AN ABSTRACT OF THE DISSERTATION OF

Catalin E. Doneanu for the degree of Doctor of Philosophy in Chemistry presented on September 18, 2002. Title: Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes

Abstract approved:

Redacted for privacy

Douglas F. Barofsky

The DNA-binding domains of *E. coli* uracil-DNA glycosylase (Ung) and human replication protein A (hRPA) were studied using a general protocol developed in our laboratory for probing protein-DNA interactions. The procedure involves purification and mass spectrometric analysis of the nucleopeptide-products of a tryptically digested UV-crosslinked protein-nucleic acid complex.

In the case of Ung x dT_{20} nucleoprotein complex, three nucleopeptide isomers having the same peptide backbone (T_{18} peptide) but with dinucleotides attached to different aminoacids were separated. The tandem mass spectra from the isomers provided new structural information about Ung binding to DNA. Specifically, His_{187}, Ser_{189}, and His_{194} from T_{18} nucleopeptide were putatively identified as sites that photocrosslink to dT_{20}.

Photochemical crosslinking of hRPA to oligonucleotide dT_{30} produced two covalent hRPA70 x dT_{30} complexes involving one of the protein's subunits (hRPA70). Three crosslinked tryptic peptides were isolated using the same protocol as used for Ung and MALDI-TOF and nanoLC-ESI-MS/MS analyses revealed the identity of these peptides as T_{43}, $T_{28/29}$, and a truncated $^{*}T_{24/25}$ (without the last 5 aminoacids from the C-teminal). Additional experiments showed that at least one amino acid from the sequence 383-VSDF-386 (located in T_{43}), at least one residue from 235-ATAFNE-240 ($^{*}T_{24/25}$), and at least one residue from F269/T270 ($T_{28/29}$) is involved in crosslinking. Aromatic residues contained in these peptides (F238, F269 and F386), which can form base stacking interactions with the DNA, are the residues most likely to be involved in crosslinking. These observations are in good agreement with previously published data regarding the single stranded-DNA binding site of hRPA obtained from crystal structure and from site-directed mutagenesis experiments.

Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes

by

Catalin E. Doneanu

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented on September 18, 2002
Commencement June 2003

Doctor of Philosophy dissertation of Catalin E. Doneanu
presented on September 18, 2002.

APPROVED:

Redacted for privacy

Major Professor, representing Chemistry

Redacted for privacy

Chair of the Department of Chemistry

Redacted for privacy

Dean of Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for privacy

Catalin E. Doneanu, Author

ACKNOWLEDGMENTS

I would like to acknowledge first my research advisor, Professor Douglas F. Barofsky, for giving me the opportunity to pursue this research in his laboratory, and also for his scientific guidance, stimulating discussions and constructive criticism of my work. Thank you for ensuring that I became an independent scientist.

Many thanks to Lilo Barofsky for running MALDI samples and to Brian Arbogast and Don Griffin for their help with the LC/MS setup.

I would like to acknowledge Dr. Phil Gafken, Dr. Sam Bennett and Professor Dale Mosbaugh for their help and assistance offered for the protein purification protocols.

I would also like to thank those who served on my graduate committee: Dr. Max Deinzer, Dr. James Ingle, Dr. Michael Schimerlik and Dr. John Simonsen.

For all my friends who provided help, ideas and encouragement, I want to express my appreciation.

I would like to thank all my fellow graduate students with whom I have been through all the hardship and excitement of learning, research and life.

I wish to thank my family for their love, devotion, and moral support.

Lastly and most importantly I wish to thank my wife Angela. Without her encouragement, faith, love and support, this work would surely not have been possible.

CONTRIBUTION OF AUTHORS

I have been responsible for most of the experimental work presented in this thesis, including the design and construction of the nanoscale liquid chromatography system described in Chapter 2 and performing the experiments and data analysis described in Chapter 3 and 4.

Dr. Phil Gafken, Dr. Sam Bennett and Professor Dale Mosbaugh provided help and assistance for Ung and hRPA purification.

