
ss-DNA binding region of pol β	[126]
Adenovirus DNA binding protein	[127]
Bacteriophage T4 regA protein-nucleic acid complex	[128]
TFIIIα and 5S DNA	[129]
HIV-RT	[130]
Klenow Fragment with azido ATP	[131]
Bacteriophage T4 gene 32	[132]
Histone H3	[133, 134]
Histones 2A and 2B in chromatin	[135]
E. coli SSB to DNA	[136]
DNA-histones	[137, 138]
fd gene 5 protein to fd DNA	[139-141]
Protein L4 and 23S RNA	[142]

Table 1.1 Protein-nucleic acid complexes analyzed by UV-crosslinking

1.7 SPECIFIC AIMS

Our aims in the work described here were to develop and utilize novel protocols to explore the protein binding sites of protein-nucleic acid complexes at the peptide and amino acid level by using low-intensity and nanosecond-pulsed laser UV light to induce covalent bonds between specific amino acids and nucleic acid bases, and to use advanced mass spectrometric techniques to explore the nature of the crosslink. The protocols we

developed improved upon and were used in conjunction with more traditional methods. We were able to use a variety of chemical and enzymatic nucleases to digest away the irradiated oligonucleotide, which allowed for less ambiguous results. We also circumvented some of the problems associated with irradiated oligonucleotides by judicious selection of oligonucleotide sequences. In Table 1.2, we describe how mass spectrometry was used in addition to the well-established methods. As there has been scant evidence in the literature regarding the use of nanosecond-pulsed laser UV crosslinking to analyze contact regions of protein-nucleic acid complexes at the amino acid/peptide level, we used our existing protocols to compare and contrast the results we obtain with both low-intensity and high-intensity UV crosslinking.

Method	Traditional	New
UV crosslinking	Low intensity UV light	Nanosecond-pulsed UV light
Protein-oligonucleotide complex detection	SDS-PAGE	MALDI-TOF MS
Peptide-oligonucleotide complex detection	HPLC/Scintillation Counting	HPLC/MALDI-TOF MS
Sequencing of Crosslinked peptides	Edman Sequencing	MALDI-PSD/Tandem-MS/MALDI/ Carboxypeptidase Y digestion
Mass spectrometric sequencing of peptide-nucleotide complexes	Edman Sequencing	MALDI-PSD/Tandem MS/MALDI/ Carboxypeptidase Y digestion

Table 1.2 New protocol for analyzing UV-crosslinked protein-nucleic acid complexes

Mass spectrometry can clearly play a significant role in these protocols. Some of the advantages of utilizing mass spectrometry can be summarized as follows:

- *Mass spectrometry is very sensitive.* Normally only pmol amounts of proteins and oligonucleotides are needed to determine accurate masses. This is doubly advantageous as it eliminates the need for radiolabeled oligonucleotides.
- *Mass spectrometry allows for much less ambiguous interpretations of the sequencing data.* Indeed, we show that modified amino acid residues can be determined directly, rather than implicitly, as what is done by Edman sequencing.
- *The protocols established should be applicable to many other protein-nucleic acid systems.* One does not need to know anything about the structure of the protein *a priori.* All that is needed is the primary sequence of the protein, and experimental evidence (gel retardation, fluorescence quenching, footprinting, etc.) that, in fact, binding does occur.

A recent use of photochemical crosslinking and mass spectrometry to deduce the contact points of protein-nucleic acid complexes was carried out by a collaborative effort of the labs of Douglas Barofsky and Dale Mosbaugh from Oregon State University and Professor Peter von Hippel at the University of Oregon. Two recent papers describe the results of their mass spectrometric studies of photochemically crosslinked proteins and nucleic acids. The first paper describes the use of MALDI-TOF MS to analyze the NPUV crosslink formed between T4 phage gene 32 protein and oligo $(dT)_{20}$[143]. The second paper used MALDI-TOF MS to determine (1) if, and to what extent, the protein was crosslinked to oligo $(dT)_{20}$, and (2) to locate peptide-oligonucleotide complexes stemming from a proteolytic digest of uracil-DNA glycosylase complexed with $(dT)_{20}$[121]. Both papers suggest the power and potential of mass spectrometry in protocols such as these.

In order to evaluate the potential for using LIUV- and NPUV-crosslinking in conjunction with mass spectrometry, we designed and characterized several systems which would "model" a protein-nucleic acid interaction. Chapter 2 is devoted to the

results obtained from irradiating peptide/nucleoside and peptide/nucleotide systems. Chapter 3 describes and compares the results from LIUV- and NPUV-crosslinking of the single-stranded DNA binding subunit of rat DNA polymerase β and oligonucleotide d(ATATATA).

1.8 EXPERIMENTAL CONSIDERATIONS

I now describe some of the experimental and analytical difficulties that one faces when trying to determine the nature of nucleic acid-amino acid (and ultimately, nucleic acid, protein) contact points using UV crosslinking and mass spectrometry. Perhaps most daunting of all is the extraordinary number of products that can be formed between reacting species. This is especially true if one uses a low-pressure Hg germicidal lamp (equipped with a Vycor shield) at the UV light source, which emits principally at 253.7nm. This wavelength closely corresponds to a number of overlapping transitions for nucleoside bases which normally occur around 258-265nm[144].

When irradiating nucleosides with UV light, a number of different products can be formed, both intramolecular and intermolecular; those products, in turn, can undergo secondary photochemical reactions. As one introduces other compounds into the system, such as amino acids, the problem (in principle) only becomes more challenging. Irradiating oligonucleotides in the presence of proteins can be prohibitively complex.

Another problem is that of potential chemical lability in the crosslinks formed. Are they stable under the conditions used in chromatographic workup procedures? For instance, a very popular method of isolating and desalting photoproducts is that of reversed-phase high performance liquid chromatography, (RP-HPLC). Normally, the aqueous solvent contains a small amount (0.1 percent by volume, or about 13mM) trifluoroacetic acid, which would be essentially completely dissociated in an aqueous solution (pKa ~ 0), hence giving an estimated pH around 2.0. Compounds sensitive to acid hydrolysis could be broken down during the course of the HPLC run. One way

around this problem is to use a system consisting of volatile and neutral ion-pairing buffers, such as triethylammonium acetate, or triethylammonium bicarbonate. In addition, while many photoreactions are carried out at 4-5°C, most HPLC and mass spectrometric instruments run at room temperature (20-23°C). Any compounds sensitive to heat might decompose.

Another item to consider is the nature of the UV light. In many cases, the length of the photochemical irradiation using a conventional UV light source (such as a germicidal lamp) can vary from a few minutes to a few hours, depending of the amounts irradiated, the wavelength of light used, etc. We have to take into consideration that the on/off rates for proteins and nucleic acids is on the order of nanoseconds to microseconds. In other words, during the course of the reaction, the proteins have the opportunity of undergoing innumerable different conformational shifts, as well as jumping on and off the nucleic acid, over an extended time period. If we impose a covalent restraint, the system might become perturbed such that a new ensemble of complexes are formed, which might perturb amino acid/nucleic acid contacts such that amino acids once found in the binding pocket might now be too far away to form a covalent bond with a nucleic acid residue (a false negative); conversely, amino acids that were not proximal to the oligonucleotide before might now be close enough to react (a false positive). Also, since it is the nucleic acid which absorbs almost all the incident photons, ensuing pyrimidine dimer formation or other photoproducts can cause significant distortions in the nucleic acid, possibly affecting the binding properties of the protein bound to it. One of the methods of crosslinking that could overcome these limitations is that of NPUV crosslinking, which was discussed previously.

One problem commonly enountered in these types of studies is in quantitatiing yields of protein-oligonucleotide, peptide-oligonucleotide and peptide-nucleotide crosslinks. The former can usually be quantitated by gel scanning and densitometry and estimating amounts of yields by comparing the amount of crosslinked material to that of a

known amount of unreacted protein. The latter two, however, are more challenging, as it is difficult to quantitate against a known "standard." Some possible methods to quantitate peptide-oligonucleotide crosslinks include HPLC and monitoring the eluant at 260nm The crosslink can be quantitated against a known amount of irradiated oligonucleotide. However, poly(dT), which is the oligonucleotide of choice is many crosslinking studies, loses a majority of its chromophores which absorb UV light in this region, and is notoriously difficult to chromatograph on reversed-phase and anion-exchange chromatography. A second possibility is to use ^{32}P labeled oligonucleotides, which greatly increases the signal-to-noise ratios of crosslinked conjugates. A third possibility is to use Edman sequencing of the peptide-oligonucleotide conjugates, and quantitate the PTH amino acid yield, which should be indicative of the amount of crosslinked material. In the case of peptide-nucleotide crosslinks, the best method would be to use HPLC and monitor at 215nm. To a first approximation, one would expect that the peak area of a peptide-nucleotide conjugate per unit amount would be the same as the sum of the two reacting species.

Another area where errors in interpretation of the results could arise is Edman sequencing. In some cases, there are several peptides that might be covalently bound to the oligonucleotide. If these peptide-oligonucleotide complexes cannot be resolved chromatographically, then the Edman sequencing results could become quite difficult to interpret. Also, by convention, a *negative* result (little or no yield of a PTH-amino acid), traditionally constitutes evidence of crosslinking of that amino acid. But there can be a number of reasons why there are low observed yields of a particular PTH-amino acid. If the amino acid is somehow modified either chemically or photochemically, the retention time would, of course, be different, and could give rise to the same "result." Of course, one could indirectly check to see if the candidate amino acids were involved in binding by doing *e.g.* site-directed mutagenesis and check for reduced or abolished binding, or

directly comfirm the proximity of the amino acid and the nucleic acid base in question by examining the X-ray or NMR data (if available).

This raises the questions, "just what constitute 'proximal' residues?" Just because amino acid-nucleotide contacts may be "proximal," are they essential for specific/non-specific DNA recognition? What are the advantages of using NPUV over LIUV crosslinking? Even under ideal photochemical reaction conditions, will all amino acids that lie on the interface be crosslinked to the nucleic acid? I hope that this thesis has provided at least partial answers. I also hope that the results obtained can be used as a foundation for other crosslinking experiments and structural studies involving proteins and nucleic acids.

CHAPTER TWO

MASS SPECTROMETRIC CHARACTERIZATION OF THE PEPTIDE SPSYSPT IRRADIATED IN THE PRESENCE OF THYMIDINE AND RELATED COMPOUNDS

2.1 INTRODUCTION

The goals of these studies are to establish and evaluate the protocols for the isolation and mass spectrometric analysis of nucleoside- and nucleotide-modified peptides. Our strategy is to design simple systems which could mimic a protein-nucleic acid interaction, induce specific covalent bonds using UV crosslinking, and evaluate the potential for chromatographic separation and mass spectrometric analysis. Photochemical crosslinking of protein-nucleic acid complexes is an attractive approach towards determining relevant binding sites[53-55]. The peptide-nucleotide system represents what could also be considered as an ideal end product of a protein-nucleic acid system which has been photochemically crosslinked, after this crosslinked complex has been completely digested with a judicious selection of proteases and nucleases.

The replacement of thymine with 5-bromouracil in crosslinking experiments has shown excellent promise with regards to locating amino acids involved in protein-nucleic acid interactions. Because the nature of the UV light used (>290nm as opposed to 254nm) there is less of a chance of forming unwanted secondary photoproducts, as well as an enhanced reactivity and photosensitivity to protein-DNA crosslinks[145]. There have been a number of reports in which bromouridine-substituted nucleic acids were used in crosslinking experiments[67, 78-88].

The photochemistry of tyrosine and thymidine, and tyrosine and 5-bromo-2'-deoxyuridine have already been investigated[69, 75, 76]. The proposed mechanisms for the formation of a tyrosine-thymidine and tyrosine-2'-deoxyuridine (the bromine atom is lost in the photochemical reaction) conjugates are described elsewhere[69, 75, 76]. There is ample evidence that tyrosine, with its aromatic ring and phenolic hydroxyl group, could be involved in a number of protein-nucleic acid interactions: for instance, crystal and solution structures of bacteriophage fd gene 5[146-149] and bacteriophage T4 gene 32[36, 150] reveal tyrosines in the nucleic acid binding pocket.

The model peptide used in these studies is a heptad repeat found in the C-terminal domain of RNA polymerase II. This heptapeptide, SPSYSPT, found in eukaryotic cells consisting of 17-52 tandem repeats (with minor variations of some of the amino acids in the heptad reported), is conserved across species[151, 152]. The primary sequence suggests that it has two overlapping SPXX motifs, which might give it unusual β-turns. Experiments attempting to deduce the secondary structure, as determined by several NMR and CD studies of tandem repeats of varying length, suggest that the peptide is largely disordered, but that it does, indeed, contain some β-turn structure[153-155]. Work done by Suzuki has shown that the 8-mer, YSPTSPSY, could bind to double stranded DNA[153]. Huang and co-workers have postulated that the polypeptide might interact with the DNA double helix by partial aromatic stacking of the tyrosine rings with the Watson-Crick base pairs[68]. More recently, Khait, *et al.* were able to show through a combination of NMR and molecular dynamics simulations that the peptide YSPTSPSY can adopt a number of stable conformations both in the presence and absence of a small double-stranded oligonucleotide[156], all of which might allow for the bisintercalation of the tyrosines into the DNA double helix.

A definitive descrption of the role of this repeating heptad, however, remains elusive: it was initially speculated that the polypeptide could, theoretically, stabilize the RNA polymerase transcription system by anchoring the tyrosines in the DNA double

helix[153], as well as bind to transcription factors[157]. It has also been shown to be extensively phoshorylated[158], and this phosphorylation is cell-cycle dependent[158]. Further evidence of other kind of interactions or roles is unclear, as it has also been shown that transcription *in vitro* can be accomplished even with this repeating heptad partially or totally removed[159, 160]. Yet its involvement in the transcription process appears to be so vital, at least in mammals, that all mutations in this region are silent[161].

In this report we describe the generation and mass spectral analysis of a number of peptide-nucleotide complexes by irradiating a tyrosine-containing peptide in the presence of thymidine, three thymidine analogs and a thymidine-containing dinucleotide. After chromatographic separation and purification, we used LSIMS(+) and high-energy tandem CID mass spectrometry to determine the location of the adduct. The use of mass spectrometry to analyze these types of complexes is a fairly novel concept, and there have only been a few reports of using mass spectrometry to measure masses of peptide-oligonucleotide crosslinks[121] as well as to obtain sequence information from a synthetic peptide-oligonucleotide conjugate[162]. Recently, Lipton and co-workers used triple-quadrupole electrospray mass spectrometry to deduce the site of modification of thymidine which had been irradiated in the presence of the bioactive peptide precursor Angiotensin I (sequence NRVYIHPFHL) using a ^{60}Co γ-ray source[163]. Our results go beyond this report and results obtained previously in our lab[69].

2.2 MATERIALS AND METHODS

General Considerations

Much of the work done in this report with thymidine-related compounds is an extension of the work done by Shaw, et al[69]. The brominated nucleotide work expands on some preliminary results by Dietz and Koch[75, 76]. Irradiations at 254nm were

carried out at 23°C using a 15W germicidal lamp equipped with a Vycor shield. Irradiations at $\lambda > 290$nm were carried out at 23°C using a Rayonet 3000 lamp equipped with a Pyrex shield. The peptide Ser-Pro-Ser-Tyr-Ser-Pro-Thr (SPSYSPT) was synthesized using Fmoc chemistry on a Rainin PS3 peptide synthesizer, with HBTU used as the coupling reagent. Protected amino acids were purchased from Novabiochem USA. Peptides were purified by RP-HPLC and the mass of the peptide was verified by LSIMS(+). Thymidine, thymidine-5'-monophosphate, thymidylyl(3'->5')-2'-deoxyadenosine, 5-bromo-2'-deoxyuridine and 5-bromo-2'-deoxyuridine-5'-monophosphate were all purchased from Sigma. Oxygen was removed by bubbling water-saturated nitrogen (99.997%) through the stoppered vessel. The vessels consisted of either a 2.0cm cylindrical quartz CD cell, or a 600µL capacity quartz rectangular cell (2.0mm width x 30mm height x 10mm depth), and sealed with a rubber septum. UV spectroscopic measurements were performed on a Hewlett-Packard HP8452A diode array spectrometer. All peptide and nucleotide solution concentrations were measured spectroscopically assuming the following extinction coefficients (Table 2.1) at pH = 8.0 and room temperature[164, 165].

Photoproducts were isolated and purified using RP-HPLC. The system consisted of two Rainin Rabbit Pumps controlled by a Dynamax (Version 1.2) system package. The eluant was monitored at either 215nm or 260nm using a Kratos 783 variable-wavelength detector equipped with a deuterium lamp. The HPLC column used was a Vydac C-18 analytical (4.6 x 250mm) column with a 300 Å pore size. The flow rate was 1.0mL/min. The solvent system consisted of degassed HPLC grade water (Fisher) with 0.10% (v/v) trifluoroacetic acid (TFA) (sequencing grade, Pierce, Solvent "A") and HPLC grade acetonitrile (Fisher) with 0.088% (v/v) TFA (Solvent "B"). All peptides, nucleosides and nucleotides were analyzed for purity (and repurified, if necessary) by

Species	λ (nm)	ε (M^{-1} cm^{-1})
Tyrosine (as SPSYSPT)	274	1,300
Thd, TMP	260	8,700
BrUd, BrUdMP	274	6,500
TpdA	260	23,400

Table 2.1. Extinction coefficients used for quantitation of tyrosine and thymidine derivatives

HPLC prior to use. The amounts of crosslinked material were estimated spectroscopically by assuming the 215nm absorbance per unit amount of crosslinked material was the same as the sum of the two original reacting species.

Peptide molecular weights were determined using LSIMS(+) on a Kratos Analytical (Manchester, UK) MS-50S double-focusing mass spectrometer equipped with a 23kG magnet, postacceleration detector, and fitted with a cesium ion source.[92] Samples for LSIMS(+) were run on a coolable introduction probe[166, 167] in a matrix of 1:1 glycerol/thioglycerol in 0.1M HCl. Tandem mass spectrometry was performed on a Kratos Concept IIHH four-sector EBEB mass spectrometer equipped with a cesium ion source and an electrooptical array detector[100]. In general, candidate crosslinked peptide-nucleotide complexes (approximately 0.5 to 1nmol) were split into two aliquots for molecular weight determination and subsequent sequencing by high-energy tandem CID. For CID analysis, lyophilized samples were taken up in 10μL of an aqueous solution containing TFA (0.1%, v/v), acetonitrile (5%, v/v) and thioglycerol (5%, v/v) and loaded onto the instrument using a flow probe[168].

Photoreaction of SPSYSPT with thymidine

A 500μL solution that was 500μM (500pmol/μL) in both SPSYSPT and thymidine in 5.0mM Na^+-phosphate buffer at pH = 8.0 was added to the 2.0cm

cylindrical quartz vessel and sealed with a rubber septum. The solution was degassed for 60 minutes with water-saturated N_2. The flat section of the cylindrical vessel was then placed flush (about 2.0cm away from the center) against a Vycor-shielded germicidal UV lamp emitting principally at 254nm. The reaction was monitored by taking out 5.0μL aliquots before and after irradiation, using the following RP-HPLC gradient: hold at 98% "A", 2% "B" for 7 minutes, ramp to 30% "B" in 28 minutes, for a total of 35 minutes per run. The reaction was stopped after 30 minutes, when it appeared that most (>90%) of the thymidine had been consumed.

Subsequent to irradiation, the sample was removed from the vessel and placed in a 1.7mL polypropylene eppendorf tube. The sample was spun at 10,000g for 10 minutes. The sample was then transferred to a clean polypropylene eppendorf tube and lyophilized to dryness, reconstituted in 50μL of 0.1% TFA, and chromatographed over HPLC employing the same conditions as before.

Photoreaction of SPSYSPT with thymidine-5'-monophosphate

A 500μL solution that was 500μM (500pmol/μL) in both SPSYSPT and thymidine-5'-monophosphate was prepared as mentioned previously. The UV irradiation was stopped after 10 minutes, when it appeared that most (>90%) of the thymidine-5'-monophosphate had been consumed. The mixture was worked up as mentioned previously.

Photoreaction of SPSYSPT with 5-bromo-2'-deoxyuridine

A different protocol was used for the irradiation of halogenated nucleoside and nucleotides. A 500μL solution that was 500μM in both SPSYSPT and 5-bromo-2'-deoxyuridine (BrdU) in 10mM Na^+-phosphate buffer (pH = 7.0) were added to a 600μL quartz UV cell (2mm width by 30mm height by 10mm depth) and sealed with a rubber septum. The system was deoxygenated for 60 minutes and irradiated at room temperature approximately 5cm from a Pyrex-shielded Hanovia lamp emitting principally at 313nm.

It was anticipated that the reaction mixture would not be quite as complex, as the nature of the UV light is of much longer wavelength (313nm), and should produce fewer secondary photoproducts. However, the lower amount of the UV light absorbed also meant much longer irradiation times. The reaction was monitored by removing and analyzing 10-μL aliquots at 0, 1, 2, 4, 18 and 40 hours, after which the reaction was terminated. Approximately 50% of the BrdU had reacted.

Photoreaction of SPSYSPT with 5-bromo-2'-deoxyuridine-5'-monophosphate

The same protocol was used for the irradiation with the 5-BrUd. Ten-μL aliquots were removed at 1, 2, 4, 18 and 33 hours, after which the reaction was terminated. Approximately 50% of the BrdUMP had reacted.

Photoreaction of SPSYSPT with thymidylyl (3'->5') -2'-deoxyadenosine

A 1.00mL solution containing 200μM SPSYSPT and 200μM thymidylyl-(3'->5')-2'-deoxyadenosine (TpdA) was prepared in 10mM phosphate buffer at pH = 8.0, introduced into the cylindrical quartz cuvette, and degassed as mentioned previously. Due to the overwhelming number of photoproducts formed from this reaction, the eluant was monitored at 215nm and at 260nm, and a more shallow HPLC gradient was used. The gradient was initially 5% "B" for five minutes ramped to 20% "B" in 30 minutes (0.5%/min instead of 1.0%/min). The reaction was terminated after 60 minutes, when it appeared that >90% of the TpdA had been consumed.

2.3 RESULTS AND DISCUSSION

In all five irradiations, candidate crosslinked species were isolated by HPLC and characterized by mass spectrometry. Figures 2.1-2.2 show a typical HPLC run of a peptide-nucleotide mixture before and after irradiation. Peaks marked with an arrow were likely candidates for crosslinked species, as those peaks were not present in separate irradiations of the peptide and the nucleoside or nucleotide.