Dr. Ajoy Velayudhan derived the mathematical equation for the exponential dilution method involving two mixing chambers.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

MASS SPECTROMETRIC ANALYSIS OF UV-CROSSLINKED PROTEIN-NUCLEIC ACID COMPLEXES

1. INTRODUCTION

1.1. ANALYSIS OF PROTEIN-DNA INTERACTIONS

The study of the interaction of proteins, whether they are enzymes, regulatory proteins or structural proteins, with nucleic acids is one of the most exciting areas of modern molecular biology. Structural studies of DNA-binding proteins and their complexes with DNA have proceeded at an accelerated pace over the last two decades due to important technical advances in molecular genetics, DNA synthesis, UV-catalyzed crosslinking, protein crystallography, and nuclear magnetic resonance spectroscopy. The classical direct methods used to investigate protein-DNA interactions are X-ray diffraction and NMR spectroscopy, but both these methods have certain limitations. A critical prerequisite of the crystallography method is the preparation of a well-ordered crystalline nucleoprotein complex that provides X-ray diffraction data of sufficient resolution. Few of the many protein-DNA systems investigated produce high quality X-ray crystals. In addition, the nonphysiological conditions often required to produce X-ray quality crystals diminish the biological relevance of the study. Even with these severe limitations,

X-ray crystallography is still the most powerful tool used to visualize protein-DNA interactions at the atomic level. Multidimensional high resolution NMR spectroscopy was developed in the last decade and used to provide information about the kinetic and thermodynamic properties of protein-DNA systems. However, the complexity of the NMR spectra limits this method to proteins with a molecular weight below 40 kDa.

The emergence of mass spectrometry as an analytical tool in biochemical research has opened up new possibilities for investigating protein-DNA interactions. It has been shown that intact nucleoprotein complexes can be analyzed by MALDI [1] and electrospray mass spectrometry [2].

A new direct method for studying protein-DNA complexes combining low intensity UV crosslinking with mass spectrometry was introduced by Barofsky et al in 1994 [3]. Soon after, laser-induced crosslinking [4,5] and chemical crosslinking [6] were used in combination with mass spectrometry to investigate protein-DNA interactions. This method had become increasingly popular in recent years [7-14].

1.2. UV CROSSLINKING

Early in the 1960s Smith [15] and Alexander and Moroson [16] published experimental results indicating that absorption of UV light could induce crosslinking of proteins to DNA in living cells. Over the next 40 years, UV photochemical crosslinking became a well-established biochemical procedure used for determining the contact domains between nucleic acids and proteins in nucleoprotein complexes [17-19]. According to this procedure, UV light (in the range of 250-270 nm), can be used to induce "zero-length" covalent bonds between DNA nucleobases and certain amino acids from the protein situated in close contact with the DNA. Upon UV irradiation, DNA bases can absorb photons and produce excited state radicals with lifetimes on the order of a few picoseconds to a few microseconds. Because the time scale for the motion of the nucleoprotein complex is on the order of milliseconds, these reactive radicals have the opportunity to form covalent bonds with certain amino acids that are in contact with the DNA, thereby "freezing" the interaction between the biopolymers. The crosslinked complexes can then be isolated for further characterization.

Any nucleotide present in DNA could in principle be involved in photoreactions, but the reactivity of different nucleotides varies greatly. Thymine appears by far to be the most reactive nucleobase and for this reason deoxythymidines are used frequently as DNA substrates in UV crosslinking experiments.

Several side reactions normally occur during the UV crosslinking process. These side reactions include protein denaturation, dimerization or degradation, as well as DNA modification or degradation. Crosslinking yields are relatively low and usually less than 10% of a given protein is crosslinked. Higher crosslinking yields due to minimized side reactions can be obtained if UV lasers are used instead of low intensity UV lamps.

Crosslinked nucleoprotein complexes are usually analyzed first by denaturing polyacrylamide gel electrophoresis (PAGE). The crosslinked protein migrates slower than the free protein and produces a shifted gel band. For this reason, the technique employed to detect crosslinked complexes is called electrophoretic mobility shift assay. Several small-scale electrophoretic mobility shift assays have to be conducted in order to optimize the irradiation time and the ratio of oligonucleotide to protein to form a stable protein-DNA complex.

1.3. MATRIX-ASSISTED LASER DESORPTION-IONIZATION MASS SPECTROMETRY

Since the first description of matrix-assisted laser desorption-ionization (MALDI) mass spectrometry by Karas and Hillenkamp [20,21], MALDI MS has become a widespread analytical technique for the analysis of peptides, proteins and oligonucleotides. The design of the MALDI-TOF instrument is illustrated in Figure 1.1. The principal components include a pulsed UV laser, a beam attenuator,

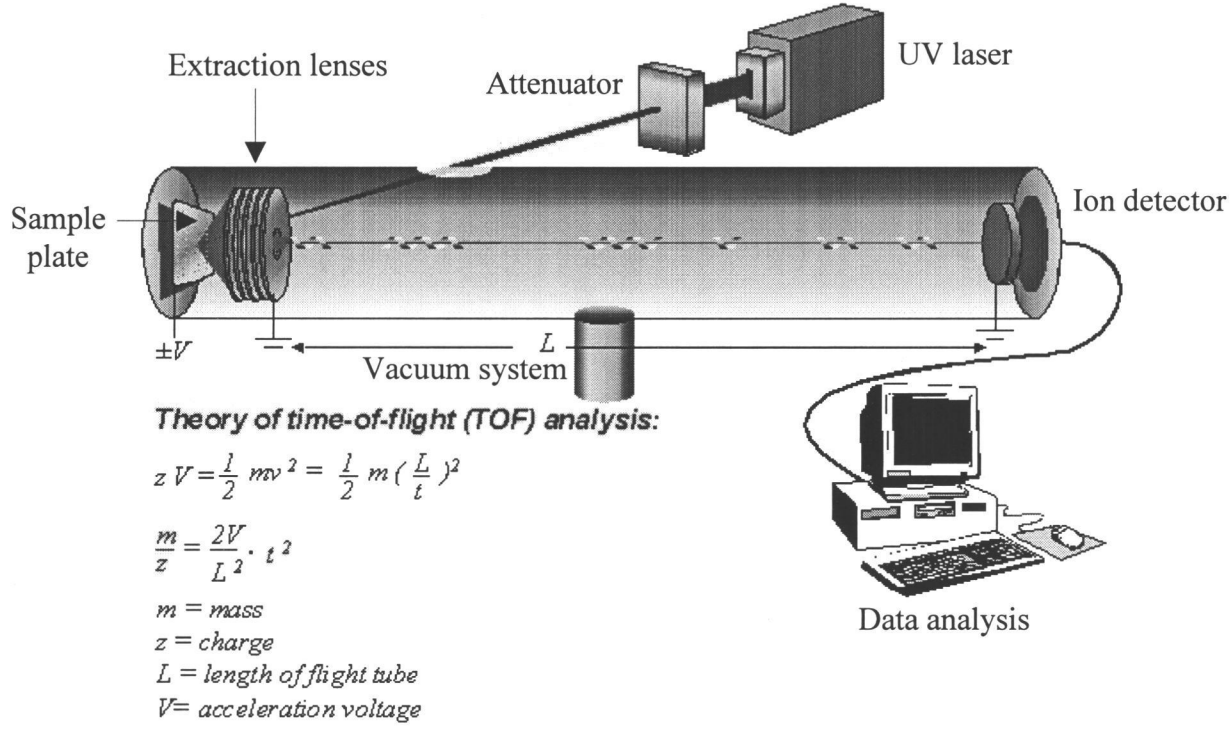


Figure 1.1. The schematic diagram of the MALDI instrument. Peptide ions produced by laser desorption are extracted and analyzed by a time-of-flight instrument. The flight time for each ion species is proportional to the square root of its *m/z* ratio.

a sample plate, the vacuum system, extraction lenses, the flight tube, the ion detector, and the data analysis system.

In the MALDI experiment, analyte molecules are diluted in a matrix solution (to about 1:5000 molar ratio) and spotted on the sample plate. After the solvent is evaporated and sample crystallization has occured, the plate is transferred to the vacuum system of the mass spectrometer. The function of the matrix is to isolate the sample molecules from each other and to absorb photons from the pulsed UV laser. Absorption of the laser energy causes rapid desorption of the matrix molecules into the gas phase carrying the analyte molecules with them. Ionization is thought to occur primarily via protonation in the plume above the sample plate [112]. These protonated molecules ($[M + H]^+$) are analyzed by a time-of-flight (TOF) mass spectrometer, because this mass analyzer is well-suited to the pulsed nature of laser desorption. All ions are accelerated to a uniform energy (20-30 keV) and then allowed to fly in a field-free flight tube to reach the detector. The total flight-time is measured from the time of the laser pulse until the ions reach the detector and generate a signal. This time-scale is converted into a mass scale in order to obtain a mass spectrum. As seen from the formula from Figure 1.1, the flight time of each ion species is directly related to its mass to charge ratio (*m/z*). In practice, instead of using this formula, the mass scale is calibrated by using well-defined molecular weight standards.