Chromatographic isolation of crosslinked compounds

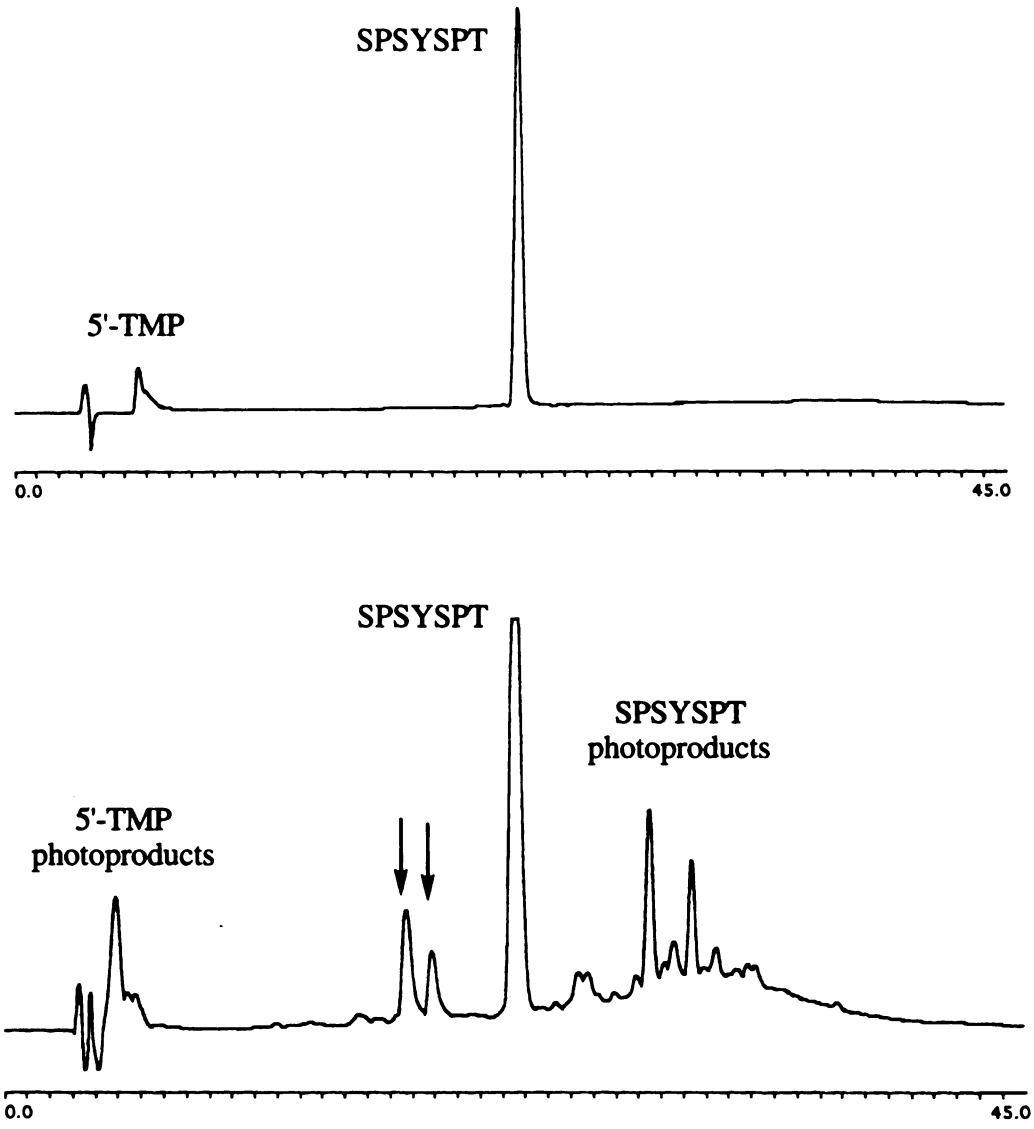
SPSYSPT-ThD conjugates. Under the HPLC conditions mentioned in the Materials and Methods section, thymidine elutes at 12.4 minutes, while the unreacted SPSYSPT elutes at 21.4 minutes. By irradiating solutions containing only thymine or only SPSYSPT, we determined that thymidine photoproducts elute at 8.1 and 9.3 minutes; while peptide photoproducts elute after the unreacted peptide. A majority of these photoproducts appear to be from the peptide itself, though a few are thought to be due to secondary photoproducts involving the peptide and the nucleoside. Further characterization of these photoproducts has not been undertaken. A candidate crosslinked peak eluted just before the unreacted peptide.

SPSYSPT-TMP conjugates. Thymidine-5'-monophosphate elutes at 5.7 minutes; thymidine monophosphate dimers elute shortly after the void volume; peptide secondary photoproducts elute after the peptide. An example of an HPLC profile before and after irradiation is shown in Figures 2.1-2.2. Two candidate peaks were collected, one having a retention time of 18.4 minutes, the other, 19.1 minutes.

SPSYSPT-Ud conjugates. The crosslinked peak is expected to have HPLC retention properties similar to that of the SPSYSPT-thymidine conjugate. Candidate crosslinked peaks were collected and lyophilized to dryness.

SPSYSPT-UdMP conjugates. In this photoreaction two candidate crosslinked complexes were observed, having retention times of 19.1 and 21.0 minutes. Both peaks were collected and lyophilzed to dryness.

SPSYSPT-TpdA conjugates. Under the different HPLC conditions used (see Materials and Methods), TpdA elutes at 10.9 minutes, while SPSYSPT elutes at 16.9 minutes. Only one candidate species was observed, which eluted just before the peptide at 16.5 minutes.



Figures 2.1-2.2 HPLC chromatograms of SPSYSPT and thymidine-5'-monophosphate before and after irradiation with a low-pressure Hg lamp. Peaks marked with arrows are likely candidates for crosslinked complexes.

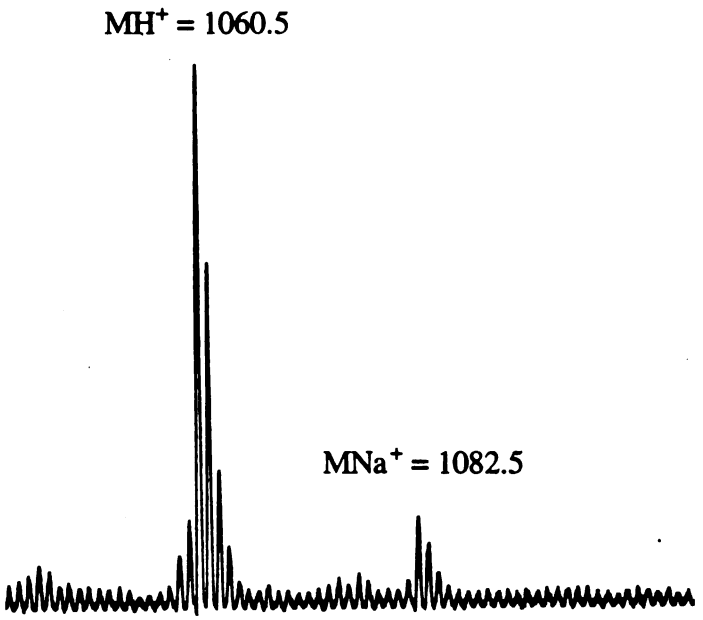


Figure 2.3 LSIMS(+) of a candidate peptide-nucleotide species. The peak at m/z = 1060.5 corresponds to that of the peptide plus that of TMP. The peak at m/z = 1082.5 is that of the same crosslink with a Na^+ adduct.

Molecular weight determinations of the crosslinked species

LSIMS(+) can be used to determine molecular weights of peptide-nucleoside and peptide-nucleotide adducts with better than unit mass precision. In the case of LSIMS(+) and high-energy tandem CID, one needs approximately 100-500pmol of the crosslinked material to get data that unambiguously localizes the attachment site of the nucleoside or nucleotide.

It should be noted that in the photochemical reactions involving thymidine and tyrosine, the predicted masses of the adducts formed can be predicted simply by summing up the masses of the individual species. This can be rationalized by examining the proposed mechanism of formation of the tyrosine-thymidine adduct, which involves a proton transfer, followed by radical combination[69].

In the photoreaction between SPSYSPT (monoisotopic mass = 737.4 Da) and ThD (monoisotopic mass = 242.1 Da), only one candidate peak was observed and

collected. LSIMS(+) revealed the protonated monoisotopic mass (MH^+) of the crosslink to be 980.4 Da, in agreement with the sum of the masses of the reacting species, plus a proton. The photoreaction between SPSYSPT and TMP appeared to produce two candidate crosslinked species. In this case, it appears that the presence of the phosphate group allows for resolution of the two peaks not only from the unreacted peptide, but also from each other (Figures 2.1-2.2). Both photoproducts were determined to have the same monoisotopic mass. An example of a molecular weight determination by LSIMS(+) of one of the photoadducts formed between SPSYSPT and TMP (monoisotopic mass = 322.1 Da) is shown in Figure 2.3. The observed m/z of 1060.5 corresponds to the sum of the two, which indicates that this species is a covalent SPSYSPT-TMP adduct. The peak at m/z = 1082.5 is indicative of an MNa^+ species, which is not unusual to find, even when the photoproducts have been desalted by RP-HPLC.

In the photoreaction between SPSYSPT and TpdA (monoisotopic mass = 555.2 Da) only one candidate peak was observed and collected. The chromatographic isolation of this adduct proved to be somewhat challenging, as it was observed that the crosslinked peak nearly coeluted with the unreacted peptide and a TpdA photodimer (data not shown). The observed monoisotopic mass of the adduct (MH^+ = 1293.6 Da) is in agreement with that of the sum of the two reacting species, plus a proton. Therefore, in each of the three cases, the monoisotopic mass of the photoproduct(s) formed is consistent with a photoaddition reaction.

In the systems involving the adducts formed between SPSYSPT with BrUd and BrUdMP, the masses observed do not correspond to the sum of the two reacting species. The observed MH^+ of the SPSYSPT-Ud conjugate is 964.4 Da, while the observed MH^+ of the two SPSYSPT-UdMP conjugates were found to be 1044.5 Da. The photoproducts formed which would be consistent with a photoreaction that has lost the bromine atom. It would also be consistent with a photoreaction that does not involve saturation of the 5,6-double bond of the pyrimidine. The proposed mechanism[75] does not involve a proton

transfer, and so the new bond formed between tyrosine and uracil displaces two hydrogen atoms that would normally be attached. So the expected mass of the crosslinked species should be the sum of the masses of the peptide and the uracil-nucleoside (or nucleotide) minus two, in accord with our mass spectrometric results.

Determination of the location of the photoadduct by high-energy tandem CID

Tandem CID mass spectrometry is a powerful analytical tool in determining the sites of modification of covalent peptide-nucleotide complexes. In general, CID of peptides results in fragmentation along the peptide backbone, giving characteristic fragments which can be categorized by whether the charge is maintained on the C-terminus (x, y, z) or the N-terminus (a, b, c)[169]; see Figure 1.15. In addition, we also observed loss of side chains and of the nucleoside (or nucleotide) itself.

In all cases, the mass spectra revealed that it was, indeed, the tyrosine residue modified. Figures 2.4-2.8 show representative CID fragmentation spectra for the five crosslinked species. All spectra reveal the characteristic C-terminal and N-terminal fragments in the low mass region, indicating that none of the amino acids at the C- or N-terminus were modified. In addition, all spectra reveal a region in the middle of the spectrum in which there were few observed peaks, which might indicate the loss of a rather large molecular weight species, such as what would happen from a loss of a nucleoside- or nucleoside-modified amino acid.

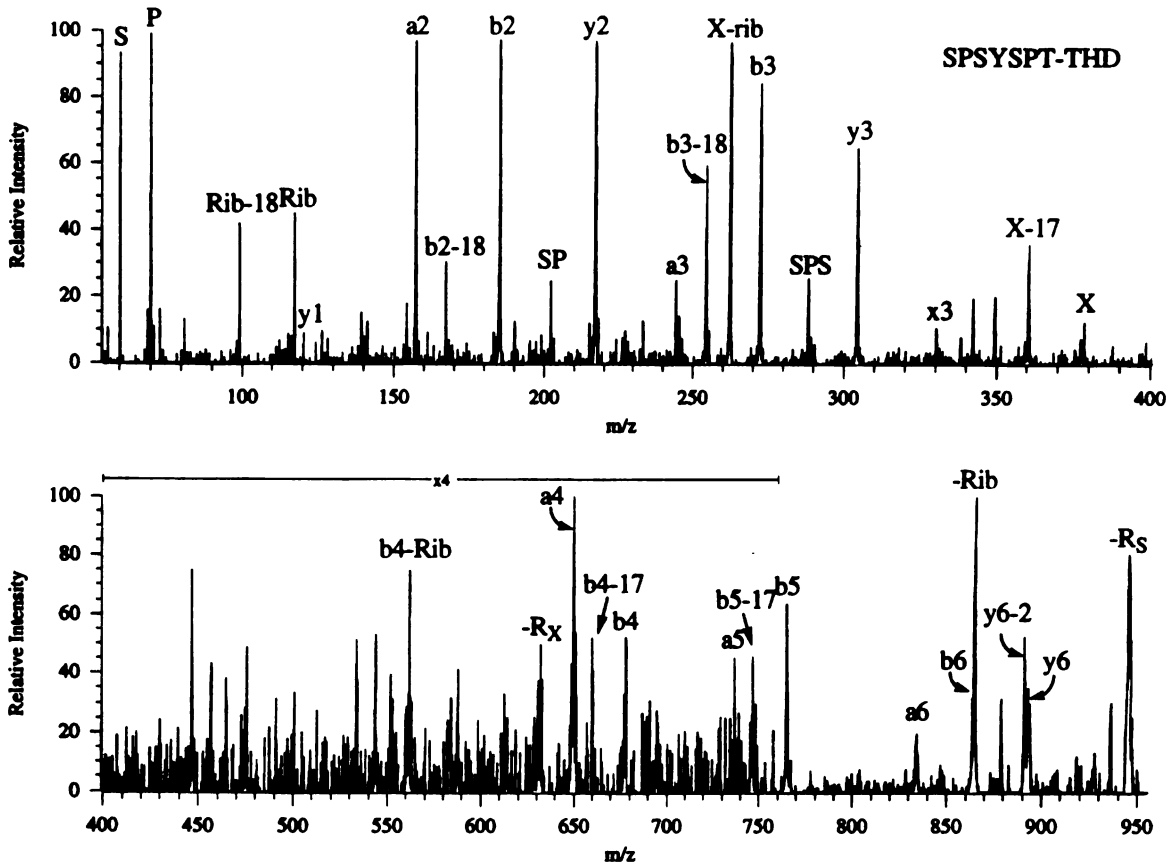


Figure 2.4 High-energy tandem CID of the suspected crosslinked species having an MH^+ of 980.4 Da. The nomenclature for the peak assignments are described in the text. The units for the peak intensities are arbitrary.

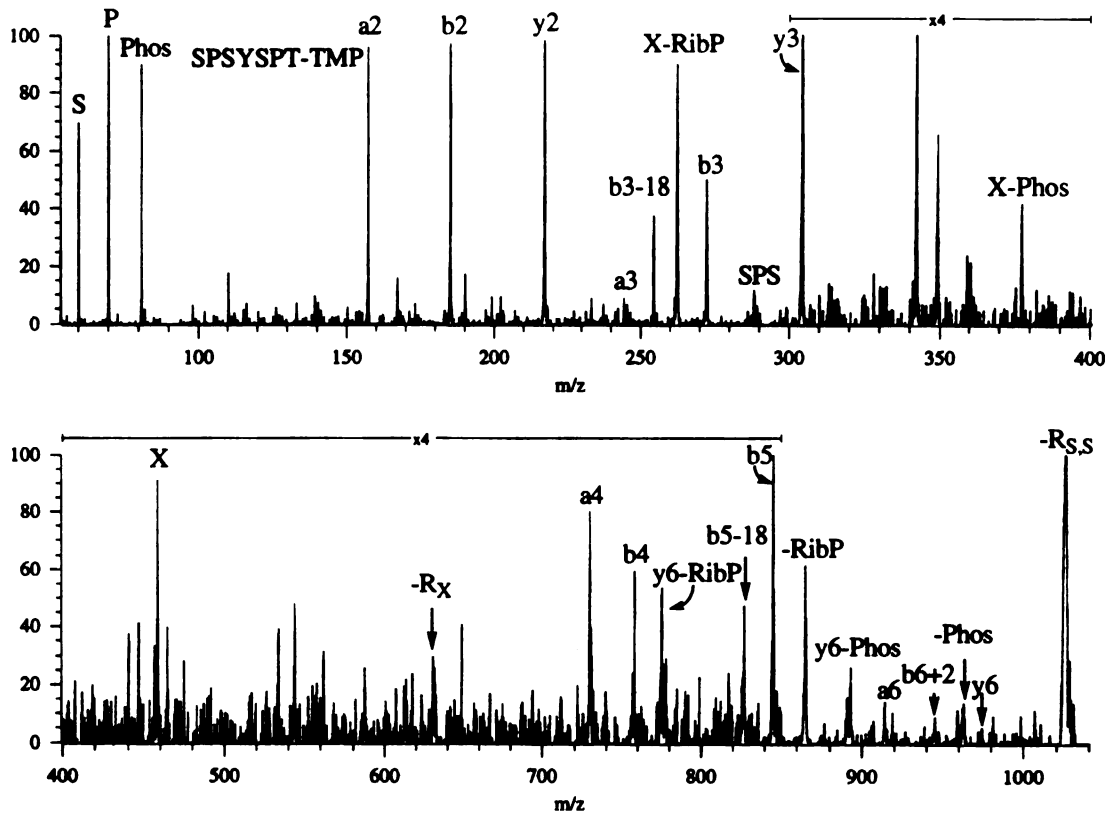


Figure 2

Figure 2.5 High-energy tandem CID of the suspected crosslinked species having an MH^+ of 1060.5 Da. The imino-(tyrosine-thymidine monophosphate) ion at m/z = 458.2 is clearly present.

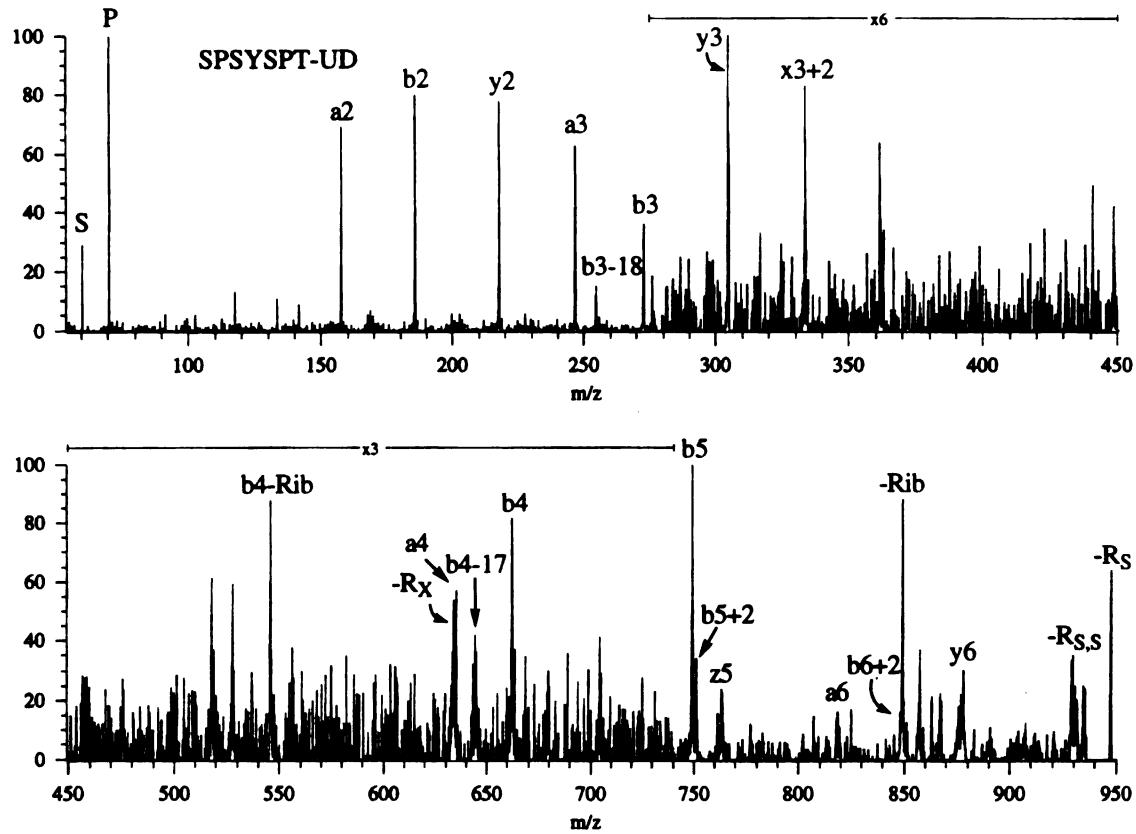


Figure 2.6 High-energy tandem CID of the suspected crosslinked species having an MH^+ = 964.4 Da. The bromine is lost in the photochemical reaction.

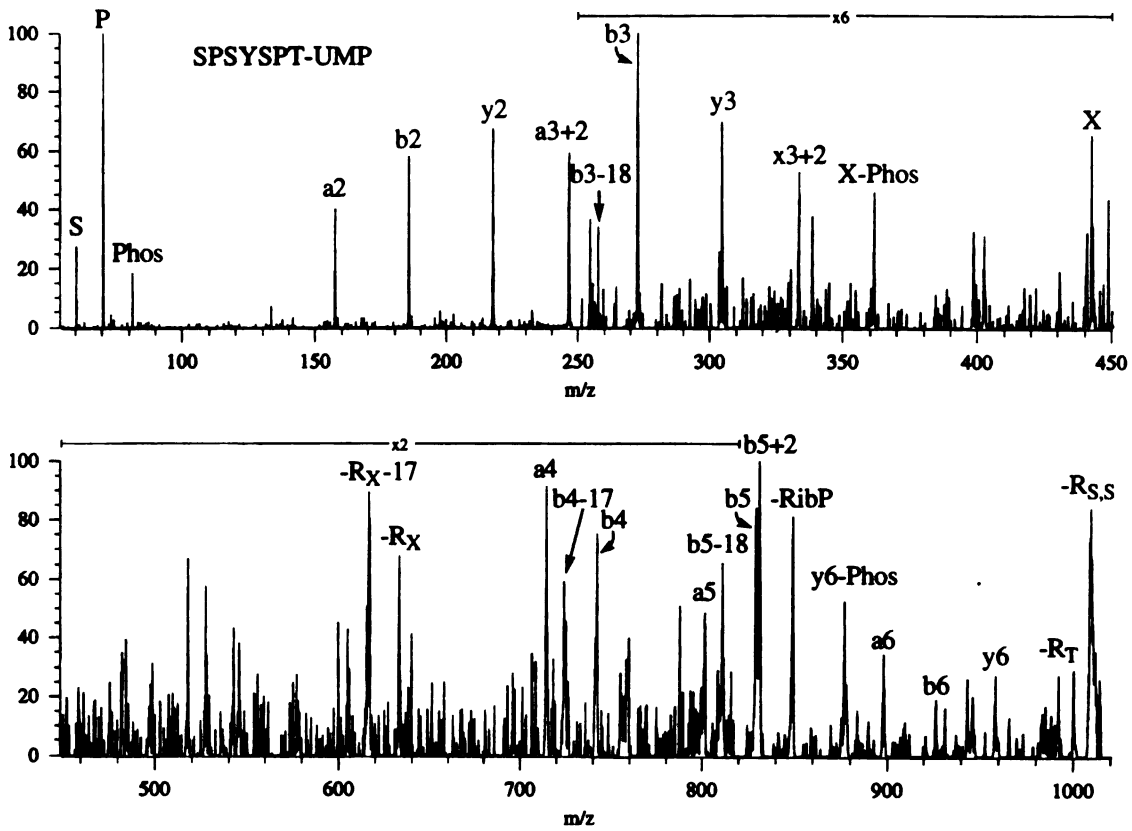


Figure 2.7 High-energy tandem CID of the suspected crosslinked species having an MH^+ of 1044.5 Da. The imino-(tyrosine-uridine monophosphate) ion at m/z = 442.2 is clearly present.