The TOF analyzer is highly efficient because all ions of different *m/z* ratios arising from a single laser shot are measured; they simply arrive at the ion detector at

different times. Performance specifications of commercial MALDI-TOF instruments include a resolving power of 500 (in the linear mode), a mass determination accuracy of about 100 ppm, a sensitivity of better than 1 pmol, a mass range of 500-300,000 Da and a spectrum acquisition time of 1 min. The poor resolving power of the MALDI-TOF instrument in the linear mode is a consequence of the initial temporal distribution and initial velocity distribution of the generated ions. The temporal distribution of ions (observed for ions formed at different times) can be significantly reduced by forming the ions with a short laser pulse (1-10 ns) on a flat surface (MALDI sample plate) and using fast, high-performance electronic devices for data recording. The initial velocity distribution of ions will cause the ions with the same *m/z* ratio to have different flight times. The resolution of MALDI spectra was significantly improved in 1995 [22,23] by the application of a time-delayed ion extraction technique originally introduced in 1953 [113]. A short delay (20-200 ns) is introduced between the laser pulse and the application of the extraction voltage so that the ions from the source can drift apart by virtue of their initial velocity differences. Initially faster ions penetrate deeper into the source volume than initially slower ions of the same mass. When the extraction voltage is applied after the delay time, the faster ions, which are now closer to the extraction grid, experience a lower extraction potential than the slower ions. Consequently, the initially slower ions acquire more kinetic energy during the extraction/acceleration step than the initially faster ions. By adjusting the delay time and/or the extraction potentials one can arrange for the initially slower ions to

catch up with the initially faster ions of the same mass exactly at the detector plane (space-focus plane) in order to obtain maximum instrumental resolution.

The resolution of the instrument can be improved further by using an electrostatic ion mirror (reflectron) for focusing ions with an initial kinetic energy distribution [24]. The ion mirror is placed at the end of the field-free region and creates a retarding electric field, causing ions to slow down and reverse direction. Ions of a given mass with higher kinetic energy penetrate deeper into the mirror before they turn around, and have a longer flight-path through the reflector than ions of the same mass with a lower kinetic energy. Thus, a higher kinetic energy is compensated for by a longer flight-path. By using delayed extraction in combination with an ion mirror, the resolving power of a MALDI-TOF instrument can be increased to greater than 20,000 over a fairly broad mass range (500-5,000 Da). The MALDI spectra presented in this thesis were recorded on a home-build instrument, equipped with delayed extraction and an ion mirror [1].

MALDI-MS is a routine technique used for protein identification by peptide mapping. MALDI-TOF peptide mapping operates efficiently when a significant number of peptide peaks can be measured, but until quite recently it possessed almost no power to characterize, i.e. sequence, an individual peptide whose signal appeared in the spectrum. Characterization of individual peptide ions in MALDI-TOF instruments has been possible for some time with the postsource decay (PSD) technique [25]. However, the low yield of fragment ions significantly reduces the utility of PSD analysis in all but a few cases. The possibility to interrogate the

sequence of any peptide peak observed in a MALDI peptide map by tandem mass spectrometry was demonstrated almost at the same time on a quadrupole-time of flight (QTOF) [26], ion trap instruments [27,28] and on a MALDI TOF/TOF tandem mass spectrometer [29]. The later instrument, which uses high-energy (> 1 keV) collision induced disociation (CID) to fragment peptide ions, was made possible by ion optics invented at Oregon State University and Uppsala (Sweden) University [115].

1.4. ELECTROSPRAY IONIZATION MASS SPECTROMETRY

Electrospray ionization (ESI) coupled with mass spectrometry was first demonstrated in the early 1980s by Fenn et al. for small molecules [30] and in the late 1980s for proteins [31]. In the most general sense, electrospray is a method of generating a very fine liquid aerosol through electrostatic charging. In the case of biomolecules, electrospray ionization provides a way to produce ions in the gas phase, at atmospheric pressure, directly from solution. These ions can be analyzed by a mass spectrometer. The essentials of an electrospray mass spectrometer are shown in Figure 1.2. The analyte is dissolved in solution and pumped through a metallic needle at flow rates between 5 and 10 μL/min. A high voltage power supply is connected to the electrospray needle placed in front of the instrument. When the high voltage (1-5 kV) is turned on, the liquid being pumped through the

needle is transformed into a fine continuous mist of droplets that fly rapidly toward the mass spectrometer's counter-electrode commonly held at ground potential.