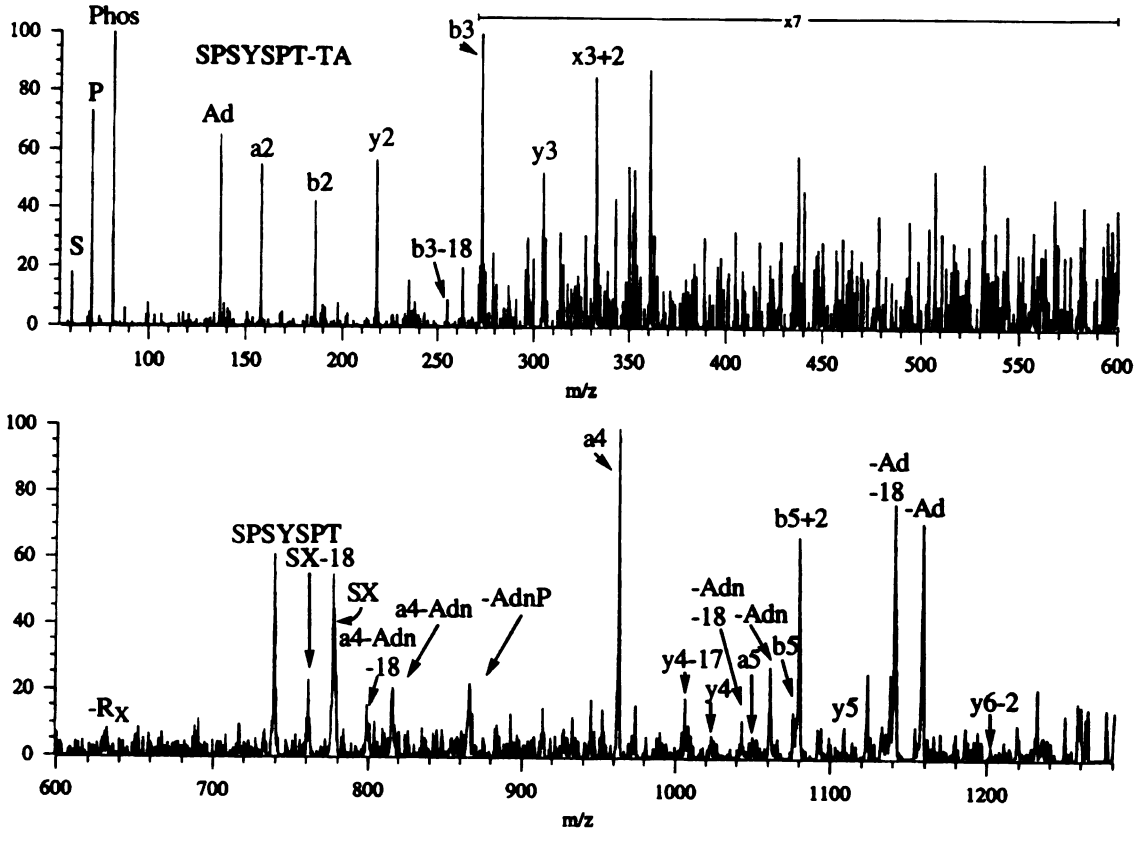


Figure 2.8 High-energy tandem CID of the suspected crosslinked species having an MH^+ = 1293.6. In this spectra, one observes fragmentation of both the peptide and the dinucleotide. See discussion for details.

At the high mass region the spectra become a little more difficult to intrepret, most likely due to a combination of side chain (denoted -R_S for the loss of a serine hydroxyl group and -R_T for the loss of a threonine side chain) and nucleoside or nucleotide fragmentation off the intact species. All but one spectrum reveal the loss of ribose (denoted -rib) or a ribose-phosphate (-ribP) from the intact crosslink. The CID spectrum of the SPSYSPT-TpdA crosslinked species (Figure 2.8) reveals loss of an adenine (-Ad), adenosine (-Adn) and adenosinemonophosphate (-AdnP) which would be indicative that it was the thymidine, and not the adenosine residue, that was covalently attached to the peptide.

In the cases involving the SPSYSPT-TMP and SPSYSPT-UdMP photoproducts (Figures 2.5 and 2.7), the imino-tyrosine (nucleotidemonophosphate), or X, peak is evident (m/z = 458.2 Da for the tyr-TMP conjugate; m/z = 442.2 Da for the tyr-UdMP conjugate). This information, in conjunction with the mass spectrometric sequencing results, lends corroborating evidence to the assignment of the tyrosine as the modified residue. In both of these cases, two photoproducts were isolated. Both LSIMS(+) and tandem CID produced essentially identical spectra (mass spectra for the second crosslinked species are not shown), which gives evidence that we had isolated two isomers, and both the isomers in each of the irradiated systems had the nucleotides attached to the tyrosine residue.

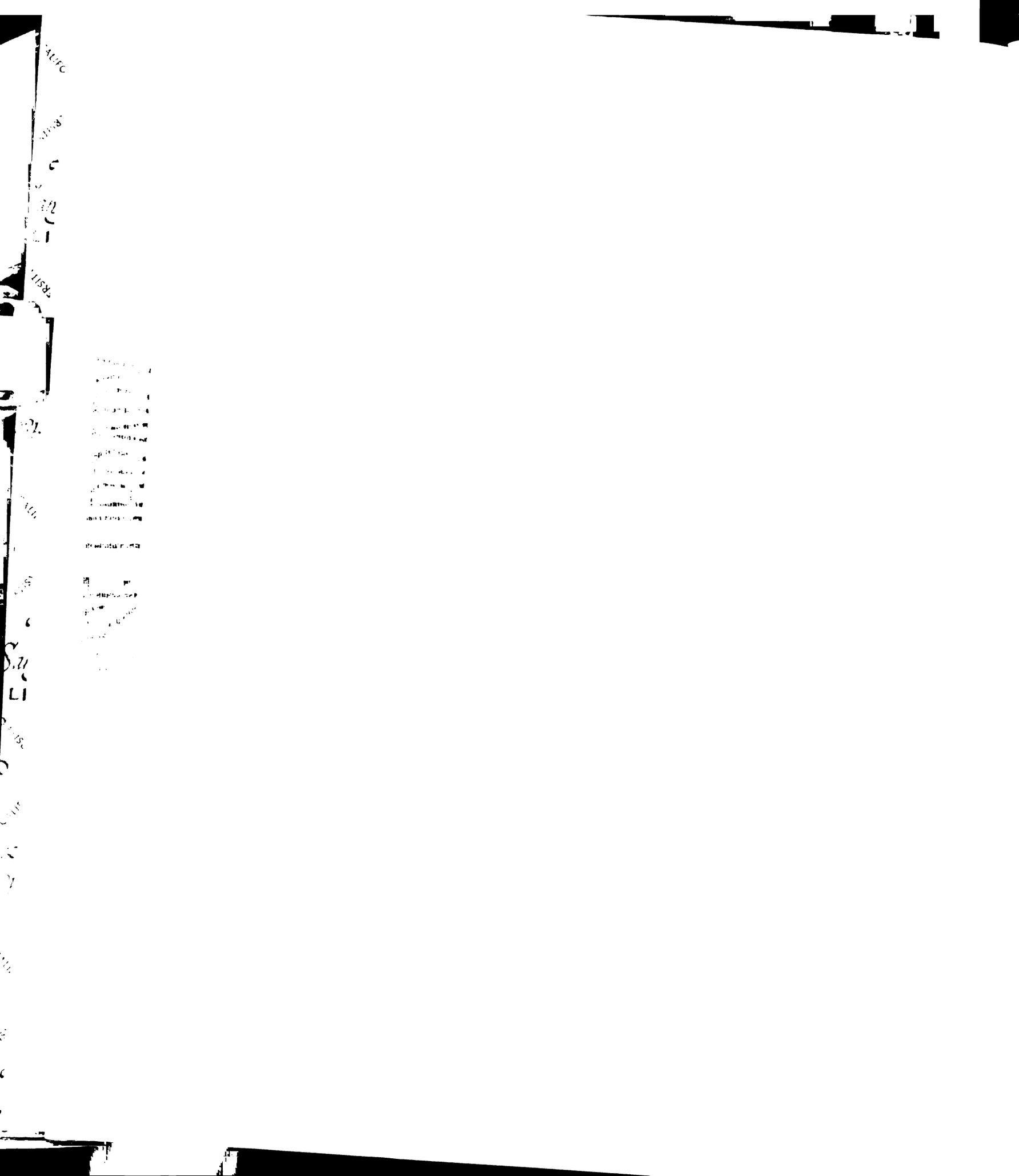
The poor yields of the TpdA-SPSYSPT photoproduct (approximately 200pmol from a 200nmol irradiation as compared to about 2-5nmol from the other irradiations) might be attributed, at least in part, to a decrease in the number of "successful" collisions among the reacting species. Adding a relatively unreactive nucleotide, however, appeared to lower the yields of the reaction considerably, either due to a dissapative process, or perhaps the crosslinked complex itself was still photoreactive. If one assumes that the system had undergone random collisions during the irradiations, then by placing

an unreactive nucleoside adjacent to the thymidine (or analog), one can envision the crosslinking as a "pseudo third order" reaction in which the tyrosine, the photon and the thymidine must all be in close proximity. The probability of the formation of these exciplexes will, of course, drop as the number of unreactive residues increases. In addition, there is always competition with intramolecular reactions, dimerizations, and secondary photoproducts, as well as a shielding effect of adding more nucleic acid bases to the reaction mixture.

2.4 CONCLUSIONS AND FUTURE PROSPECTS

Our results show that mass spectrometry can unambiguously identify the amino acids which have been photochemically crosslinked to nucleosides and nucleotides. The presence of the phosphate group or an additional nucleotide does not appear to alter the ability to perform CID on the crosslink. The polar phosphate group appeared to make the chromatographic separation a little easier, as we were able to resolve nucleotide-modified peptides from unreacted peptides much more readily than nucleoside-modified peptides. Our methods are applicable to both systems which involve the irradiation of traditional and brominated nucleotides, as both types of covalent crosslinks formed appear to be amenable to chromatographic and mass spectrometric analysis.

Some of the limitations of this method are that we have implied that a protein-nucleic acid crosslink can be completely digested by chemical or enzymatic means, leaving only a small peptide and a covalently-bound nucleotide (or, possibly, a dinucleotide). The practical upper limit of this mass spectrometric method is around 2500 Da. Yet it is possible that even under ideal conditions, some of the crosslinked species will have an m/z > 2500, which would make high-energy tandem CID difficult to perform. There is also the concern of whether the fragmentation of the oligonucleotide might dominate the spectra, and we would, therefore, be unable to sequence the peptide. Finally, given the enormous complexity of photoproducts resulting from a UV-irradiation



of a protein-nucleic acid complex, it would be difficult to presume that all the resultant crosslinks could be detected and sequenced by high-energy tandem CID mass spectrometry. Many of these limitations have recently been overcome with the use of more versatile mass spectrometric techniques, which are discussed in Chapter 3.

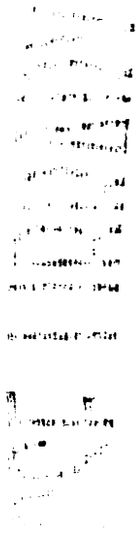
CHAPTER THREE

PROBING THE BINDING REGION OF THE ssDNA BINDING SUBUNIT OF RAT DNA POLYMERASE β WITH UV CROSSLINKING AND MASS SPECTROMETRY

3.1 INTRODUCTION

Of particular interest in this chapter is the use of low intensity ultraviolet (LIUV) and nanosecond-pulsed laser UV (NPUV)[113-117, 170-172] crosslinking to probe non-covalent interactions between proteins and nucleic acids by creating covalent bonds between photoreactive residues, followed by applications of mass spectrometry to locate the peptides and, ultimately, the amino acids that are covalently bound to the nucleic acid. We will compare the results obtained from NPUV crosslinking directly with other structural studies done on our protein. To date, NPUV crosslinking has been used on a number of protein-nucleic acid systems. Harrison and co-workers[172] successfully crosslinked *Escherichia coli* RNA polymerase to T7 DNA using a single-pulse from a KrF laser emitting principally at 248nm. Perhaps the most significant contributions to the field has been work done by von Hippel, Hockensmith and their co-workers[113, 114, 116, 117]. Using a neodymium:yttrium-aluminum-garnet (Nd:YAG) laser emitting principally at 266nm (the same laser used in our studies), they explored the individual protein-nucleic acid interactions of the five-protein T4 phage replication holoenzyme[114], as well as the becteriophage T4 gene 32 system[113, 115].

There have been a few papers which use mass spectrometry for analyzing photochemically crosslinked protein-nucleic acid systems. Sastry and co-workers used



mass spectrometry to locate a peptide from a tryptic digest of T7 RNA polymerase and psoralen-tagged $p(dT)_{12}$ that had been LIUV-crosslinked[173]. Jensen and co-workers used MALDI-TOF MS to analyze the photoproduct formed between T4-phage gene 32 ssDNA binding protein and oligo $p(dT)_{20}$ that had been crosslinked *via* NPUV[143]. A subsequent report from the same labs shows how MALDI-TOF MS can be used to analyze crosslinked peptide-oligonucleotide complexes isolated from a tryptic digest of LIUV-mediated crosslinked complexes of uracil-DNA glycosylase and oligo $p(dT)_{20}$[121].

The protein used in our study is the ssDNA binding subunit of recombinant rat DNA polymerase β. DNA polymerases, enzymes which repair and replicate genomic DNA, have received considerable attention recently regarding their specific primary functions in the living cell, and the mechanisms by which they catalyze the formation of phosphodiester bonds. Mammalian DNA polymerase β, a small (approximately 40 kDa) protein, has been cloned and overexpressed in *E. coli*, and purified to homogeneity[174, 175].

The crystal structure of DNA polymerase β complexed with a template primer has been recently solved[44, 45]. Curiously, the crystal structure reveals that the hinge portion extends away from the core polymerase, while hydrodynamic studies indicate that the structure is relatively compact in solution[176]. The enzyme appears to be divided into two distinct domains separated by a proteolytically sensitive and flexible hinge: the N-terminus, consisting of the first 85 or so amino acids, has ss-DNA binding but no apparent catalytic activity[175]; the rest of the protein binds to ds-DNA and has catalytic activity[177]. One can infer the possible roles of each of the domains solely on this basis--the core polymerase catalyzes the formation of the phosphodiester bonds, while the flexible single-stranded binding domain might serve to stabilize the open regions of the complex and might allow the core polymerase easier access to the DNA gaps which need filling. While the crystal structures of DNA polymerase β complexed with primers have

clearly revealed the catalytic site, there were still many unanswered questions left regarding the specific role of the ss-DNA binding region, and the mechanisms for DNA replication *in vivo*.

In an initial photochemical crosslinking study[126], Prasad and co-workers irradiated the ss-DNA binding subunit of this protein (hereafter referred to as pol β, whereas the entire protein will be referred to as DNA polymerase β) with a LIUV lamp in the presence of oligo $[^{32}P]p(dT)_{16}$. After proteolytic digestion and subsequent PTH-amino acid sequencing, they concluded that residues Ser^{30} and His^{34}, by virtue of their comparatively low yields in the sequencing, were believed to be crosslinked to the oligonucleotide, and thus resided in the ssDNA-binding domain. These results have been reproduced in our lab (see discussion below). NMR spectroscopic studies of the ^{15}N and ^{13}C double-isotope labeled protein-$p(dT)_8$, complex, however, reveal that the peptide segment containing these amino acids actually lies in a rather flexible loop, and that the (more rigid) DNA binding domain appears to be situated much closer to the C-terminus[47].

In this chapter we will compare the results obtained from our LIUV and NPUV crosslinking studies of this protein complexed with the heptameric oligonucleotide d(ATATATA), with those obtained from previous crosslinking and spectroscopic studies. The selection of the oligonucleotide is not arbitrary. The reasons for this selection were twofold: first, we could eliminate many of the difficulties associated when digesting and chromatographing irradiated poly(dT) oligonucleotides, while still retaining at least a portion of the pyrimidine reactivity. Second, we could use the relative non-reactivity of the intervening adenosines as chromophores when chromatographing the irradiated complexes (by monitoring at 260nm), rather than having to rely on radiolabeled oligonucleotides. Eliminating this need for radiolabeled material can be doubly advantageous, because if one were to carry out a chemical or enzymatic nucleolytic

digestion on these peptide-oligonucleotide complexes, then the ^{32}P label would almost certainly be lost. Even under our HPLC conditions using a wide-bore FPLC column (see Materials and Methods), we can consistently detect irradiated oligo d(ATATATA) in the 10-picomole range (data not shown). Since MALDI-TOF MS has at least the same level of sensitivity, we can use both these methods in conjunction to identify peptides, and ultimately the amino acids that are covalently crosslinked to the oligonucleotide.

While our attempts at determining the points of contact between LIUV- and NPUV-crosslinked pol β with d(ATATATA) at the peptide level were successful, we were not able to use mass spectrometry to definitively pinpoint an amino acid-nucleobase comtact point from a protein-oligonucleotide complex which had been irradiated with UV light. Our results show the potential for such an analysis being performed, as we were successful in obtaining *partial* sequence information of both the peptide and the oligonucleotide from *in situ* digests of a peptide-oligonucleotide crosslink.

3.2 MATERIALS AND METHODS

General considerations

Recombinant rat DNA polymerase β (single-stranded DNA binding subunit) was a generous gift from Drs. Rajendra Prasad and Samuel Wilson at the University of Texas Medical Branch in Galveston. Protocols for cloning and isolation are described elsewhere[174, 175].

All photoproducts were isolated and purified using high-performance liquid chromatography (HPLC) equipped with a 100μL injection loop. The system consisted of two Rainin Rabbit Pumps controlled by a Dynamax (Version 1.2) system and data acquisition package. The eluant was monitored at either 215nm (reversed-phase HPLC) or 260nm (anion-exchange HPLC) using a Kratos 783 variable-wavelength detector equipped with a deuterium lamp. For reversed-phase HPLC, a Vydac C-18 analytical

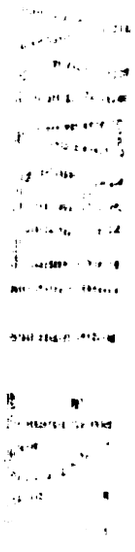
(4.6 x 250mm) column with a 300 Å pore size was used. The flow rate was 1.0mL/min. The solvent system consisted of degassed HPLC grade water (Fisher) with 0.10% (v/v) trifluoroacetic acid (TFA) (sequencing grade, Pierce, solvent "A") and HPLC grade acetonitrile (Fisher) with 0.088% (v/v) TFA (Solvent "B"). For anion-exchange HPLC, a Mono Q HR 5/5 FPLC anion-exchange column (Pharmacia) was used. Flow rates were 0.70ml/min. The solvent system used was the following: solvent "A" consisted of 50mM NH_4^+-bicarbonate, pH = 8.5, while solvent "B" consisted of solvent "A" + 1.0M NH_4^+-acetate, pH = 8.5. The rationale behind using ammonium salts was that they were easily removed by lyophilization. All solutions for anion-exchange chromatography were filtered and degassed prior to use, stored at 4°C in sealed bottles when not in use, and made fresh every two days.

Tandem mass spectrometry was performed on a Kratos Concept IIHH four-sector mass spectrometer equipped with a cesium ion source and an electrooptical array detector[100]. To prepare samples for tandem mass spectrometric sequencing, approximately 500pmol of the lypophilized peptides were taken up in 5.0µL of 0.1% TFA. This solution was mixed with an equal amount of a solution containing 5% (v/v) acetonitrile, 5% (v/v) thioglycerol, and 0.1% (v/v) TFA.

MALDI-TOF MS was performed on a PerSeptive Biosystems Voyager Elite mass spectrometer. In order to prepare samples for MALDI-TOF MS, the samples were reconstituted in an appropriate amount of doubly-distilled (DD) H_2O, so that the concentration of crosslinked species was approximately 10pmol/µL. In general, 1.0µL of the sample was mixed with 1.0µL of 100mM diammonium hydrogen citrate (pH = 5.1) (Aldrich) and 2.0µL of a solution of 20mg/mL 2,4,6-trihydroxyacetophenone (Aldrich) in neat HPLC grade acetonitrile (Fisher), to give a final solution containing 25mM citrate and 10µg/µL of the 2,4,6-trihydroxytacetophenone. Samples were vortexed and centrifuged briefly. The MALDI matrix was recrystallized from (90:10) DD

H_2O:ethanol, air-dried and stored at 4°C prior to use. All matrix solutions were made fresh daily. A 1.0μL aliquot (or approximately 2-3pmol sample, as the solution has been diluted 1:4) was removed from the mixture and placed on a MALDI sample well and allowed to air dry. For the protein-oligonucleotide complexes, the mass spectrometer was internally calibrated with *E. Coli* thioredoxin (MH^+ = 11,676.4 average mass), and the samples were prepared for mass spectrometry without any prior chromatographic cleanup. For peptide-oligonucleotide mass measurements, the mass spectrometer was externally calibrated with oligonucleotides d(ATATATA) (MH^+ = 2104.5 Da avg.) and d(GGTTTTTTGG) (MH^+ = 3081.1 Da avg). The oligonucleotides were purchased from Genset Technologies.

TPCK-treated trypsin and Endo Glu-C (from *Saphylcoccus aureus*) protease were purchased from Sigma. One milligram of each was weighed out and dissolved in 1.00mL HPLC H_2O, and pipetted out in 10μL (10μg) or 100μL (100μg) aliquots, lyophilized to dryness, and stored at -20°C until use. In general, digestions were done at a pol β concentration of 10pmol/μL. For peptide mapping, we used the following protocols. For tryptic digests, a 1.0nmol aliquot (approximately 10μg) of pol β was added to a 100μL solution containing 100mM NH_4^+-bicarbonate, pH = 8.5. The solution was brought to 37°C. A 0.2μg aliquot of trypsin (for an initial protease:protein ratio of 1:50, by weight) was added, and the solution was incubated for two hours at 37°C. Subsequently, a second aliquot of 0.2μg of trypsin (for a final protease:protein ratio of 1:25, by weight) was added, and the solution was incubated overnight. For Glu-C digests, a 1.0nmol aliquot of pol β was added to a solution containing 50mM Na^+-phosphate buffer (pH = 7.4) and equilibrated at 37°C. A 0.5μg aliquot of Glu-C (1:20, by weight) was added to the solution and incubated for two hours. A subsequent aliquot of 0.5μg Glu-C (total of 1:10, by weight) was added and the digest was allowed to proceed overnight.



For peptide mapping, following a 1:25 (w/w) tryptic or 1:10 (w/w) Glu-C digest, the mixture was lypohilized to dryness, reconstituted with 50μL of 0.1% TFA, and chromatographed over RP-HPLC using the following gradient: initially 2% "B", ramp to 52% "B" in 50 minutes. Peptides were collected and lyophilized to dryness.

For proteolytic digestion of protein-oligonucleotide crosslinked complexes, a 1:20 (w/w) aliquot of trypsin was added to the solution and the mixture was incubated at 37°C for two hours. Subsequently, a second aliquot of trypsin (1:20, by weight, for a final protease:protein ratio of 1:10) was added to the solution, and the solution was incubated at 37°C overnight.

Peptide sequence analysis was performed on an Applied Biosystems 470A gas-phase sequencer equipped with an on-line ABI model 130A PTH analyzer. All sequencer reagents and related sequencer supplies were purchased from Applied Biosystems, with the exception of HPLC grade acetonitrile, which was from Fisher.

LIUV crosslinking

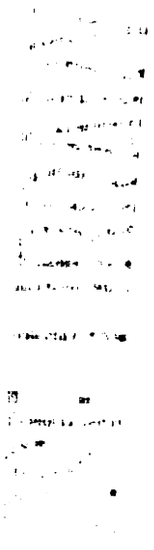
All irradiations were done in the cold room (4°C) with a low-pressure 15W germicidal lamp equipped with a Vycor shield, emitting principally at 254nm. In neither the LIUV nor the NPUV crosslinking experiments was oxygen removed. In general, a 30nmol aliquot (approximately 300μg) of the protein and 30nmol of the oligonucleotide (approximately 65μg) were added to a solution containing 10mM Tris buffer pH = 8.0, 10mM NaCl, and 1.0mM EDTA. The solution was allowed to incubate at room temperature for approximately 15 minutes. The sample was the transferred into a 1.0cm by 4.0cm quartz cuvette with a 1.0mm path length (Hellman) and equilibrated in the cold room at 4°C. The optical density of this solution is approximately 0.82 at 260nm. The cuvette was then placed flush against a germicidal UV lamp, aproximately 2.0cm from the center of the lamp. Ten-μL aliquots were then taken out at 0, 1, 5, 10, 20 and 30 minutes irradiation and run on denaturing SDS-PAGE (20%) in a running buffer

consisting of 3.0g/L Tris base, 14.4g/L glycine and 0.1% SDS. The stacking gel (6.0% acrylamide) consisted of 19.5g/L Tris•Cl, pH = 6.8 and 0.1% SDS; the running gel (20% acrylamide) consisted of 58.7g/L Tris•Cl, pH = 8.8 and 0.1% SDS. The gel dimensions were 7cm by 8cm by 0.75mm. The mini-gel apparatus (Mighty Small) was purchased from Hoefer Scientific Instruments. The gel was stained with Coomassie blue. It is expected that any crosslinked species formed will migrate slightly slower than that of the unreacted protein. The gel was visualized and the amount of crosslink was quantitated by scanning the gel on a Personal Densitometer from Molecular Dynamics.

NPUV crosslinking

All NPUV laser irradiations were done in collaboration with Dr. Mark Young in the laboratory of Professor Peter von Hippel at the University of Oregon. For the NPUV crosslinking experiments, a Nd:YAG laser (model DCR-3G, Spectra-Physics, Mountain View, CA) was employed. Details for the schematics of the NPUV laser are described elsewhere[116]. In general the Nd:YAG emits light principally at 1064nm, which is frequency-quadrupled through a Brewster-angle prism to 266nm. For our experiments, the "single-pulse" method was employed. The single pulses were approximately 8 nanoseconds in duration. The energy of the pulse was measured with a pyroelectric detector (model 70825, Oriel, Stratford, CT) and was monitored with a digital storage oscilloscope (model 2201, Tektronix, Wilsonville, OR).

Approximately 2.0mg (200nmol) of the ss-DNA binding subunit of rat DNA pol β, along with 200nmol oligo d(ATATATA) were dissolved in 2.0mL of a buffer containing 10mM Tris (pH = 8.0), 10mM NaCl, and 1.0mM EDTA so that the final concentration of the reactants was approximately 100pmol/μL. The reaction mixture was allowed to incubate at room temperature for at least 15 minutes. Ten microliter aliquots (which corresponds to approximately 1.0nmol of each of the protein and the oligonucleotide) of this solution were taken out and introduced into 1.7mL polypropylene



eppendorf tubes. The samples were centrifuged briefly at ~1000g (to ensure the solution was at the bottom of the tube), and subsequently placed on ice until ready for irradiation. An initial set of samples was given 0, 1, 2, 4, 8 and 16 pulses to determine if, and to what extent, crosslinking occurs. After it was determined that a single NPUV pulse resulted in crosslinking, the remaining samples were each irradiated with a single pulse. Subsequent to irradiation, samples were pooled and stored at -70°C until ready for chromatographic separation.

Digestion and chromatographic workup of the LIUV- and NPUV-crosslinked peptide-oligonucleotide complexes

In general, we have developed a protocol that avoids extremes of heat and of pH, so as to maintain the integrity of as many crosslinks as possible. It consists of a two-step chromatographic process which removes the unreacted oligonucleotide first, followed by a proteolytic digestion and second chromatographic step which removes the unreacted peptides.

A 20nmol (LIUV) or a 100nmol (NPUV) aliquot of the sample was then lyophilized to dryness and reconstituted in sufficient solvent "A" to bring the concentration of pol β to 1.0nmol/μL. The mixture was then injected (either entirely or in 20nmol aliquots) onto anion-exchange HPLC using the following gradient at a flow rate of 0.70mL/min: isocratic 20% "B" for 15 minutes, then ramped to 100% "B" in 40 minutes. Given the fact that this protein has a net charge of +10 under these conditions, the unreacted pol β and the pol β-d(ATATATA) crosslinked species (with a net charge of +4) elute in the void volume, while the unreacted oligonucleotide elutes much later (approx. 70% "B"). The void volumes were collected, pooled and lyophilized to dryness, reconstituted with 1.0mL of 10mM NH_4^+-bicarbonate, and lyophilized again.

Following a 1:10 (w/w) tryptic digest, the solution was again lyophilized to dryness. The pellet was resuspended in an appropriate amount of Solvent "A" and

chromatographed exactly as previously mentioned. Uncrosslinked peptides elute at or near the void volume. Candidate crosslinked peptide-oligonucleotide complexes should elute sometime near that of the unreacted oligonucleotide, owing to the number of negative charges as compared to the number of positively-charged amino acids per peptide from a tryptic digest. By virtue of having a number of relatively non-reactive adenosines in the oligonucleotide, there is still sufficient absorbance at 260nm to detect at the low picomole level without the need for ^{32}P-labeling. Candidate peaks were collected and lyophilized, reconstituted in 1.00mL of 10mM NH_4^+-bicarbonate, and lyophilized again. If necessary, the samples were further desalted using a Nensorb 20 cartridge (DuPont), following manufacturer's instructions with some minor modifications. The cartridges were initially prewetted with 1.0mL HPLC methanol (Fisher), and then 3.0mL DD H_2O. Samples were reconstituted in 250µL DD H_2O, and passed through the column. After extensive washing with 5mL DD H_2O, samples were eluted with 1.0mL of a 50/50 (v/v) mixture of methanol/water. The eluted samples were then diluted to 1.5mL with DD H_2O and lyophilized to dryness. The resultant samples were split into two equal portions, one for Edman sequencing and the other for mass spectrometric analysis. MALDI-TOF MS was performed on the candidate peptide-oligonucleotide complexes as mentioned previously.

C-terminal sequencing of peptides and peptide-oligonucleotide complexes with carboxypeptidase Y

C-terminal sequencing of peptides and candidate peptide-oligonucleotide complexes was done with carboxypeptidase Y using a Sequazyme™ kit (developed by Patterson and co-workers[112]) available from PerSeptive Biosystems, following manufacturer's instructions.

Phosphodiesterase I digestion of oligonucleotides and peptide-oligonucleotide complexes

The 5'-endonuclease, phosphodiesterase I (from *Crotalus atrox* venom, Sigma), was used to digest the oligonucleotide. 1.2mg of the lyophilized protein (which corresponded to 0.52 units) was dissolved in 1.2mL DD H_2O. This solution was partitioned out in 10^{-3} unit (2.3μL) aliquots and lyophilized to dryness. (A unit is defined as the amount which will hydrolyze 1.0μmol of thymidine-5'-monophosphate-*p*-nitrophenyl ester per minute at pH = 8.9 and 25°C).

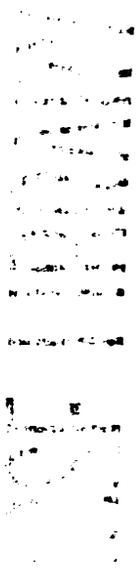
In general, candidate crosslinked peptide-oligonucleotide complexes were diluted to a concentration of 10pmol/μL and 50mM NH_4^+-citrate, pH = 9.4. A droplet of 0.5μL was placed directly on the MALDI sample well. To that sample was added 0.5μL of a solution containing the phosphodiesterase I (10^{-4} to 10^{-6} unit/μL, depending on the nature and extent of the digestion desired) in 50mM NH_4^+-citrate, pH = 9.4.

To digest candidate crosslinked peptide-oligonucleotide complexes, the samples were diluted with an appropriate amount of DD H_2O, so that the concentration of oligonucleotide ws approximately 10-20pmol/μL. An aliquot of 1-1.5μL was taken, diluted with an equal volume of 100mM citrate buffer (pH = 9.4), and digested for 20 minutes using the above protocol. Subsequent to digestion, 1.0μL of the matrix solution containing 20mg/mL 2,4,6-trihydroxyacetophenone in neat acetonitrile was added, and the sample was allowed to dry.

3.3 RESULTS AND DISCUSSION

Protein sequencing

Our initial characterization of the single-stranded binding subunit of rat DNA polymerase β revealed a discrepancy regarding its primary sequence. According to the published sequence[175], the protein should have an average MH^+ of 9482 Da, while MALDI-TOF MS showed that the average MH^+ was 9468 Da (Figure 3.1). This



discrepancy was resolved by examining the profile of the Glu-C digest of the protein. The mass error was traced to the C-terminal peptide, which had a measured monoisotopic MH^+ of 1389.6, 14 mass units lower than what would be predicted for the peptide (MH^+ = 1403.6 Da) given the published sequence[175]. Using a combination of Edman (N-terminal) and C-terminal sequencing, we were able to unambiguously show that the C-terminal amino acid of this peptide should be an asparagine, not the predicted lysine (data not shown). The residue weight of Asn is 14 Da less than that of Lys, thus explaining the discrepancy.

LIUV irradiation of pol β with d(ATATATA)

Figure 3.3 shows a 20% denaturing SDS-PAGE gel of the pol β/d(ATATATA) mixture after 0, 5, 10, 20 and 30 minutes. The gel shows the appearance of a second band with a molecular weight slightly higher than that of the unreacted protein. The yield of this apparent crosslink increases with irradiation time (Figure 3.4), but the reaction mixture appears to get a little more complicated as well, with the apparent formation of secondary photoproducts with higher molecular weights. It was determined visually, (taking into account of maximum yields of desired crosslink while minimizing undesirable crosslinks) that the optimum time for irradiation was approximately 20 minutes.

NPUV irradiation of pol β with d(ATATATA)

Figure 3.5 shows a 20% denaturing SDS-PAGE gel of the pol β/d(ATATATA) mixture after receiving 0, 1, 2, 4, 8 and 16 pulses. After a single NPUV pulse, a second band appears that migrates slower than that of the unreacted protein, with an apparent molecular weight of around 12kDa, which would correspond closely to that of the protein and the oligonucleotide. The yield of this apparent crosslink increases as the number of pulses increases; however, the reaction mixture appears to grow more complex as well. The gray area at the top of the gel is an artifact of the staining process.

MALDI-TOF MS of the NPUV-crosslinked protein-oligonucleotide complex

The MALDI-TOF mass spectrum of the reaction mixture (Fig. 3.6) reveals peaks that correspond to that of the unreacted oligonucleotide, the unreacted protein, and the protein-oligonucleotide complex. Due to the overwhelming number of different potential photoproducts that could be formed between these two reacting species, and lack of knowledge of the exact chemical nature of the crosslinks, it would be essentially impossible to predict the mass of the pol β-d(ATATATA) complex *a priori*. The best guess we can make is that the "predicted" mass of the crosslink is the sum of the protein and the oligonucleotide. As Table 3.1 shows, the observed mass of the putative crosslink does, indeed, appear to have this mass. The inset in Figure 3.6 reveals that the crosslink is apparently homogeneous. However, we would not be able to detect any mass transfers that may have resulted from the photochemical crosslinking of the protein and the oligonucleotide, or any other type of intramolecular crosslinking, such as a photorreaction between two proximal amino acids within the protein.

Edman sequencing of the LIUV-crosslinked peptide-oligonucleotide species

After a proteolytic digestion and anion-exchange chromatography, the candidate LIUV-crosslinked peptide-oligonucleotide complexes were submitted for PTH amino acid sequencing. The result from Edman sequencing is shown in Figure 3.7. We were able to identify the tryptic peptide sequence NVSQAIHK... (Asn^{28}-Lys^{35}), which is the same peptide identified by Prasad, *et al*[126]. There appeared to be a drop in yields of Ser^{30} and His^{34} (which was also noted by Prasad, *et al*[126]). However, we also noted an elevation in PTH-dehydroserine in cycle 3, which may have resulted from either a chemical or a photochemical loss of a water molecule from Ser^{30}. Unlike the results obtained previously, however, we were able to sequence this tryptic peptide up to Arg^{40} (NVSQAIHKYNAYR). There is an unusual drop in PTH-amino acid yield after Lys^{35},

which may indicate we were sequencing two peptides with the same N-terminus. The MALDI-TOF MS results confirm these findings, which are discussed later.

Edman sequencing of the NPUV-crosslinked peptide-oligonucleotide species

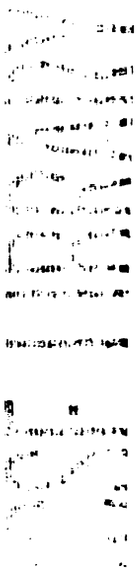
Edman sequencing of the NPUV-crosslinked peptide-oligonucleotide complexes proved to be fairly difficult to perform because of the low yields, even given the rather large amount of starting materials. The problem is compounded by the apparent coleution of several crosslinked species with each peak. From the size of the peaks from the HPLC chromatographic profile, and assuming the extinction coefficient for these complexes would be about the same as what one would find for irradiated oligo d(ATATATA), (the relative contribution from the protein at 260nm should be negligible) we have estimated that we were only able to recover about 100pmol crosslinked peptide-oligonucleotide complexes from a 50nmol irradiation, or about 0.2%. This amount, however, appears to be distributed rather equally among a number of different photoproducts; thus the yield of individual species can only be considerably less than 0.2%. Figures 3.8-3.11 show the results of PTH-amino acid sequencing of one of the peaks collected from anion-exchange HPLC. It appears there are a number of conjugates that coelute, which makes accurate quantitation of the peptide yields difficult. One of the crosslinked peptides tentatively identified, Asn^{28}-Lys^{35} (Figure 3.8), is the same peptide identified in a previous crosslinking paper[126] and in our own lab using low-intensity UV-crosslinking (Figure 3.7). The sequencing of the peptide reveals the same characteristic drops in PTH-amino acid yields of Ser^{30} and His^{34}. The yield of Asn^{28} is not reported, however, due to an unknown compound which coeluted with the PTH-Asn in the first cycle, making identification and quantitation of this particular amino acid impossible.

A second crosslinked peptide, YNAYR (Tyr^{36}-Arg^{40}), was also identified (Figure 3.9). The yields of Asn^{37} and Arg^{40} were lower than that of the other amino

acids, suggesting some form of modification of those amino acids. A third peptide, LPGVGTKIAEK (Leu^{62}-Lys^{72}), was also tentatively identified by Edman sequencing as being crosslinked (Figure 3.10). In this peptide, there were no reported yields of Val^{65}, Lys^{68} or Ile^{69}, any of which might indicate some form of chemical or photochemical modification. However, after the first three cycles, the yields of PTH-amino acid drops precipitously, which may also indicate some modification that results in a failure of PTH to couple to Val^{65}. A fourth crosslinked peptide, IDEFLATGK (Ile^{73}-Lys^{81}), was also identified (Figure 3.11). After an unusually high yield of the first amino acid, Ile^{73}, the yields of subsequent amino acids also drop considerably. In all, four peptides were tentatively identified as being crosslinked to oligo d(ATATATA), *via* Edman sequencing.

MALDI-TOF MS of the LIUV-crosslinked peptide-oligonucleotide complexes

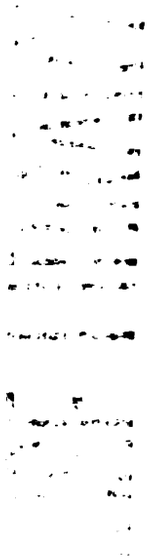
Results from MALDI-TOF MS of the LIUV crosslinked peptide-oligonucleotide complexes are shown in Figures 3.12-3.13. From the size of the peaks eluting from the HPLC column, and comparing these peaks with that of a known amount of irradiated d(ATATATA), it is estimated that about 1.0nmol of each of the crosslinks were recovered. The mass spectrum revealed three peaks that likely correspond to d(ATATATA) crosslinked with peptides Asn^{28}-Lys^{35} (expected monoisotopic MH^+ = 2998.9 Da), Asn^{28}-Tyr^{36} (expected monoisotopic MH^+ = 3161.9 Da) and Asn^{28}-Arg^{40} (expected monoisotopic MH^+ = 3666.2 Da). The second peptide was somewhat unusual, because it would have required cleavage after a tyrosine residue; it is possible, however, that there may have been some chymotryptic activity in the proteolytic digest. These crosslinked species are apparently not homogeneous, because in addition to the expected peak due to the d(ATATATA)-Asn^{28}-Arg^{40} crosslink, other peaks corresponding to losses of approximately 18, 26 and 45 mass units are also present (Figure 3.13), which implies that photochemical crosslinking results in a variety of species. At this point we are unable to explain the origin of these peaks, except that they are most likely due to a



series of photochemical reactions. We have indirect evidence that these multiple species are related to the same crosslinked peptide-oligonucleotide species, however. Both phosphodiesterase I and carboxypeptidase Y digestions of these crosslinks, which result in sequential losses of mononucleotides and amino acids, respectively, show that each of the resultant crosslinked species maintains the same heterogeneous distribution of masses (Figures 3.20-3.21, discussion below). The presence of peaks corresponding to the loss of 18 mass units in each of the three crosslinked species lends further evidence to the results of a loss of a water molecule from Ser^{30}, which had been revealed by Edman sequencing.

MALDI-TOF MS of the NPUV-crosslinked peptide-oligonucleotide complexes

The low yields (<100pmol, total) of the peptide-oligonucleotide conjugates from NPUV crosslinking make them fairly difficult to characterize by MALDI-TOF MS. A summary of the peaks found, along with tentative peptide assignments, can be found in Table 3.2. The mass spectra of five peptide-oligonucleotide complexes are shown in Figures 3.14-3.17. We believe that this is the first example in the literature where the peptides participating in NPUV protein-nucleic acid crosslinking have been identified at the peptide level. One of the peaks we found had a measured m/z of 2999, which corresponds almost exactly to that of d(ATATATA) crosslinked with Asn^{28}-Lys^{35} (predicted average MH^+ = 3000.2 Da, Figure 3.14). Moreover, another peak was detected having an m/z of 3162, which may correspond to d(ATATATA) crosslinked to the peptide Asn^{28}-Tyr^{36} (predicted average MH^+ = 3163.6 Da, Figure 3.14). As in the case of LIUV crosslinking, the cleavage at the Tyr^{36} might have been the result of some chymotryptic activity in the tryptic digest. The peak at m/z = 3137 apparently is a photoproduct of the d(ATATATA)-Asn^{28}-Tyr^{36} complex, and was also detected in the LIUV-crosslinked sample. The peak at m/z = 3017 is about 18 mass units higher than the d(ATATATA)-Asn^{28}-Lys^{35} conjugate, but we have no further evidence that it is a



photoproduct of this complex. We also detected a peak with an m/z of 2788, which might correspond to d(ATATATA) crosslinked to the peptide Tyr^{36}-Lys^{40} (predicted average MH^+ = 2789.6 Da). In addition, we also detected a peak with an observed m/z of 2666, which might correspond to a d(ATATATA)-Ser^{55}-Lys^{60} crosslink (predicted average MH^+ = 2665.9 Da). Another peak we detected had an observed m/z of 3219, which corresponds closely to that of d(ATATATA) crosslinked to the peptide Leu^{62}-Lys^{72} (predicted average MH^+ = 3216.0 Da). Finally, we detected a peak with an m/z = 3537, which might correspond to d(ATATATA) crosslinked to the peptide Ile^{69}-Lys^{81} (predicted average MH^+ = 3539.0 Da, Figure 3.17).

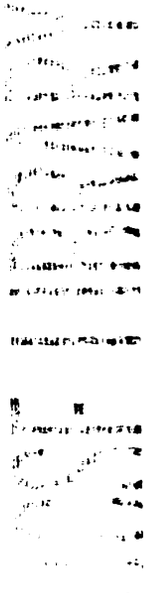
MALDI-TOF MS also revealed two peaks that had an m/z of 2408 and 2536. Interestingly, there are two equally likely explanations for these observations. First, these peaks could be thought of as d(ATATATA) possibly crosslinked to two very small peptides, RK (or KR) and KRK (predicted average MH^+ = 2406.7 and 2534.9 Da, respectively). While there are a number of regions in the protein where there are two consecutive basic residues, there is only one that has three, which is at the N-terminus, **SKRK**APQE... If these are indeed crosslinked peptide-oligonucleotide complexes, then we may have located an additional ssDNA binding region. This is entirely possible, as this basic region at the N-terminus should have few physical constraints and would be free to interact, at least transiently, with the ssDNA phosphate backbone. However, this peptide would had to have been cleaved after a serine residue, which is highly unlikely in a tryptic digest. Closer examination of the data revealed that these peaks may also be the relatively acidic peptides resulting from the tryptic digestion of uncrosslinked pol β. The peaks might be peptides Ala^5-Lys^{25} and Lys^4-Lys^{25} (predicted average MH^+ = 2391.7 and 2519.9 Da, respectively), plus an oxygen atom (+16 Da). It is possible that, under the conditions that we irradiated (O_2 was not excluded), that the Met^{18} residue was oxidized, but was not investigated any further. These results are corroborated by Edman

sequencing of the two peptides KAPQETL... and APQETL... (data not shown), and the fact that there are smaller peaks in the mass spectrum with masses that correspond to the unreacted peptides. Under the HPLC conditions used, these acidic peptides could conceivably coelute with peptide-oligonucleotide conjugates having the same net charge. This ambiguity could be resolved in future studies by either using a larger oligonucleotide in future crosslinking studies involving pol β, or performing a nuclease digestion of this species.

Usually, one uses a slightly larger oligonucleotide--one with more negative charges--so as to ensure that there would be sufficient separation between and acidic peptides and crosslinked peptide-oligonucleotide complexes. Our choice of the length of the oligonucleotide at the beginning of these experiments was directed more towards being able to assign the mass of the complexes accurately enough to determine the identities of any covalently bound peptides. As we have become more familiar with mass spectral characterization of synthetic oligonucleotides, however, the length of the oligonucleotide is becoming less of an analytical concern, and it is quite likely that larger oligonucleotides will be used in future studies.

In many cases, our Edman sequencing and MALDI-TOF MS results agree, but there are a number of significant differences. For instance, Edman sequencing clearly detected the tryptic peptide Ile^{73}-Lys^{81}, which could have been sequenced from the peptide itself or possibly Ile^{73}-Arg^{83} or Ile^{73}-Lys^{84}, but we didn't detect any of the corresponding crosslinked complexes using MALDI-TOF MS (predicted average MH^+ = 3097.4, 3366.8 and 3495.0 Da, respectively). The peaks that most closely corresponded to any of these masses was a series of peaks with m/z of 3340, 3348 and 3356. It is unknown if any of these peaks correspond to crosslinked species.