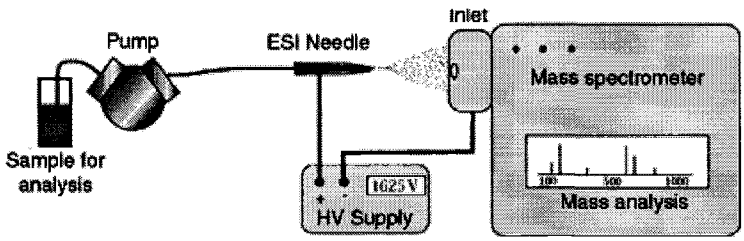


Figure 1.2. Design of the electrospray ionization source. The main components are the pump, the electrospray needle and the high voltage supply.

The liquid charges up as it exits the needle and assumes a conical shape, referred to as the "Taylor" cone [32]. At the tip of the cone, the liquid changes its shape into a fine jet of liquid. Then, at some point in front of the instrument, the liquid jet becomes unstable and breaks up into a mist of tiny droplets (less than 10 μm in diameter) that are highly charged and repel each other very strongly. For this reason, the droplets fly apart from each other and cover a wide surface area. As solvent molecules evaporate, the droplets will shrink rapidly decreasing the distance between the electrical charges on the droplet. When the electrostatic repulsion between electrical charges overcomes the surface tension, the droplet

blows apart into smaller droplets. This phenomenon, which is known as a Coulomb explosion, is responsible for producing smaller and smaller droplets. Repetitive Coulomb explosions can produce, according to some authors [33,34], nanometer size droplets containing only one biomolecule. According to this model, the solvent molecules will totally evaporate, leaving the macromolecules as gas phase ions. Fenn et al. proposed a different ionization model [35,36]. They suggested that a charged part of a biomolecule penetrates the surface of a small, highly charged droplet produced by Coulomb explosions from an initially larger droplet. The electrostatic repulsion between this part of the biomolecule and the droplet surface would pull the entire biomolecule out of the droplet and would generate a gas phase molecular ion. The essential difference between these two ionization models is that in Fenn's model, the macromolecule plays an active role in the ionization process, while in the first model the ionization process is passive (does not depend on the analyte). Scientists are still debating over which of these models prevails.

Since the total charge has to be conserved, the electrospray process should also involve an electrochemical conversion of ions to electrons. This conversion occurs at the liquid-metal interface of the capillary through oxidation reactions. The electrospray process can be viewed as an electrolytic cell in which the ion transport does not occurs through uninterrupted solution, but rater as charged droplets and later as ions in the gas phase [37]. In 1995 Van Berkel and Zhou proved that the electrospray ion source behaves like a controlled-current electrolytic cell [38].

A theoretical model for the electrostatic dispersion of liquid in electrospray, was developed and verified experimentally by Wilm and Mann in 1994 [39]. According to this model, the threshold voltage (U_T) required to obtain a stable electrospray depends on the distance from the liquid jet to the counter-electrode (d), the surface tension of the liquid (γ) and the dielectric constant of the medium (ε_0):

$$U_T = 0.863\ (\gamma\, d / \varepsilon_0)^{0.5} \qquad (1.1.)$$

The same authors demonstrate the following formula (1.2.), which relates the liquid flow rate (F) to the emission radius (r) of the droplets emitted at the tip of the Taylor cone:

$$r = \left(\frac{\rho}{4\,\pi^2 \gamma \tan(\pi/2-\theta)[(U_a/U_T)^2 - 1]} \right)^{1/3} F^{2/3} \qquad (1.2.)$$

where:

ρ - is the density of the liquid expressed in g/cm^3;

θ – is the liquid cone angle;

U_a – is the applied voltage.

Formula (1.2.) predicts a higher desorption efficiency at lower flow rates because the solvent evaporation is more efficient if the first generation of droplets formed from the liquid jet are smaller. This would produce a higher ionization efficiency and ultimately a higher sensitivity in electrospray mass spectrometry. The overall efficiency, defined as the number of analyte ions recorded at the detector divided by the number of molecules analyzed, is on the order of 10^{-5} with conventional electrospray sources coupled to quadrupole mass analyzers [40]. This overall efficiency takes into account the desorption efficiency, the ionization efficiency and the transfer efficiency into the vacuum system of the mass spectrometer. Theoretically, the overall efficiency should be improved considerably at lower flow rates. To verify experimentally these predictions, Wilm and Mann introduced the nanoelectrospray (nanospray) source [39,40] designed to work at very low flow rates (~ 25 nL/min). They obtained an overall efficiency of 5×10^{-3} with the nano-ESI source, but the nanospray technique failed to become a routine technique for most of the analytical labs performing protein analysis. The coupling of nanoscale liquid chromatography with low flow (0.2 – 0.5 μL/min) microelectrospray mass spectrometry turned out to be a better choice. An explanation from the last statement will be given in the following section (1.4.).