When the NPUV-crosslinked peptides are compared with the NMR data identifying amino acid residues that have undergone a ^1H chemical shift of >0.1ppm and



a ^{15}N chemical shift of >0.4ppm when complexed with p(dT)8[47], the proposed protein-oligonucleotide contact regions are in very good agreement (Figure 3.19). Nearly the entire binding region has been covered, with the exception of amino acid residues Lys^{41} and Ala^{47}. The Lys^{41} residue lies just outside the region (Asn^{28}-Arg^{40}) we have identified by both Edman sequencing and MALDI-TOF MS. In addition, MALDI-TOF MS detected a species with an m/z of 2748, which is 15 Da less than that of what would be predicted for a d(ATATATA)-Ala^{42}-Lys^{48} complex (predicted MH^+ = 2763.1 Da). While it would be tempting to conclude that this is, indeed, the a peptide-oligonucleotide complex which had undergone some type of photochemical reaction, the Edman sequencing results on this peptide were inconclusive. Therefore, we cannot conclude that we have found this "missing" peptide. There are several reasons why certain peptides "expected" to crosslink may have not been identified. Among them are the following:

Poor Yields. The yields of some particular peptide-oligonucleotide complexes were too low to detect by either MALDI-TOF MS or Edman sequencing, possibly because only relatively unreactive amino acids were accessbile to excited nucleobases.

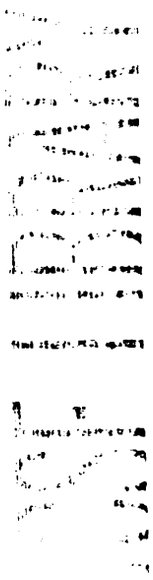
O_2 quenching. By not excluding molecular O_2 from our irradiations, we may have quenched some of the photoreactions at potentially crosslinkable sites.

Unstable crosslinks. Under the conditions used some of the crosslinks formed may have been chemically unstable to our chromatographic or mass spectrometric analysis.

Photomodification. Photocrosslinking reactions could have produced one or more species with molecular weights not equal to the sum of the peptide and the oligonucleotide, thus making identification much more difficult.

Carboxypeptidase Y and Phosphodiesterase I digestion of a LIUV-crosslinked complex

Figures 3.20 and 3.21 show the results of our initial attempts to digest away unreacted amino acids and nucleotides from the crosslinked complex. In the case of the carboxypeptidase Y digestion, we were able to sequence three amino acids from the C-terminus of the d(ATATATA)-Asn^{28}-Arg^{40} complex, revealing loss of masses that correspond to Arg^{40}, Tyr^{39} and Ala^{38} (Figure 3.20). The phosphodiesterase I digestion of the d(ATATATA)-Asn^{28}-Arg^{40} crosslinked species reveals sequential losses of adenosine-PO_4 and thymidine-PO_4 (Figure 3.21). However we have been unsuccessful to date in obtaining definitive mass spectrometric data determining points of contact of a protein-oligonucleotide complex at the amino acid level. It is possible that the enzymes were unable to penetrate far enough into the complex because of steric hindrance, or some chemical or photochemical modification that may have rendered the substrate unrecognizable by the enzyme.



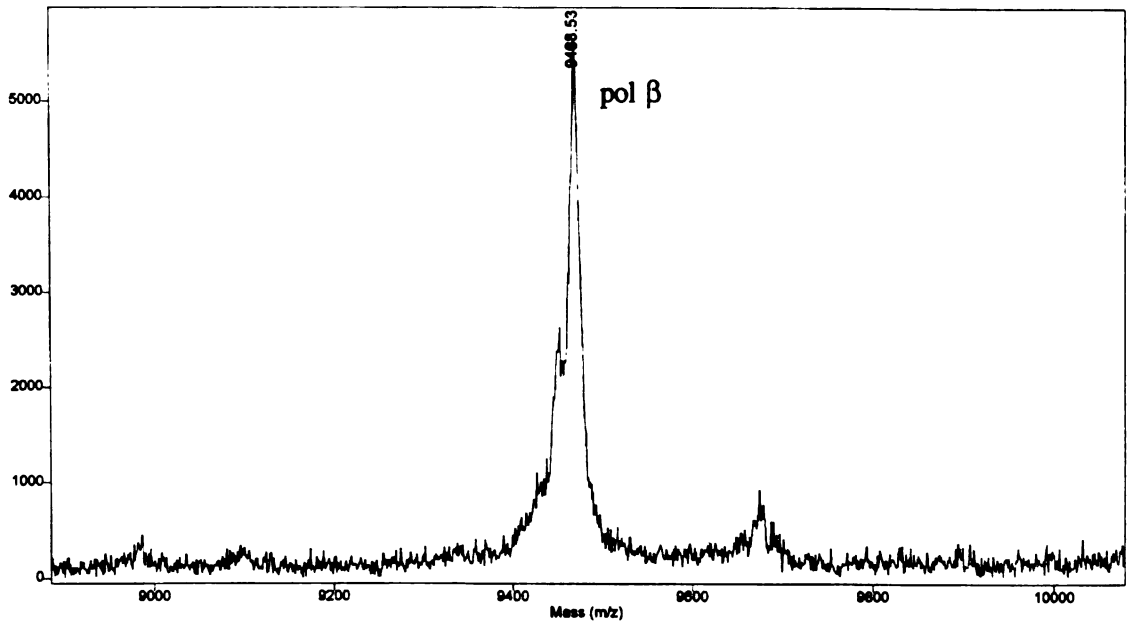


Figure 3.1 MALDI-TOF mass spectrum of the ssDNA-binding subunit of recombinant rat DNA pol β. The measured MH^+ of 9468 Da is 14 mass units too low given the amino acid sequence from reference[178].

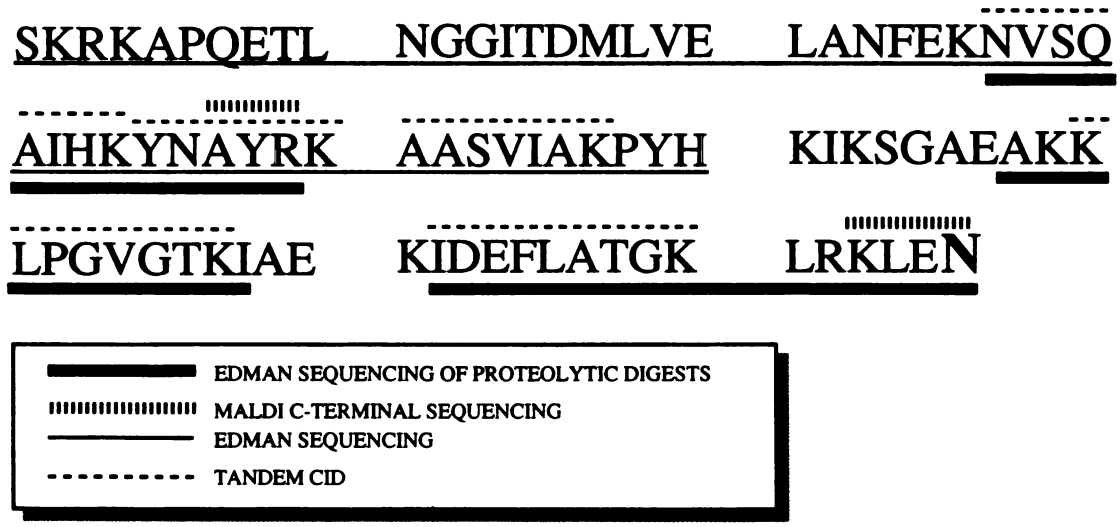


Figure 3.2 Primary sequence of the ss-DNA binding subunit of rat DNA polymerase β. See text for discussion of the C-terminal amino acid in boldface.

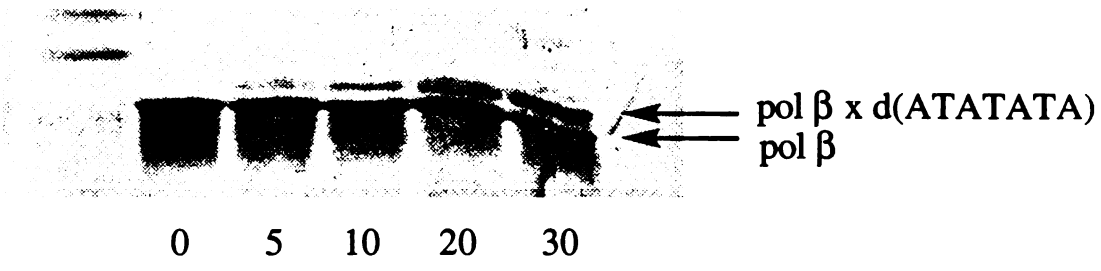


Figure 3.3 Denaturing SDS-PAGE of the LIUV-crosslinking of pol β with d(ATATATA)

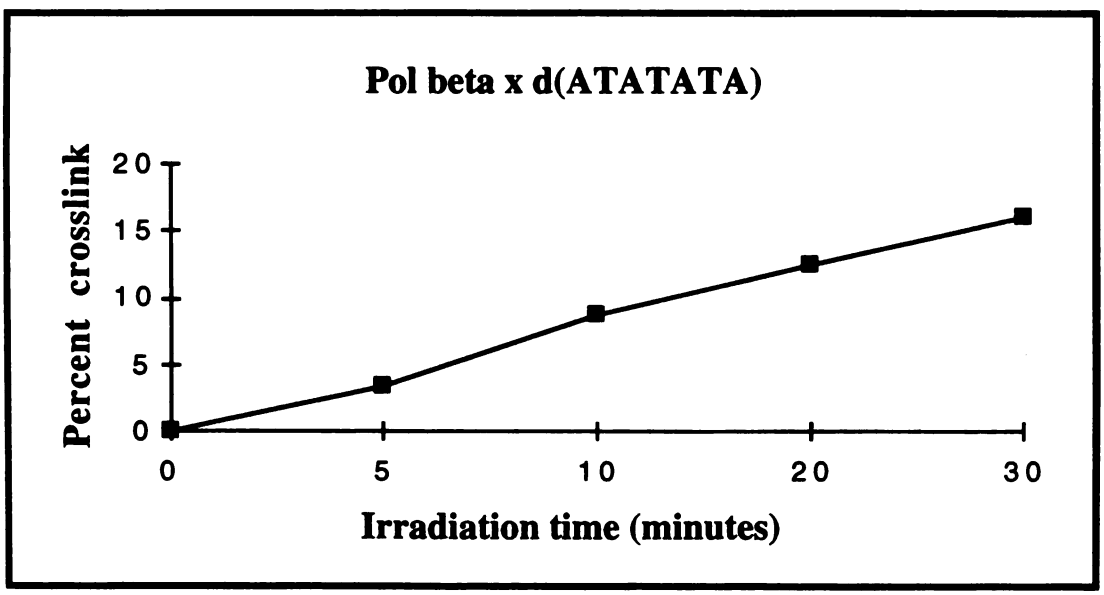


Figure 3.4 Yields of the LIUV-mediated pol β-d(ATATATA) crosslink.

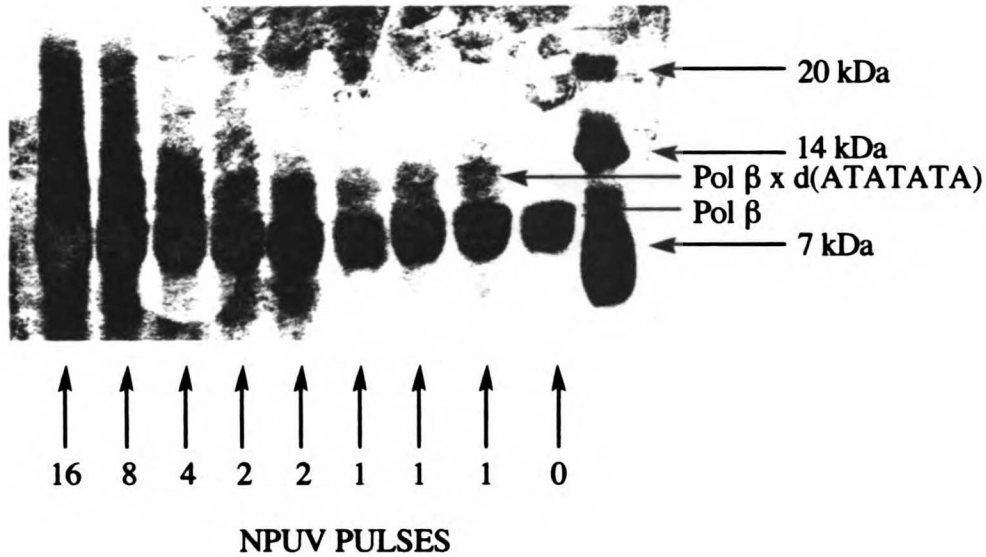


Figure 3.5 Denaturing SDS-PAGE of the pol β before and after NPUV crosslinking with 1, 2, 4, 8 and 16 pulses.

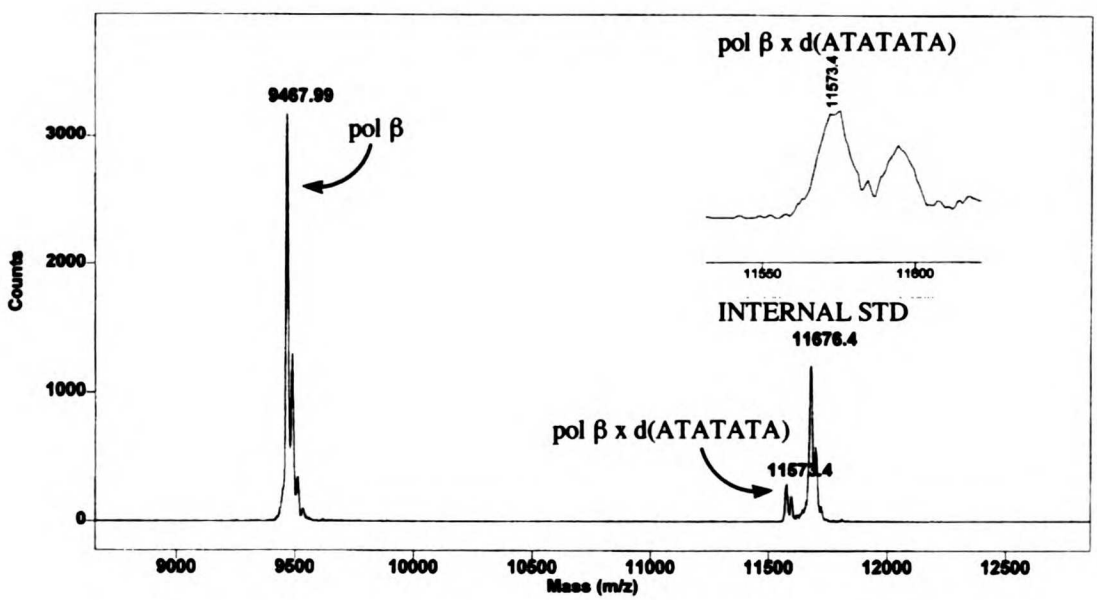


Figure 3.6 MALDI-TOF mass spectra of the pol β x d(ATATATA) irradiated mixture

Crosslinked Species	Predicted mass (avg)	Observed mass (avg)
Unreacted pol β	MH^+ = 9468.0 Da	MH^+ = 9468 ± 1 Da
Unreacted oligo d(ATATATA)	MH^+ = 2104.3 Da	MH^+ = 2104 ± 1 Da
pol β-d(ATATATA) complex	MH^+ = 11571.2 Da	MH^+ = 11573 ± 2 Da

Table 3.1 NPUV-crosslinked protein-oligonucleotide complexes analyzed by MALDI-TOF MS.

Crosslinked Species	Predicted mass (avg)	Observed mass (avg)
d(ATATATA) x Asn^{28}-Lys^{35}	MH^+ = 3000.3 Da	MH^+ = 2999 ± 2 Da
d(ATATATA) x Asn^{28}-Tyr^{36}*	MH^+ = 3163.5 Da	MH^+ = 3162 ± 2 Da
d(ATATATA) x Tyr^{36}-Arg^{40}	MH^+ = 2790.0 Da	MH^+ = 2788 ± 2 Da
d(ATATATA) x Ser^{55}-Lys^{60}#	MH^+ = 2665.9 Da	MH^+ = 2666 ± 2 Da
d(ATATATA) x Leu^{62}-Lys^{72}	MH^+ = 3216.6 Da	MH^+ = 3219 ± 2 Da
d(ATATATA) x Ile^{69}-Lys^{81}#	MH^+ = 3539.0 Da	MH^+ = 3537 ± 2 Da

Table 3.2 NPUV peptide-oligonucleotide crosslinks identified by MALDI-TOF MS.
* Likely a result of chymotryptic activity.
# Only partial sequences found by Edman degredation.

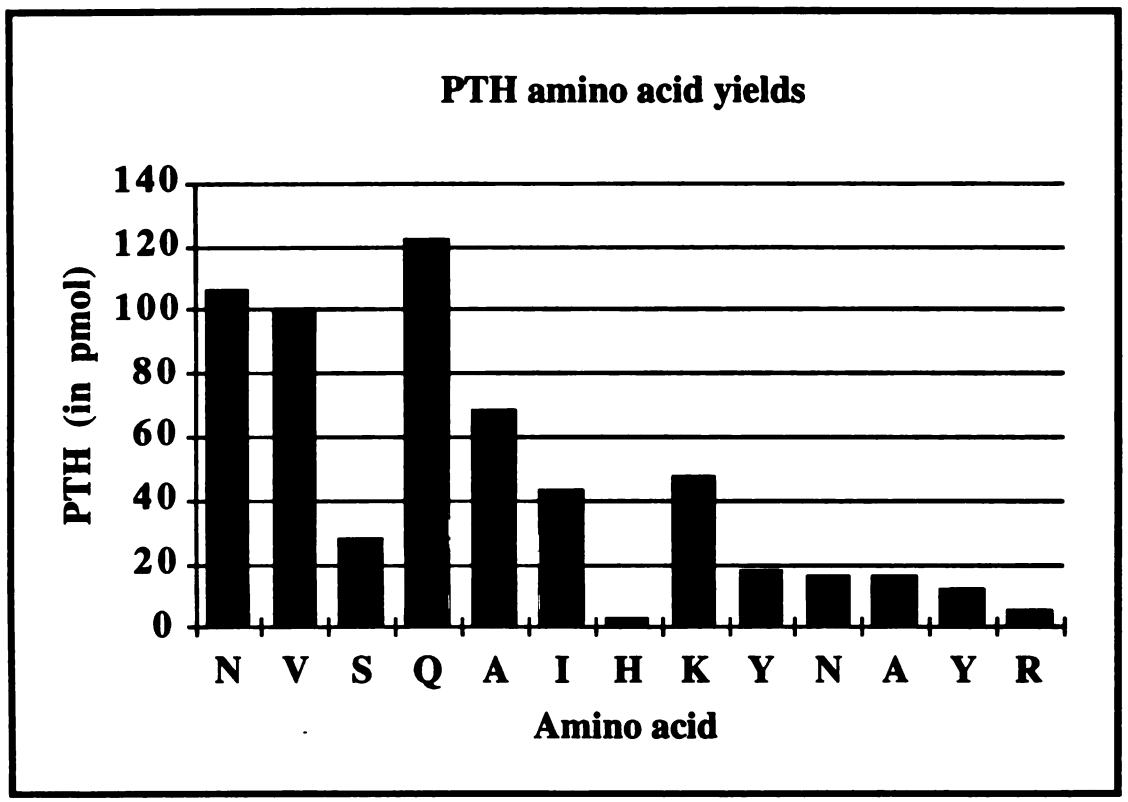
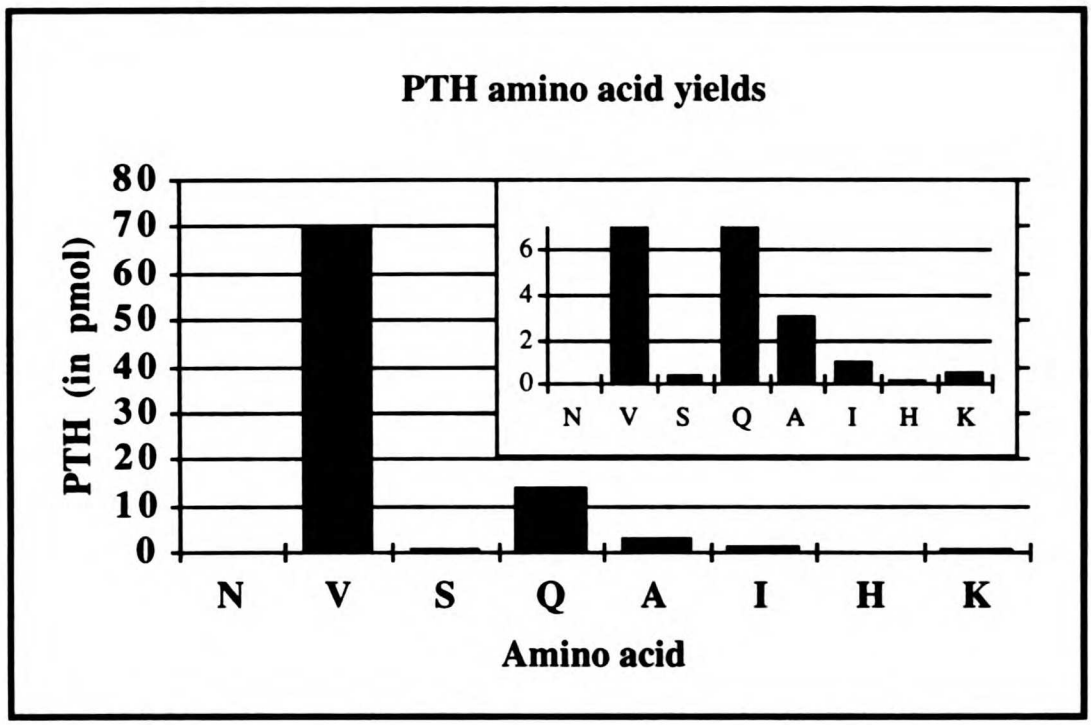
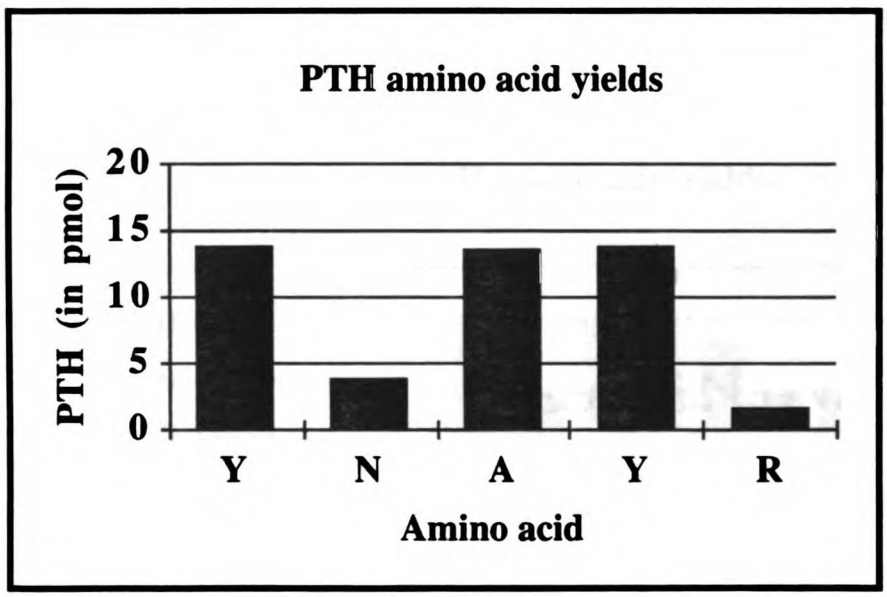
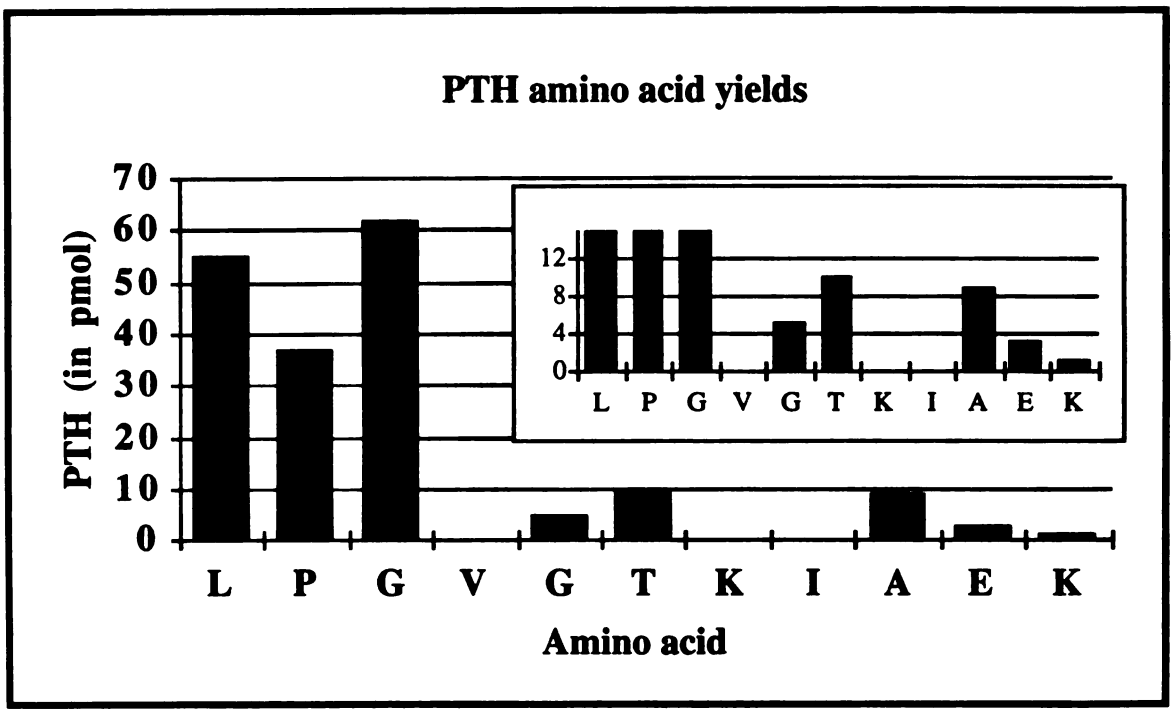