Electrospray ionization has been coupled to all known mass analyzers (magnetic sectors, quadrupoles, time-of-flight, Fourier transform ion cyclotron resonance (FTICR) and ion traps [41]. Tandem mass spectrometers like triple-quadrupoles [42] and QTOFs (quadrupole/time-of-flight) [43] have became the most popular

electrospray instruments used for protein identification and proteomics research. Both instruments can be used to isolate peptide ions produced by electrospray and fragment them under low energy (< 100 eV) CID conditions in order to obtain sequencing information. Recently, a new dissociation method called electron-capture dissociation (ECD) was interfaced to electrospray ionization on an FTICR instrument [44] and shown to produce near complete MS sequence coverage for peptides and small proteins.

The electrospray mass spectrum of a protein indicates the presence of a family of charged states, averaging roughly one positive charge per 1,000 Da [45]. The signals produced by all charged states between +20 and +30 of a 30 kDa protein, would appear in the mass spectrum between *m/z* 1000 and *m/z* 1500. Each member of the series of charge states for a particular protein is related to its nearest neighbor by having one more or one less proton. The molecular weight of the parent molecule can be determined from any two consecutive charge states [46].

The electrospray spectra presented in this thesis were recorded on an LC-Q ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) and on an API III triple-quadrupole instrument (PE Sciex, Ontario, Canada).

1.5. NANOSCALE LIQUID CHROMATOGRAPHY - MASS SPECTROMETRY

The development of highly sensitive matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) mass spectrometric techniques for the analysis of peptides, proteins, and other biomolecules during the last decade generated the concomitant development of miniaturized separation techniques. Since electrospray is a continuous ionization technique, it can be more easily coupled on-line to liquid chromatography. Although the initial interest in nanoscale separations arose because of their increased separation efficiencies in comparison with the analytical columns, the major driving forces behind this coupling were the higher mass sensitivity that can be achieved for a given amount of sample and the limited amount of sample available for analysis.

Packed fused-silica capillary column with less that 100 μm i.d. were introduced by McGuffin and Novotny in early eighties [47] and the direct coupling with ESI/MS was first demonstrated by Hunt et al in 1993 [48]. The term nanoscale liquid chromatography (nano-LC), used for describing separations performed at flow rates in the range of 10-1000 nL/min, was introduced by Moseley et al [49]. Later on, Chervet presented the instrumental requirements for nanoscale liquid chromatography [50].

In nano-LC, like in other microcolumn LC techniques, all chromatographic volumes must be down-scaled by a factor (*f*):

$$f = \left(\frac{\text{i.d.}_{\text{conv}}}{\text{i.d.}_{\text{nano}}} \right)^2 \approx 3800 \qquad (1.3.)$$

where $\text{i.d.}_{\text{conv}}$ and $\text{i.d.}_{\text{nano}}$ are the internal diameters of the conventional (analytical) and nano-LC columns, respectively. This down-scaling is inevitable in order to maintain the chromatographic performance of the nano-LC system. Down-scaling from a 4.6 mm i.d. analytical column to a 75 μm i.d. nanocolumn, requires a factor of about 4000. This factor applies to all components and parameters of the LC system, such as flow rates, injection and detection volumes, and connecting capillaries. Nanoscale liquid chromatography is performed using the following parameters: flow rates between 0.2 and 0.4 μl/min, injection and detector volumes of 5-10 nL, peak volumes of 50-100 nL and connection tubing volumes below 15 nL. The injection volume can be increased substantially in order to obtain adequate sensitivity if sample focusing techniques are used for gradient separations. Dead volumes associated with the coupling of the electrospray needle to the nanocolumn can be eliminated if a fused-silica column with an integral frit - "PicoFrit" is packed with the stationary phase. The column is connected to an injector placed in a plastic box and the spraying voltage is applied at the injector after the sample is injected.

The direct coupling (with no make-up flow) of nano-LC with ESI-MS turned out to be a highly sensitivite technique for peptide analysis. One reason for the high sensitivity of nano-LC/ESI-MS in comparison with conventional micro-LC/ESI-

MS is that electrospray ionization produces gas phase ions more efficiently at lower flow rates (formula (1.2.)). More importantly, nanoscale separations produce very narrow and very concentrated analyte bands, with peak volumes on the order of 50-100 nL. ESI-MS behaves like a concentration-sensitive detector, and for this reason, one can achieve sensitivities in the fmole/μL range for peptide analysis. Nanospray is not as sensitive as nano-LC/ESI-MS because the level of the chemical noise from the sample lowers the signal/background ratio in the MS spectrum. In addition, because of the competitive nature of the electrospray ionization, not all peptides present in the sample will produce a signal above the background level. Nano-LC/ESI-MS became the preferred electrospray technique for peptide analysis because it is more sensitive and more reliable than nano-ESI/MS and has the potential to become a high-throughput analysis technique.

One of the limitations of nanoscale liquid chromatography coupled to electrospray tandem mass spectrometry (nano-LC/ESI-MS/MS) is generated by the narrowness of the chromatographic peak. The elution profile of a peptide peak is in the range of 7-15 s and frequently coeluting peptides are presented to the mass spectrometer simultaneously. The instrument makes instant decisions as to which peptide ion to select for the MS/MS event, and because peptide separation is a dynamic process, some of the peptide ions (usually those with lower intensity) will not be selected for fragmentation. For example, an ion trap mass spectrometer is able to record about 30 MS/MS peptide spectra/min. To overcome this problem, Davis and Lee [51] introduced the concept of variable flow nanoscale separation in which peptides

are separated at 200 nL/min, but the MS/MS spectra are recorded at a "peak parking" flow rate of 25 nL/min.

In order to exploit the advantages of nano-LC, it is necessary to have a micropumping system that can provide gradient elution at flow rates below 1 μL/min while maintaining accurate proportioning, homogeneous mixing with minimal delay, and reproducible delivery of the mobile phase. The first nanoscale LC systems commercially available were UltiMate (LC Packings, San Francisco, CA) and Agilent 1100 (Agilent Technologies, Palo Alto, CA) introduced in early 2000. The most common way to achieve gradient elution in the nano-LC domain is to insert a flow-splitter between a conventional HPLC solvent delivery system and a nanocolumn. The inlet ports of most microinjectors designed for capillary LC applications operating at flow rates of 5-10 μL/min have dead volumes between 5-15 μL. In nano-LC applications, dead volumes of this magnitude act as mixing chambers that introduce delays in establishing the gradient and alter the delivery system's gradient profile through solvent-mixing. Rather than taking inordinate measures to minimize the unavoidable dead volumes between the splitting point and the microinjector, one can use simpler, less expensive pumps to deliver solvents and take advantage of the exponential dilution [52] that occurs in the dead volumes to produce useful chromatographic gradients at flow rates less than 1 μL/min.

By pursuing this later alternative, I constructed an exponential gradient nano-LC system and coupled it online to ESI-MS and off-line to MALDI and nanospray MS.

A detailed description of the system [53] is presented in the second chapter of my thesis.

1.6. BIOLOGICAL SYSTEMS INVESTIGATED

The research conducted for this thesis was focused on elucidating the DNA-binding domains of *E. coli* uracil-DNA glycosilase and human replication protein A. The following sections provide a short biological background on both of these proteins.

1.6.1. *E. coli* Uracil-DNA Glycosylase

Uracil-DNA glycosylase catalyzes the cleavage of the N-glycosylic bond that joins the uracil base to the deoxyribose phosphate backbone of DNA. The enzyme recognizes uracil residues present in DNA as a result of dUMP incorporation or of deamination of cytosine and plays an important role in initiating the uracil-DNA excision repair pathway [54].