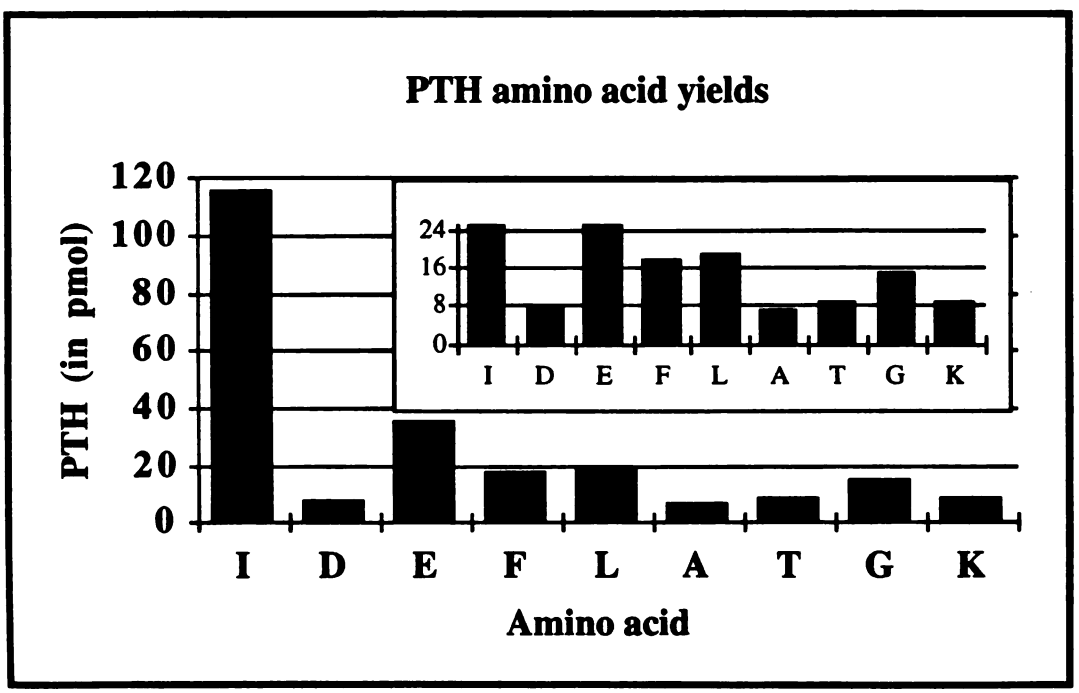
Figure 3.7 Edman sequencing of the LIUV-crosslinked peptide-oligonucleotide complex. See discussion for details.





Figures 3.8-3.9 PTH amino acid sequences. The PTH Asn^{28} is not reported because of a large unknown peak which coeluted with PTH-Asn.





Figures 3.10-3.11 Edman sequencing of NPUV-crosslinked peptide-oligonucleotide complexes

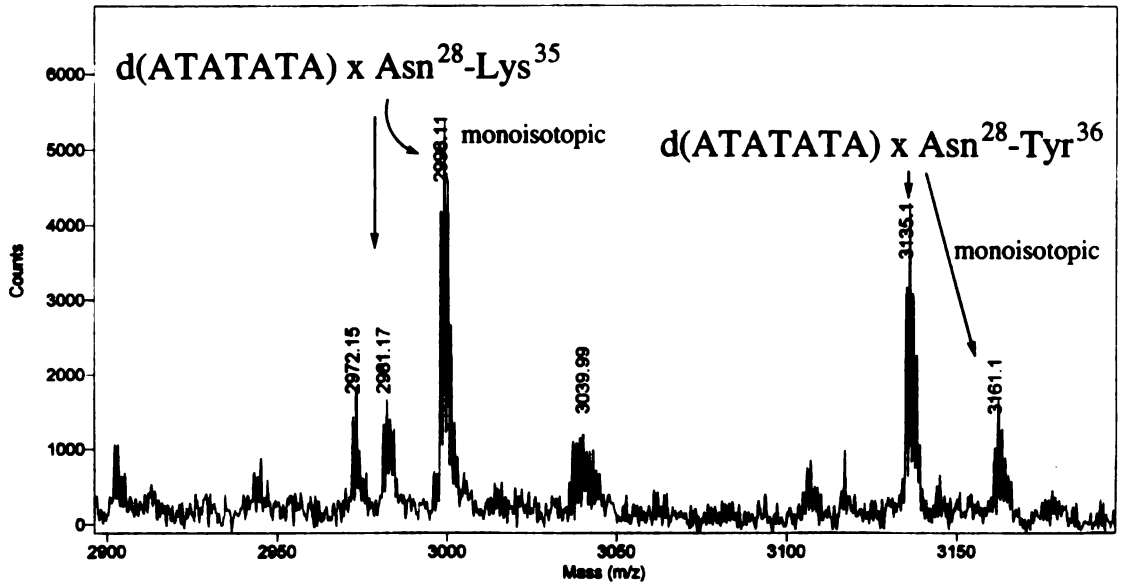


Figure 3.12 MALDI-TOF reflectron MS of peptide-nucleotide crosslinked species d(ATATATA)xAsn^{28}-Lys^{35} and d(ATATATA)xAsn^{28}-Tyr^{36}. The peaks labeled "monoisotopic" represent the species whose mass agrees with the sum of the peptide and the oligonucleotide.

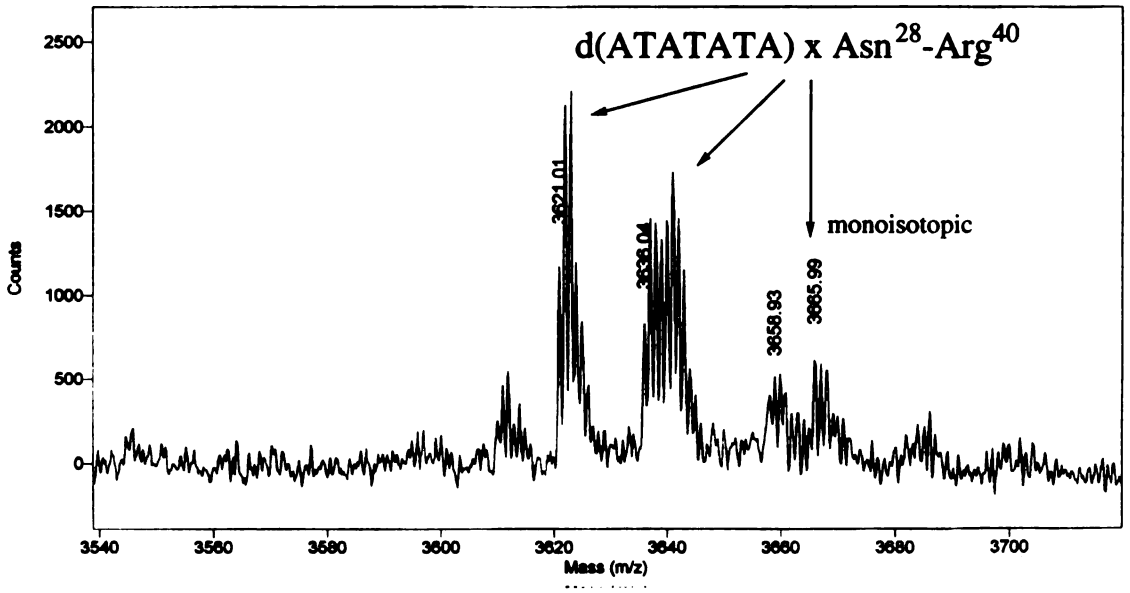
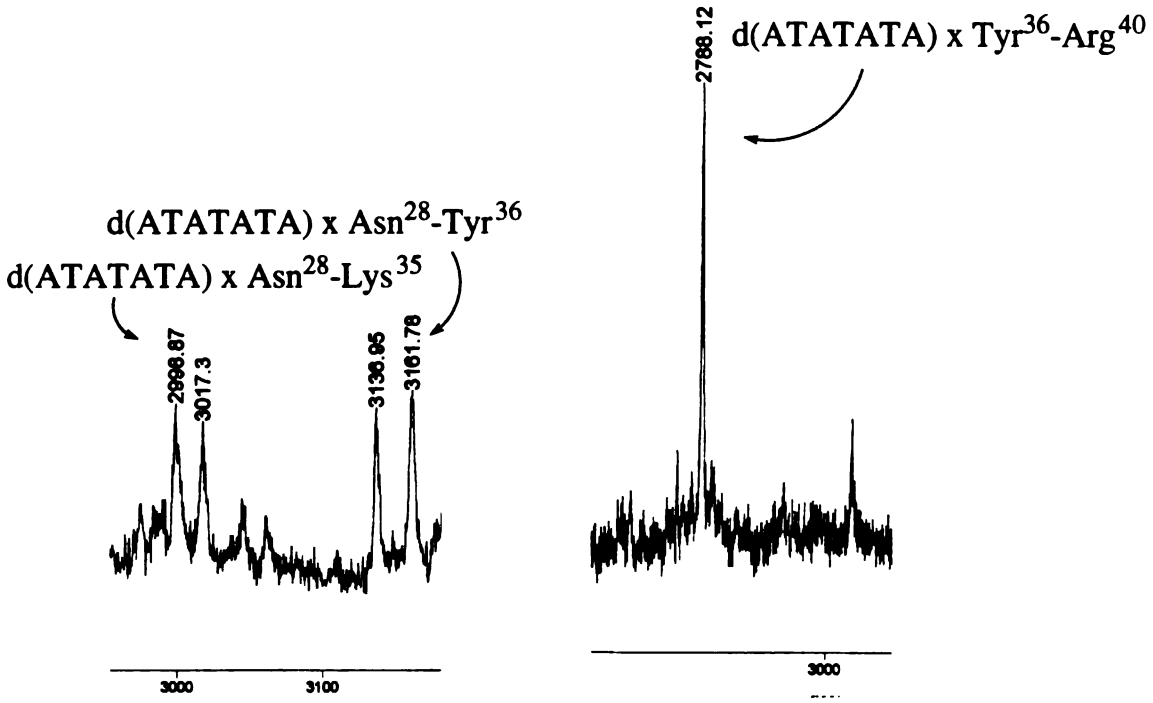
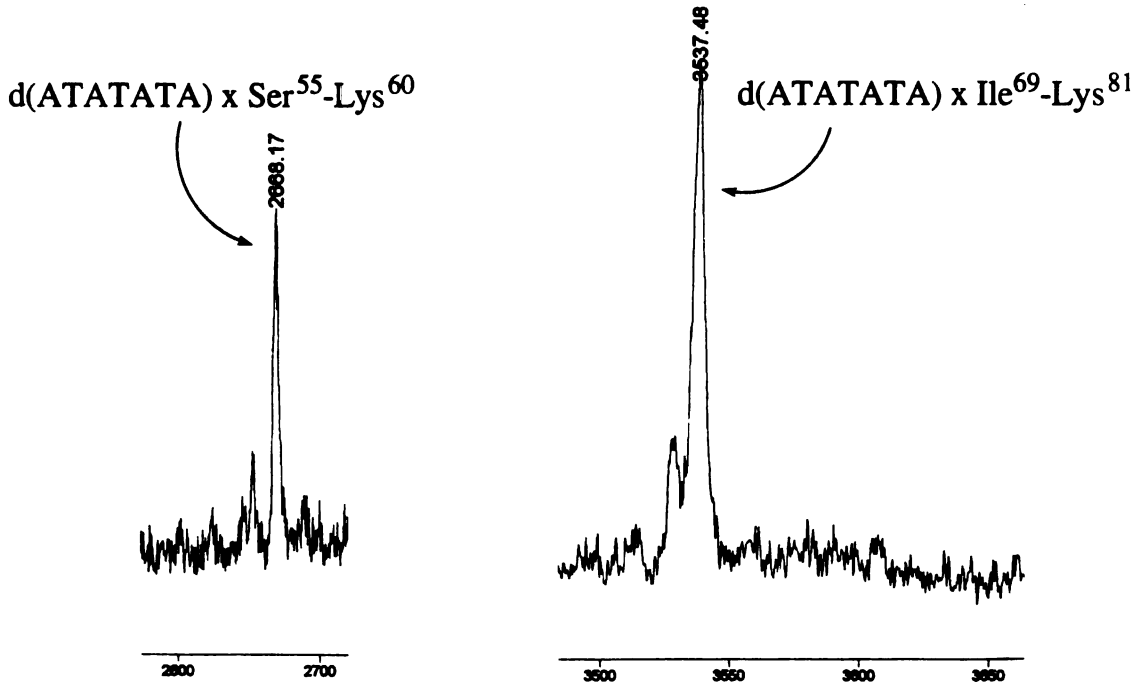
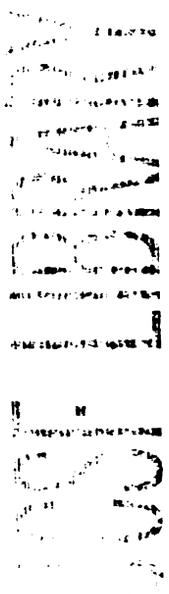


Figure 3.13 MALDI-TOF reflectron MS of the d(ATATATA)xAsn^{28}-Arg^{40} crosslink.





Figures 3.14-3.17 MALDI-TOF MS of NPUV-crosslinked peptide-oligonucleotide complexes



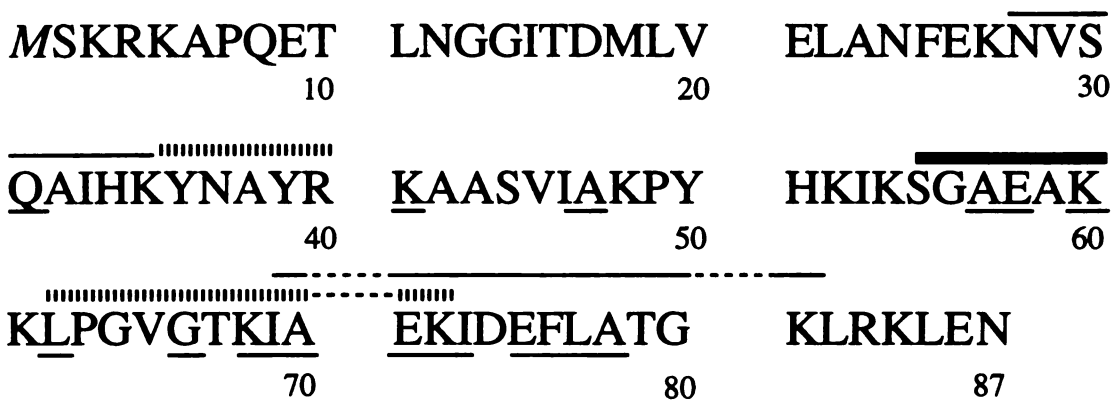


Figure 3.18 Primary sequence of pol β. The underlined residues indicate amino acids that have undergone a ^{1}H chemical shift of >0.1ppm and a ^{15}N chemical shift >0.5ppm when complexed with p(dT)8[47]. The lines overhead indicate tryptic peptide fragments tentatively identified crosslinked to oligo d(ATATATA) *via* NPUV light, as verified by MALDI-TOF mass spectrometry. The italicized Met^{1} is not in our protein; it serves to keep the numbering of the amino acids consistent with other reports.

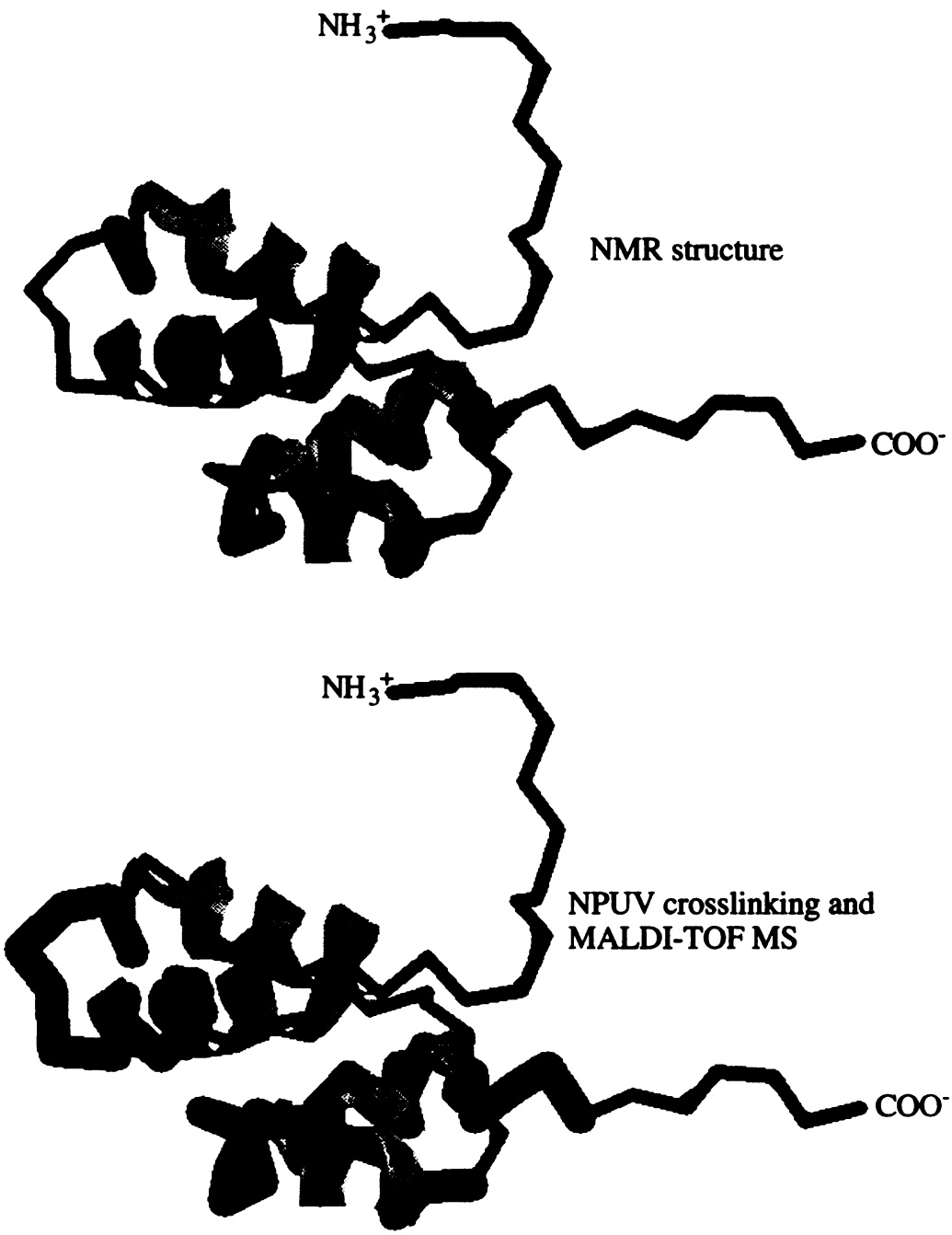


Figure 3.19 Comparison of the NMR solution structure[47] with MALDI-TOF MS sequences identified from NPUV crosslinking. The thin tube represents the peptide backbone; the thicker tube represents the peptide backbone along which are the amino acids believed to lie in the ssDNA-binding domain. The ribbons represent α-helices.

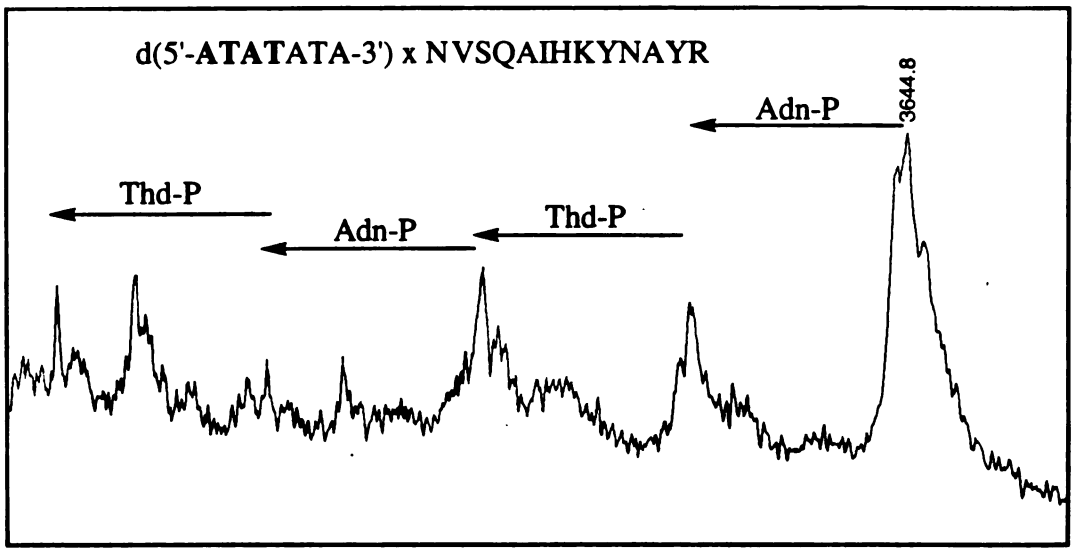


Figure 3.20 Phosphodiesterase I digestion of the LIUV-crosslinked d(ATATATA)-Asn^{28}-Arg^{40} complex.

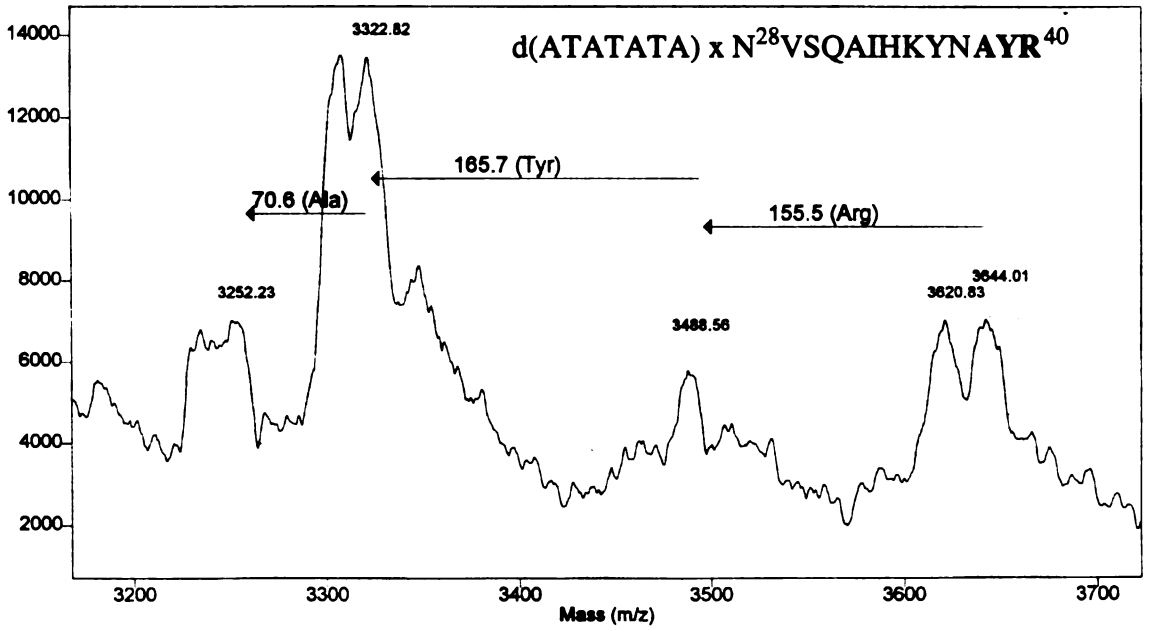
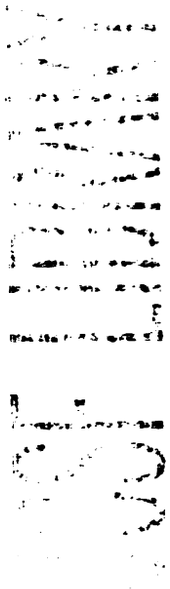


Figure 3.21 Carboxypeptidase Y sequencing of the LIUV-crosslinked d(ATATATA)-Asn^{28}-Arg^{40} complex.



3.4 CONCLUSIONS AND FUTURE PROSPECTS

Our results show that LIUV and NPUV crosslinking can be very useful methods for helping to determine the protein interface of protein-nucleic acid binding complexes. In our particular case, NPUV crosslinking appears to be superior in mapping out binding regions of the protein. The reasons for the differences between the NPUV and the LIUV crosslinking studies are unclear: it is possible that the LIUV crosslinking produces crosslinks in some areas that may preclude the formation of other covalent bonds between proximal contact points. It is also possible that the differences in the wavelengths of UV light used (254nm for LIUV and 266nm for NPUV) may have had an effect on crosslinking.

MALDI-TOF MS has been shown to be an excellent analytical tool for determining protein-oligonucleotide and peptide-oligonucleotide molecular weights. The accuracy of mass determination far exceeds that of SDS-PAGE. While MALDI-TOF MS and Edman sequencing appear to have approximately equal sensitivity, MALDI-TOF MS also has a number of advantages. For instance, MALDI-TOF MS can discriminate between a number of peptides that have the same N-terminus but different C-termini, as was the case in tryptic digests of our protein-oligonucleotide system. Another advantage is that there may be several peptide-oligonucleotide complexes in a single fraction eluting off an HPLC, which could potentially lead to ambiguous results in the interpretation of the data obtained by Edman sequencing. MALDI-TOF MS is not encumbered by modifications that may make the peptides unsuitable for PTH-amino acid coupling. However, the high mass accuracy of MALDI-TOF MS can, at times, be a disadvantage, as it appears that photochemical crosslinking results in species with molecular weights different than that of the predicted sum of the peptide and the oligonucleotide.

We have also shown that we can obtain more precise information regarding the specific amino acid-nucleic acid contact points at the amino acid level. By using non-

radiolabeled d(ATATATA) as the oligonucleotide we are now in a position to sequence the oligonucleotide itself using either chemical or enzymatic nucleases. We have shown that we can digest irradiated d(ATATATA)-peptide on the MALDI sample plate using phosphodiesterase I, thus producing a ragged end at the 5'-side of the oligonucleotide. This will allow us to sequence the oligonucleotide *in situ*. MALDI-TOF MS also allows us to sequence the C-terminus of a peptide *in situ* using carboxypeptidase Y. We have obtained partial sequence information on both the peptide and the oligonucleotide portion of crosslinked peptide-modified oligonucleotides resulting from LIUV-crosslinking of pol β with d(ATATATA). Unfortunately, the low yields of NPUV-crosslinked complexes prohibited such detailed analyses.

Future work should be directed at expanding on the preliminary results obtained from the *in situ* protease and nuclease digestions of peptide-oligonucleotide complexes. By developing these protocols on LIUV-crosslinked materials, which are easier to obtain and appear to result in much higher yields of photoproducts, one should be able to apply those protocols to NPUV-crosslinked systems.

REFERENCES

1. J. A. McClarin, C. A. Frederick, B. C. Wang, P. Greene, H. W. Boyer, J. Grable, and J. M. Rosenberg, Structure of the DNA-EcoR I endonuclease recognition complex at 3 A resolution, *Science* 234, 1526-41, 1986.

2. D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, and B. W. Matthews, The molecular basis of DNA-protein recognition inferred from the structure of cro repressor, *Nature* 298, 718-23, 1982.

3. I. T. Weber and T. A. Steitz, Model of specific complex between catabolite gene activator protein and B-DNA suggested by electrostatic complementarity, *Proceedings of the National Academy of Sciences of the United States of America* 81, 3973-7, 1984.

4. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison, Recognition of a DNA operator by the repressor of phage 434: a view at high resolution, *Science* 242, 899-907, 1988.

5. S. R. Jordan and C. O. Pabo, Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions, *Science* 242, 893-9, 1988.

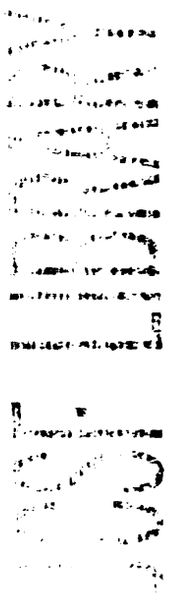
6. M. E. Churchill and M. Suzuki, 'SPKK' motifs prefer to bind to DNA at A/T-rich sites, *Embo Journal* 8, 4189-95, 1989.

7. M. Suzuki, SPKK, a new nucleic acid-binding unit of protein found in histone, *Embo Journal* 8, 797-804, 1989.

8. J. E. Anderson, M. Ptashne, and S. C. Harrison, Structure of the repressor-operator complex of bacteriophage 434, *Nature* 326, 846-52, 1987.

9. C. Wolberger, Y. C. Dong, M. Ptashne, and S. C. Harrison, Structure of a phage 434 Cro/DNA complex, *Nature* 335, 789-95, 1988.

10. K. L. Borden, M. N. Boddy, J. Lally, N. J. O'Reilly, S. Martin, K. Howe, E. Solomon, and P. S. Freemont, The solution structure of the RING finger domain from the



acute promyelocytic leukaemia proto-oncoprotein PML, *Embo Journal* 14, 1532-41, 1995.

11. G. M. Clore, A. Bax, J. G. Omichinski, and A. M. Gronenborn, Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA, *Structure* 2, 89-94, 1994.

12. K. Ogata, S. Morikawa, H. Nakamura, A. Sekikawa, T. Inoue, H. Kanai, A. Sarai, S. Ishii, and Y. Nishimura, Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices, *Cell* 79, 639-48, 1994.

13. Y. Cho, S. Gorina, P. D. Jeffrey, and N. P. Pavletich, Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations, *Science* 265, 346-55, 1994.

14. L. Fairall, J. W. Schwabe, L. Chapman, J. T. Finch, and D. Rhodes, The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition, *Nature* 366, 483-7, 1993.

15. H. Baumann, K. Paulsen, H. Kovacs, H. Berglund, A. P. Wright, J. A. Gustafsson, and T. Hard, Refined solution structure of the glucocorticoid receptor DNA-binding domain, *Biochemistry* 32, 13463-71, 1993.

16. N. P. Pavletich and C. O. Pabo, Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers, *Science* 261, 1701-7, 1993.

17. J. G. Omichinski, G. M. Clore, O. Schaad, G. Felsenfeld, C. Trainor, E. Appella, S. J. Stahl, and A. M. Gronenborn, NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1, *Science* 261, 438-46, 1993.

18. D. Neuhaus, Y. Nakaseko, J. W. Schwabe, and A. Klug, Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional 1H nuclear magnetic resonance spectroscopy. A zinc-finger structure with a third strand of beta-sheet, *Journal of Molecular Biology* 228, 637-51, 1992.

19. F. F. Damberger, J. G. Pelton, C. Liu, H. Cho, C. J. Harrison, H. C. Nelson, and D. E. Wemmer, Refined solution structure and dynamics of the DNA-binding domain of the heat shock factor from Kluyveromyces lactis, *Journal of Molecular Biology* 254, 704-19, 1995.

20. M. H. Werner, M. Clore, C. L. Fisher, R. J. Fisher, L. Trinh, J. Shiloach, and A. M. Gronenborn, The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation, *Cell* 83, 761-71, 1995.

21. X. Qiu, C. L. Verlinde, S. Zhang, M. P. Schmitt, R. K. Holmes, and W. G. Hol, Three-dimensional structure of the diphtheria toxin repressor in complex with divalent cation co-repressors, *Structure* 3, 87-100, 1995.

22. E. H. Morita, M. Shirakawa, F. Hayashi, M. Imagawa, and Y. Kyogoku, Structure of the Oct-3 POU-homeodomain in solution, as determined by triple resonance heteronuclear multidimensional NMR spectroscopy, *Protein Science* 4, 729-39, 1995.

23. W. Xu, M. A. Rould, S. Jun, C. Desplan, and C. O. Pabo, Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations, *Cell* 80, 639-50, 1995.

24. M. Delarue, A. Poterszman, S. Nikonov, M. Garber, D. Moras, and J. C. Thierry, Crystal structure of a prokaryotic aspartyl tRNA-synthetase, *Embo Journal* 13, 3219-29, 1994.

25. M. A. Schumacher, K. Y. Choi, H. Zalkin, and R. G. Brennan, Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices, *Science* 266, 763-70, 1994.

26. H. Zhang, D. Zhao, M. Revington, W. Lee, X. Jia, C. Arrowsmith, and O. Jardetzky, The solution structures of the trp repressor-operator DNA complex, *Journal of Molecular Biology* 238, 592-614, 1994.

27. V. P. Chuprina, J. A. Rullmann, R. M. Lamerichs, J. H. van Boom, R. Boelens, and R. Kaptein, Structure of the complex of lac repressor headpiece and an 11 base-pair

half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics, *Journal of Molecular Biology* 234, 446-62, 1993.

28. K. Ogata, H. Kanai, T. Inoue, A. Sekikawa, M. Sasaki, A. Nagadoi, A. Sarai, S. Ishii, and Y. Nishimura, Solution structures of Myb DNA-binding domain and its complex with DNA, *Nucleic Acids Symposium Series* 29, 201-2, 1993.

29. L. J. Beamer and C. O. Pabo, Refined 1.8 A crystal structure of the lambda repressor-operator complex, *Journal of Molecular Biology* 227, 177-96, 1992.

30. P. C. Ma, M. A. Rould, H. Weintraub, and C. O. Pabo, Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation, *Cell* 77, 451-9, 1994.

31. T. E. Ellenberger, C. J. Brandl, K. Struhl, and S. C. Harrison, The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex, *Cell* 71, 1223-37, 1992.

32. P. Konig and T. J. Richmond, The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility, *Journal of Molecular Biology* 233, 139-54, 1993.

33. K. Katayanagi, M. Miyagawa, M. Matsushima, M. Ishikawa, S. Kanaya, H. Nakamura, M. Ikehara, T. Matsuzaki, and K. Morikawa, Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution, *Journal of Molecular Biology* 223, 1029-52, 1992.

34. R. C. Robinson, L. M. Grey, D. Staunton, D. I. Stuart, J. K. Heath, and E. Y. Jones, The crystal structure of murine leukemia inhibitory factor, *Annals of the New York Academy of Sciences* 762, 179-87; discussion 187-8, 1995.

35. C. D. Mol, A. S. Arvai, R. J. Sanderson, G. Slupphaug, B. Kavli, H. E. Krokan, D. W. Mosbaugh, and J. A. Tainer, Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA, *Cell* 82, 701-8, 1995.

36. Y. Shamoo, A. M. Friedman, M. R. Parsons, W. H. Konigsberg, and T. A. Steitz, Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA, *Nature* 376, 362-6, 1995.

37. L. Pellegrini, S. Tan, and T. J. Richmond, Structure of serum response factor core bound to DNA, *Nature* 376, 490-8, 1995.

38. L. S. Beese, V. Derbyshire, and T. A. Steitz, Structure of DNA polymerase I Klenow fragment bound to duplex DNA, *Science* 260, 352-5, 1993.

39. L. S. Beese, J. M. Friedman, and T. A. Steitz, Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate, *Biochemistry* 32, 14095-101, 1993.

40. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, and T. A. Steitz, Cocrystal structure of an editing complex of Klenow fragment with DNA, *Proceedings of the National Academy of Sciences of the United States of America* 85, 8924-8, 1988.

41. F. K. Winkler, D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petratos, and K. S. Wilson, The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments, *Embo Journal* 12, 1781-95, 1993.

42. D. Suck, A. Lahm, and C. Oefner, Structure refined to 2A of a nicked DNA octanucleotide complex with DNase I, *Nature* 332, 464-8, 1988.

43. D. B. Nikolov, S. H. Hu, J. Lin, A. Gasch, A. Hoffmann, M. Horikoshi, N. H. Chua, R. G. Roeder, and S. K. Burley, Crystal structure of TFIID TATA-box binding protein, *Nature* 360, 40-6, 1992.

44. H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, and J. Kraut, Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP, *Science* 264, 1891-903, 1994.

45. M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, and J. Kraut, Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism, *Science* 264, 1930-5, 1994.

46. J. F. Davies, R. J. Almassy, Z. Hostomska, R. A. Ferre, and Z. Hostomsky, 2.3 A crystal structure of the catalytic domain of DNA polymerase beta, *Cell* 76, 1123-33, 1994.

47. D. J. Liu, R. Prasad, S. H. Wilson, E. F. Derose, and G. P. Mullen, Three-dimensional solution structure of the N-terminal domain of DNA polymerase beta and mapping of the ssDNA interaction interface, *Biochemistry* 35, 6188-6200, 1996.

48. R. M. Knegtel, R. H. Fogh, G. Ottleben, H. Ruterjans, P. Dumoulin, M. Schnarr, R. Boelens, and R. Kaptein, A model for the LexA repressor DNA complex, *Proteins* 21, 226-36, 1995.

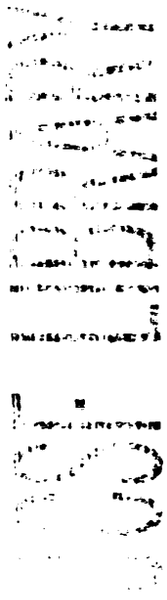
49. M. H. Werner, J. R. Huth, A. M. Gronenborn, and G. M. Clore, Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex, *Cell* 81, 705-14, 1995.

50. H. Baumann, S. Knapp, A. Karshikoff, R. Ladenstein, and T. Hard, DNA-binding surface of the Sso7d protein from Sulfolobus solfataricus, *Journal of Molecular Biology* 247, 840-6, 1995.

51. S. Morikawa, K. Ogata, A. Sekikawa, A. Sarai, S. Ishii, Y. Nishimura, and H. Nakamura, Determination of the NMR solution structure of a specific DNA complex of the Myb DNA-binding domain, *Journal of Biomolecular NMR* 6, 294-305, 1995.

52. P. W. Howe, K. Nagai, D. Neuhaus, and G. Varani, NMR studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein, *Embo Journal* 13, 3873-81, 1994.

53. M. D. Shetlar, Cross-linking of proteins to nucleic acids by ultraviolet light, in *Photochemical and Photobiological Reviews*, vol. 5, 1980, pp. 105-197.



54. I. Saito and H. Sugiyama, Photoreactions of nucleic acids and their constituents with amino acids and related compounds, in *Bioorganic Photochemistry*, vol. 1, H. Morrison, Ed. New York: Wiley and Sons, 1990, pp. 317-340.

55. K. Meisenheimer and T. Koch, Photocross-linking of nucleic acids to associated proteins, *Critical Reviews in Biochemistry and Molecular Biology* 32, 101-140, 1997.

56. A. J. Varghese, Photochemical addition of amino acids and related compounds to nucleic acid substituents, in *Aging, Carcinogenesis and Radiation Biology*, K. C. Smith, Ed. New York: Prenum Press, 1976, pp. 207-223.

57. H. Steinmaus, I. Rosenthal, and D. Elad, Photochemical and gamma-ray-induced reactions of purines and purine nucleosides with 2-propanol, *Journal of the American Chemical Society* 91, 4921-4923, 1969.

58. H. Steinmaus, I. Rosenthal, and D. Elad, Light- and -ray-induced reactions of purines and purine nucleosides with alcohols, *Journal of Organic Chemistry* 36, 3594-8, 1971.

59. J. S. Connolly and H. Linschitz, Photoaddition of alcohols to purine, *Photochemistry and Photobiology* 7, 791-806, 1968.

60. J. L. Fourrey and P. Jouin, Photoaddition of alcohols to methyl-4-pyrimidinones, *Tetrahedron Letters* 20, 951-954, 1977.

61. M. D. Shetlar, Photochemical and free radical initiated reactions of 1,3-dimethylthymine with isopropanol, *Photochemistry and Photobiology* 29, 253-9, 1979.

62. M. D. Shetlar, Photochemical and free radical initiated reactions of 1,3-dimethylthymine with ethanol, *Photochemistry and Photobiology* 32, 587-592, 1980.

63. M. D. Shetlar, J. Christensen, and K. Hom, Photochemical addition of amino acids and peptides to DNA, *Photochemistry and Photobiology* 39, 125-33, 1984.

64. M. D. Shetlar, K. Hom, J. Carbone, D. Moy, E. Steady, and M. Watanabe, Photochemical addition of amino acids and peptides to homopolyribonucleotides of the major DNA bases, *Photochemistry and Photobiology* 39, 135-40, 1984.

65. M. D. Shetlar, J. Carbone, E. Steady, and K. Hom, Photochemical addition of amino acids and peptides to polyuridylic acid, *Photochemistry and Photobiology* 39, 141-4, 1984.

66. M. D. Shetlar and K. Hom, Mixed products of thymine and cysteine produced by direct and acetone-sensitized photoreactions, *Photochemistry and Photobiology* 45, 703-12, 1987.

67. B. J. Hicke, M. C. Willis, T. H. Koch, and T. R. Cech, Telomeric protein-DNA point contacts identified by photo-cross-linking using 5-bromodeoxyuridine [published erratum appears in Biochemistry 1994 Jun 21;33(24):7744], *Biochemistry* 33, 3364-73, 1994.

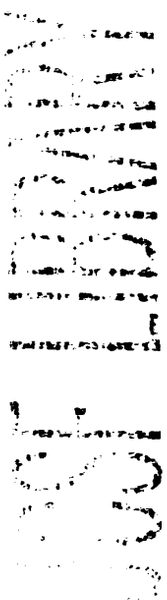
68. X. Huang, D. F. Shullenberger, and E. C. Long, Aromatic stacking and bending of the DNA helix by the individual repeat units of the carboxy-terminal domain of RNA polymerase II, *Biochemical and Biophysical Research Communications* 198, 712-9, 1994.

69. A. A. Shaw, A. M. Falick, and M. D. Shetlar, Photoreactions of thymine and thymidine with N-acetyltyrosine, *Biochemistry* 31, 10976-83, 1992.

70. M. D. Shetlar, H. N. Schott, H. G. Martinson, and E. T. Lin, Formation of thymine-lysine adducts in irradiated DNA-lysine systems, *Biochemical and Biophysical Research Communications* 66, 88-93, 1975.

71. I. Saito, H. Sugiyama, and T. Matsuura, Photoreaction of thymidine with alkylamines. Application to selective removal of thymine from DNA, *Journal of the American Chemical Society* 105, 956-962, 1983.

72. I. Saito, H. Sugiyama, N. Furukawa, and T. Matsuura, Photoreaction of thymidine with primary amines. Application to specific modification of DNA, *Nucleic Acids Symposium Series* 10, 61-4, 1981.



73. I. Saito and T. Matsuura, Chemical aspects of UV-induced cross-linking of proteins to nucleic acids. Photoreactions with lysine and tryptophan, *Accounts in Chemical Research* 18, 134-141, 1985.

74. M. D. Shetlar, K. Hom, S. Distefano, K. Ekpenyong, and J. Yang, Photochemical reactions of cytosine and 5-methylcytosine with methylamine and n-butylamine, *Photochemistry and Photobiology* 47, 779-86, 1988.

75. T. M. Dietz and T. H. Koch, Photochemical coupling of 5-bromouracil to tryptophan, tyrosine and histidine, peptide-like derivatives in aqueous fluid solution, *Photochemistry and Photobiology* 46, 971-8, 1987.

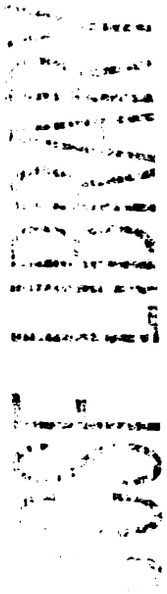
76. T. M. Dietz, R. J. von Trebra, B. J. Swanson, and T. H. Koch, Photochemical coupling of 5-bromouracil (BU) to a peptide linkage. A model for BU-DNA protein photocrosslinking, *Journal of the American Chemical Society* 109, 1793-7, 1987.

77. C. Norris, K. Meisenheimer, and T. Koch, Mechanistic studies relevant to bromouridine-enhanced nucleoprotein photocrosslinking: possible involvement of an excited tyrosine residue of the protein, *Photochemistry and Photobiology* 65, 201-207, 1997.

78. M. Ikeda and T. Kusaka, Characterization of the DNA-binding state of the rat uterine estrogen receptor by UV cross-linking, *Biochemistry and Molecular Biology International* 33, 447-56, 1994.

79. M. Katouzian-Safadi and M. Charlier, Identification of the DNA-interacting sites of proteins: microsequencing of the peptides cross-linked to 5-bromouracil substituted DNA, *Biochimie* 76, 129-32, 1994.

80. K. L. Wick and K. S. Matthews, Interactions between lac repressor protein and site-specific bromodeoxyuridine-substituted operator DNA. Ultraviolet footprinting and protein-DNA cross-link formation, *Journal of Biological Chemistry* 266, 6106-12, 1991.



81. Q. Dong, E. E. Blatter, Y. W. Ebright, K. Bister, and R. H. Ebright, Identification of amino acid-base contacts in the Myc-DNA complex by site-specific bromouracil mediated photocrosslinking, *Embo Journal* 13, 200-4, 1994.

82. I. Ahmad and D. N. Rao, Interaction of EcoP15I DNA methyltransferase with oligonucleotides containing the asymmetric sequence 5'-CAGCAG-3', *Journal of Molecular Biology* 242, 378-88, 1994.

83. M. Katouzian-Safadi, B. Blazy, and M. Charlier, Photochemical cross-linking of the cyclic adenosine 3',5' monophosphate receptor protein to Escherichia coli 5-bromouracil-substituted DNA. Role of the effectors, *Photochemistry and Photobiology* 53, 611-6, 1991.

84. E. E. Blatter, Y. W. Ebright, and R. H. Ebright, Identification of an amino acid-base contact in the GCN4-DNA complex by bromouracil-mediated photocrosslinking, *Nature* 359, 650-2, 1992.

85. J. M. Gott, M. C. Willis, T. H. Koch, and O. C. Uhlenbeck, A specific, UV-induced RNA-protein cross-link using 5-bromouridine-substituted RNA, *Biochemistry* 30, 6290-5, 1991.

86. M. Katouzian-Safadi, B. Laine, F. Chartier, J. Y. Cremet, D. Belaiche, P. Sautiere, and M. Charlier, Determination of the DNA-interacting region of the archaebacterial chromosomal protein MC1. Photocrosslinks with 5-bromouracil-substituted DNA, *Nucleic Acids Research* 19, 4937-41, 1991.

87. P. A. Boulanger, S. K. Yoshinaga, and A. J. Berk, DNA-binding properties and characterization of human transcription factor TFIIIC2, *Journal of Biological Chemistry* 262, 15098-105, 1987.

88. J. Blanco, H. Kimura, and G. C. Mueller, A method for detecting protein-DNA interactions at sites of chromatin replication, *Analytical Biochemistry* 163, 537-45, 1987.

89. T. M. Dietz and T. H. Koch, Photochemical reduction of 5-bromouracil by cysteine derivatives and coupling of 5-bromouracil to cystine derivatives, *Photochemistry and Photobiology* 49, 121-9, 1989.

90. M. D. Shetlar, R. B. Rose, K. Hom, and A. A. Shaw, Ring opening photoreactions of 5-bromouracil and 5-bromo-2'-deoxyuridine with selected alkylamines, *Photochemistry and Photobiology* 53, 595-609, 1991.

91. H. R. Morris, M. Panico, M. Barber, R. S. Bordoli, R. D. Sedgwick, and A. Tyler, Fast atom bombardment: a new mass spectrometric method for peptide sequence analysis, *Biochemical and Biophysical Research Communications* 101, 623-31, 1981.

92. W. Aberth and A. L. Burlingame, Comparison of three geometries for a cesium primary beam liquid secondary ion mass spectrometry source, *Analytical Chemistry* 56, 2915-8, 1984.

93. M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson, and M. B. Alice, Gas phase microions, *Journal of Chemical Physics* 49, 2240, 1968.

94. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science* 246, 64-71, 1989.

95. C. M. Whitehouse, R. N. Dreyer, M. Yamashita, and J. B. Fenn, Electrospray interface for liquid chromatographs and mass spectrometers, *Analytical Chemistry* 57, 675-9, 1985.

96. M. Karas and F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Analytical Chemistry* 60, 2299-301, 1988.

97. F. Hillenkamp, M. Karas, R. C. Beavis, and B. T. Chait, Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers, *Analytical Chemistry* 63, 1193A-1203A, 1991.

98. C. G. Edmonds and R. D. Smith, Electrospray ionization mass spectrometry, *Methods in Enzymology* 193, 412-31, 1990.

99. R. W. Nelson, D. Dogruel, and P. Williams, Mass determination of human immunoglobulin IgM using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry [see comments], *Rapid Communications in Mass Spectrometry* 8, 627-31, 1994.

100. F. C. Walls, M. A. Baldwin, A. M. Falick, B. W. Gibson, S. Kaur, D. A. Maltby, B. L. Gillece-Castro, K. F. Medzihradsky, S. Evans, and A. L. Burlingame, *Biological Mass Spectrometry*. Amsterdam: Elsevier, 1990.

101. K. Biemann, Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation, *Methods in Enzymology* 193, 455-79, 1990.

102. K. R. Jennings and G. G. Dolnikowski, Mass analyzers, *Methods in Enzymology* 193, 37-61, 1990.

103. K. Strupat, M. Karas, and F. Hillenkamp, 2,5-dihydroxybenzoic acid - a new matrix for laser desorption ionization mass spectrometry, *International Journal of Mass Spectrometry And Ion Processes* 111, 89-102, 1991.

104. M. Karas, H. Ehring, E. Nordhoff, B. Stahl, K. Strupat, F. Hillenkamp, M. Grehl, and B. Krebs, Matrix-assisted laser desorption ionization mass spectrometry with additives to 2,5-dihydroxybenzoic acid, *Organic Mass Spectrometry* 28, 1476-1481, 1993.

105. R. C. Beavis, T. Chaudhary, and B. T. Chait, Alpha-cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry, *Organic Mass Spectrometry* 27, 156-158, 1992.

106. K. Tang, N. I. Taranenko, S. L. Allman, C. H. Chen, L. Y. Chang, and K. B. Jacobson, Picolinic acid as a matrix for laser mass spectrometry of nucleic acids and proteins, *Rapid Communications in Mass Spectrometry* 8, 673-677, 1994.

107. U. Pieles, W. Zurcher, M. Schar, and H. E. Moser, Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and

sequence analysis of natural and modified oligonucleotides, *Nucleic Acids Research* 21, 3191-6, 1993.

108. R. Kaufmann, D. Kirsch, and B. Spengler, Sequencing of peptides in a time-of-fight mass spectrometer: evaluation of post-source decay following matrix-assisted laser desorption/ionization (MALDI), *International Journal of Mass Spectrometry and Ion Processes* 131, 355-385, 1994.

109. J. Kim and K. Kim, Identification of the C-terminal amino acid amides by carboxypeptidase Y digestion and fast atom bombardment mass spectrometry, *Biochemistry and Molecular Biology International* 34, 897-907, 1994.

110. J. Kim, K. Kim, J. Kim, J. H. Ok, and J. Kim, Application of carboxypeptidase Y and fast atom bombardment mass spectrometry for C-terminal sequencing of small peptides, *Biochemistry and Molecular Biology International* 33, 55-64, 1994.

111. B. Thiede, J. Salnikow, and B. Wittmann-Liebold, C-terminal ladder sequencing by an approach combining chemical degradation with analysis by matrix-assisted-laser-desorption ionization mass spectrometry, *European Journal of Biochemistry* 244, 750-4, 1997.

112. D. H. Patterson, G. E. Tarr, F. E. Regnier, and S. A. Martin, C-terminal ladder sequencing via matrix-assisted laser desorption mass spectrometry coupled with carboxypeptidase Y time-dependent and concentration-dependent digestions, *Analytical Chemistry* 67, 3971-8, 1995.

113. J. W. Hockensmith, W. L. Kubasek, W. R. Vorachek, and P. H. von Hippel, Laser cross-linking of proteins to nucleic acids. I. Examining physical parameters of protein-nucleic acid complexes, *Journal of Biological Chemistry* 268, 15712-20, 1993.

114. J. W. Hockensmith, W. L. Kubasek, E. M. Evertsz, L. D. Mesner, and P. H. von Hippel, Laser cross-linking of proteins to nucleic acids. II. Interactions of the bacteriophage T4 DNA replication polymerase accessory proteins complex with DNA, *Journal of Biological Chemistry* 268, 15721-30, 1993.

115. W. L. Kubasek, D. Spann, and J. W. Hockensmith, Laser cross-linking of proteins to nucleic acids: photodegradation and alternative photoproducts of the bacteriophage T4 gene 32 protein, *Photochemistry and Photobiology* 58, 1-10, 1993.

116. J. W. Hockensmith, W. L. Kubasek, W. R. Vorachek, E. M. Evertsz, and P. H. von Hippel, Laser cross-linking of protein-nucleic acid complexes, *Methods in Enzymology* 208, 211-36, 1991.

117. J. W. Hockensmith, W. L. Kubasek, W. R. Vorachek, and P. H. von Hippel, Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system, *Journal of Biological Chemistry* 261, 3512-8, 1986.

118. G. B. Zavil'gel'skii, G. G. Gurzadian, and D. N. Nikogosian, [Laser-induced covalent cross-links in DNA], *Biofizika* 30, 568-70, 1985.

119. K. R. Williams and W. H. Konigsberg, Identification of amino acid residues at interface of protein-nucleic acid complexes by photochemical cross-linking, *Methods in Enzymology* 208, 516-39, 1991.

120. D. K. Srivastava, R. K. Evans, A. Kumar, W. A. Beard, and S. H. Wilson, dNTP binding site in rat DNA polymerase beta revealed by controlled proteolysis and azido photoprobe cross-linking, *Biochemistry* 35, 3728-34, 1996.

121. S. E. Bennett, O. N. Jensen, D. F. Barofsky, and D. W. Mosbaugh, UV-catalyzed cross-linking of Escherichia coli uracil-DNA glycosylase to DNA. Identification of amino acid residues in the single-stranded DNA binding site, *Journal of Biological Chemistry* 269, 21870-9, 1994.

122. V. N. Pandey, N. Kaushik, and M. J. Modak, Photoaffinity labeling of DNA template-primer binding site in Escherichia coli DNA polymerase I. Identification of involved amino acids, *Journal of Biological Chemistry* 269, 21828-34, 1994.

123. M. A. Mullen, H. Wang, K. Wilcox, and T. Herman, Characterization of a Max:DNA complex by cross-linking to photoactive oligonucleotides, *DNA and Cell Biology* 13, 521-30, 1994.

124. M. Drelich, M. Haenggi, and J. Mous, Conserved residues Pro-109 and Asp-116 are required for interaction of the human immunodeficiency virus type 1 integrase protein with its viral DNA substrate, *Journal of Virology* 67, 5041-4, 1993.

125. C. Wenzel and W. Guschlbauer, Dam methyltransferase from Escherichia coli: sequence of a peptide segment involved in S-adenosyl-methionine binding, *Nucleic Acids Research* 21, 4604-9, 1993.

126. R. Prasad, A. Kumar, S. G. Widen, J. R. Casas-Finet, and S. H. Wilson, Identification of residues in the single-stranded DNA-binding site of the 8-kDa domain of rat DNA polymerase beta by UV cross-linking, *Journal of Biological Chemistry* 268, 22746-55, 1993.

127. V. Cleghon and D. F. Klessig, Characterization of the nucleic acid binding region of adenovirus DNA binding protein by partial proteolysis and photochemical cross-linking, *Journal of Biological Chemistry* 267, 17872-81, 1992.

128. K. R. Webster, S. Keill, W. Konigsberg, K. R. Williams, and E. K. Spicer, Identification of amino acid residues at the interface of a bacteriophage T4 regA protein-nucleic acid complex, *Journal of Biological Chemistry* 267, 26097-103, 1992.

129. D. K. Lee, R. K. Evans, J. Blanco, J. Gottesfeld, and J. D. Johnson, Contacts between 5 S DNA and Xenopus TFIIIA identified using 5-azido-2'-deoxyuridine-substituted DNA, *Journal of Biological Chemistry* 266, 16478-84, 1991.

130. R. W. Sobol, R. J. Suhadolnik, A. Kumar, B. J. Lee, D. L. Hatfield, and S. H. Wilson, Localization of a polynucleotide binding region in the HIV-1 reverse transcriptase: implications for primer binding, *Biochemistry* 30, 10623-31, 1991.

131. J. Rush and W. H. Konigsberg, Photoaffinity labeling of the Klenow fragment with 8-azido-dATP, *Journal of Biological Chemistry* 265, 4821-7, 1990.

132. Y. Shamoo, K. R. Williams, and W. H. Konigsberg, Photochemical crosslinking of bacteriophage T4 single-stranded DNA-binding protein (gp32) to oligo-p(dT)8: identification of phenylalanine-183 as the site of crosslinking, *Proteins* 4, 1-6, 1988.

133. L. P. Kurochkina, A. A. Komissarov, and G. I. a. Kolomiitseva, [Localization of lysine residue in histone H3 forming the thymine-lysine cross-link after UV-irradiation of deoxyribonucleoprotein], *Biokhimiia* 52, 1683-7, 1987.

134. T. M. Cao and M. T. Sung, Ultraviolet light induced preferential cross-linking of histone H3 to deoxyribonucleic acid in chromatin and nuclei of chicken erythrocytes, *Biochemistry* 21, 3419-27, 1982.

135. J. E. Callaway, R. J. DeLange, and H. G. Martinson, Contact site of histones 2A and 2B in chromatin and in solution, *Biochemistry* 24, 2686-92, 1985.

136. B. M. Merrill, K. R. Williams, J. W. Chase, and W. H. Konigsberg, Photochemical cross-linking of the Escherichia coli single-stranded DNA-binding protein to oligodeoxynucleotides. Identification of phenylalanine 60 as the site of cross-linking, *Journal of Biological Chemistry* 259, 10850-6, 1984.

137. I. Saito, H. Sugiyama, and T. Matsuura, Mechanism of DNA-histone crosslinks in UV-irradiated nuclei, *Nucleic Acids Symposium Series* 11, 225-8, 1982.

138. J. Sperling and R. Sperling, Photochemical cross-linking of histones to DNA nucleosomes, *Nucleic Acids Research* 5, 2755-73, 1978.

139. P. R. Paradiso and W. Konigsberg, Photochemical cross-linking of the gene 5 protein.fd DNA complex from fd-infected cells, *Journal of Biological Chemistry* 257, 1462-7, 1982.

140. P. R. Paradiso, Y. Nakashima, and W. Konigsberg, Photochemical cross-linking of protein-nucleic acid complexes. The attachment of the fd gene 5 protein to fd DNA, *Journal of Biological Chemistry* 254, 4739-44, 1979.

141. E. Anderson, Y. Nakashima, and W. Konigsberg, Photo-incuced cross-linkage of gene-5 protein and bacteriophage fd DNA, *Nucleic Acids Research* 2, 361-71, 1975.

142. P. Maly, J. Rinke, E. Ulmer, C. Zwieb, and R. Brimacombe, Precise localization of the site of cross-linking between protein L4 and 23S ribonucleic acid induced by mild

ultraviolet irradiation of Escherichia coli 50S ribosomal subunits, *Biochemistry* 19, 4179-88, 1980.

143. O. N. Jensen, D. F. Barofsky, M. C. Young, P. H. von Hippel, S. Swenson, and S. E. Seifried, Direct observation of UV-crosslinked protein-nucleic acid complexes by matrix-assisted laser desorption ionization mass spectrometry, *Rapid Communications in Mass Spectrometry* 7, 496-501, 1993.

144. P. R. Callis, An extended semi-empirical molecular orbital study of the pi pi excited states of nucleic acid bases, *Photochemistry and Photobiology* 44, 315-22, 1986.

145. J. Cadet, M. Berger, C. Decarroz, J. R. Wagner, J. E. van Lier, Y. M. Ginot, and P. Vigny, Photosensitized reactions of nucleic acid, *Biochimie* 68, 813-34, 1986.

146. N. C. Alma, B. J. Harmsen, J. H. van Boom, G. van der Marel, and C. W. Hilbers, A 500-MHz proton nuclear magnetic resonance study of the structure and structural alterations of gene-5 protein-oligo(deoxyadenylic acid) complexes, *Biochemistry* 22, 2104-15, 1983.

147. J. E. Coleman, R. A. Anderson, R. G. Ratcliffe, and I. M. Armitage, Structure of gene 5 protein-oligodeoxynucleotide complexes as determined by 1H, 19F, and 31P nuclear magnetic resonance, *Biochemistry* 15, 5419-30, 1976.

148. P. J. Folkers, J. P. van Duynhoven, A. J. Jonker, B. J. Harmsen, R. N. Konings, and C. W. Hilbers, Sequence-specific 1H-NMR assignment and secondary structure of the Tyr41----His mutant of the single-stranded DNA binding protein, gene V protein, encoded by the filamentous bacteriophage M13, *European Journal of Biochemistry* 202, 349-60, 1991.

149. P. J. Folkers, J. P. van Duynhoven, H. T. van Lieshout, B. J. Harmsen, J. H. van Boom, G. I. Tesser, R. N. Konings, and C. W. Hilbers, Exploring the DNA binding domain of gene V protein encoded by bacteriophage M13 with the aid of spin-labeled oligonucleotides in combination with 1H-NMR, *Biochemistry* 32, 9407-16, 1993.

150. R. V. Prigodich, J. Casas-Finet, K. R. Williams, W. Konigsberg, and J. E. Coleman, 1H NMR (500 MHz) of gene 32 protein--oligonucleotide complexes, *Biochemistry* 23, 522-9, 1984.

151. J. L. Corden, Tails of RNA polymerase II, *Trends in Biochemical Sciences* 15, 383-7, 1990.

152. A. Emili and C. J. Ingles, The RNA polymerase II carboxy-terminal domain: links to a bigger and better 'holoenzyme'?, *Current Opinion in Genetics and Development* 5, 204-9, 1995.

153. M. Suzuki, The heptad repeat in the largest subunit of RNA polymerase II binds by intercalating into DNA, *Nature* 344, 562-5, 1990.

154. M. M. Harding, NMR studies on YSPTSPSY: implications for the design of DNA bisintercalators, *Journal of Medicinal Chemistry* 35, 4658-64, 1992.

155. P. M. Cagas and J. L. Corden, Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II, *Proteins* 21, 149-60, 1995.

156. A. Khiat, M. Lamoureux, and Y. Boulanger, Structural differences between the free and bound states of the DNA-bisintercalating peptide YSPTSPSY, *Journal of Medicinal Chemistry* 39, 2492-8, 1996.

157. A. Yuryev, M. Patturajan, Y. Litingtung, R. V. Joshi, C. Gentile, M. Gebara, and J. L. Corden, The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins, *Proceedings of the National Academy of Sciences of the United States of America* 93, 6975-80, 1996.

158. M. E. Kang and M. E. Dahmus, RNA polymerases IIA and IIO have distinct roles during transcription from the TATA-less murine dihydrofolate reductase promoter, *Journal of Biological Chemistry* 268, 25033-40, 1993.

159. S. Buratowski and P. A. Sharp, Transcription initiation complexes and upstream activation with RNA polymerase II lacking the C-terminal domain of the largest subunit, *Molecular and Cellular Biology* 10, 5562-4, 1990.

160. W. A. Zehring and A. L. Greenleaf, The carboxyl-terminal repeat domain of RNA polymerase II is not required for transcription factor Sp1 to function in vitro, *Journal of Biological Chemistry* 265, 8351-3, 1990.

161. E. Barron-Casella and J. L. Corden, Conservation of the mammalian RNA polymerase II largest-subunit C-terminal domain, *Journal of Molecular Evolution* 35, 405-10, 1992.

162. O. N. Jensen, S. Kulkarni, J. V. Aldrich, and D. F. Barofsky, Characterization of peptide-oligonucleotide heteroconjugates by mass spectrometry, *Nucleic Acids Research* 24, 3866-3872, 1996.

163. M. S. W. Lipton, A. F. Fuciarelli, D. L. Springer, and C. G. Edmonds, Characterization of radiation-induced thymine-tyrosine crosslinks by electrospray ionization mass spectrometry, *Radiation Research* 145, 681-686, 1996.

164. M. M. Warshaw and I. Tinoco Jr., Absorption and optical rotatory dispersion of six dinucleoside phosphates, *Journal of Molecular Biology* 13, 54-64, 1965.

165. M. M. Warshaw and I. Tinoco Jr., Optical properties of sixteen dinucleoside phosphates, *Journal of Molecular Biology* 20, 29-38, 1966.

166. A. M. Falick, F. C. Walls, and R. A. Laine, Cooled sample introduction probe for liquid secondary ionization mass spectrometry, *Analytical Biochemistry* 159, 132-7, 1986.

167. A. M. Falick, G. H. Wang, and F. C. Walls, Ion source for liquid matrix secondary ionization mass spectrometry, *Analytical Chemistry* 58, 1308-11, 1986.

168. R. M. Caprioli and W. T. Moore, Continuous-flow fast atom bombardment mass spectrometry, *Methods in Enzymology* 193, 214-37, 1990.

169. K. Biemann, Appendix 6. Mass values for amino acid residues in peptides, *Methods in Enzymology* 193, 888, 1990.

170. S. I. Dimitrov and T. Moss, UV laser-induced protein-DNA crosslinking, *Methods in Molecular Biology* 30, 227-36, 1994.

171. I. G. Pashev, S. I. Dimitrov, and D. Angelov, Crosslinking proteins to nucleic acids by ultraviolet laser irradiation, *Trends in Biochemical Sciences* 16, 323-6, 1991.

172. C. A. Harrison, D. H. Turner, and D. C. Hinkle, Laser crosslinking of E. coli RNA polymerase and T7 DNA, *Nucleic Acids Research* 10, 2399-414, 1982.

173. S. S. Sastry, H. P. Spielmann, Q. S. Hoang, A. M. Phillips, A. Sancar, and J. E. Hearst, Laser-induced protein-DNA cross-links via psoralen furanside monoadducts [published erratum appears in Biochemistry 1994 Feb 15;33(6):1616], *Biochemistry* 32, 5526-38, 1993.

174. T. Date, M. Yamaguchi, F. Hirose, Y. Nishimoto, K. Tanihara, and A. Matsukage, Expression of active rat DNA polymerase beta in Escherichia coli, *Biochemistry* 27, 2983-90, 1988.

175. A. Kumar, S. G. Widen, K. R. Williams, P. Kedar, R. L. Karpel, and S. H. Wilson, Studies of the domain structure of mammalian DNA polymerase beta. Identification of a discrete template binding domain, *Journal of Biological Chemistry* 265, 2124-31, 1990.

176. S. J. Kim, M. S. Lewis, J. R. Knutson, D. K. Porter, A. Kumar, and S. H. Wilson, Characterization of the tryptophan fluorescence and hydrodynamic properties of rat DNA polymerase beta, *Journal of Molecular Biology* 244, 224-35, 1994.

177. A. Kumar, J. Abbotts, E. M. Karawya, and S. H. Wilson, Identification and properties of the catalytic domain of mammalian DNA polymerase beta, *Biochemistry* 29, 7156-9, 1990.

178. A. Basu, P. Kedar, S. H. Wilson, and M. J. Modak, Active-site modification of mammalian DNA polymerase beta with pyridoxal 5'-phosphate: mechanism of inhibition and identification of lysine 71 in the deoxynucleoside triphosphate binding pocket, *Biochemistry* 28, 6305-9, 1989.