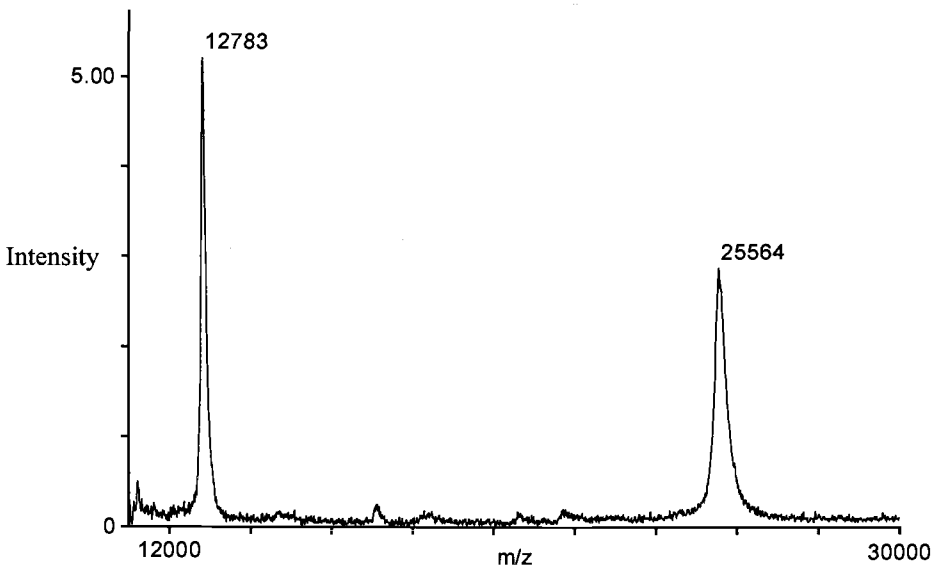
Uracil-DNA glycosilase isolated from *E. coli* (Ung) was the first enzyme of this class to be purified to homogeneity. The native enzyme is a single polypeptide containing 228 amino acids and has a relative molecular weight of M_r = 25,563 Da.

The enzyme was purified according to the purification protocol described in Chapter 3. Figure 1.3 shows the MALDI (1.3A.) and nano-ESI-MS (1.3B.) spectra recorded for Ung.

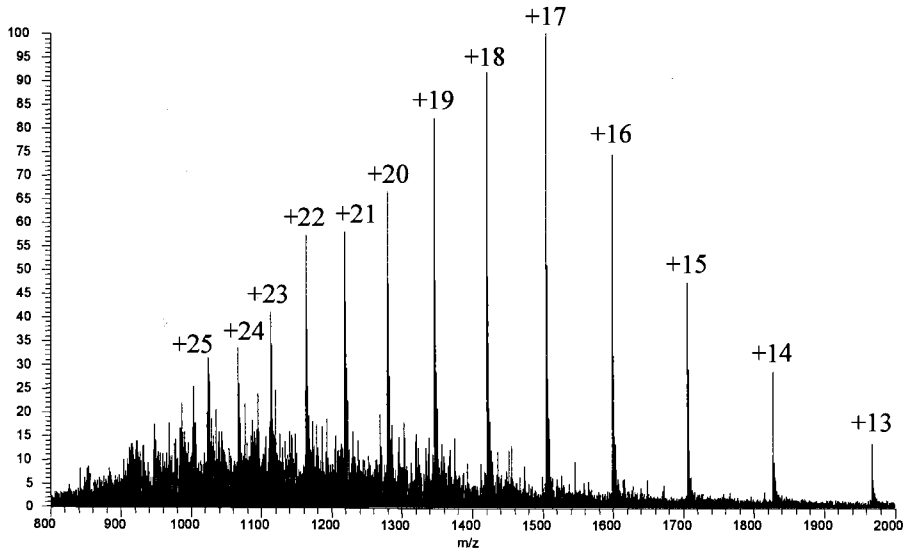
1.6.2. Human Replication Protein A

Replication protein A (RPA) is a single-standed DNA-binding protein that is highly conserved in eukaryotic cells. RPA was originally identified as a factor essential for the replication of simian virus SV40 DNA *in vitro* and subsequently has also been shown to play an important role in DNA repair and recombination [55]. Human RPA (hRPA) is a heterotrimer comprising 70, 32 and 14 kDa subunits. The 70-kDa subunit binds single-stranded DNA with high affinity and the DNA-binding domain was located between amino acids 181 and 422 [56].

hRPA was purified according to the procedure described in Chapter 4 of this thesis. Figure 1.4 displays the MALDI spectrum obtained for two of the protein's subunits, hRPA14 (M_r =13,569 Da) and hRPA32 (M_r =29,330 Da). The third subunit, hRPA70, also present in the mixture according to SDS-PAGE analysis, did not produce a signal probably because of the competitive nature of the MALDI ionization process.



A.



B.

Figure 1.3. MALDI and nano-ESI mass spectra of Ung. (A) A solution of 36 pmoles/μL of Ung was mixed with a matrix solution of saturated α-hydroxy-cinnamic acid (HCCA) in a 1:3 ratio and analyzed by MALDI. Only singly protonated (*m/z* = 25,563) and doubly protonated (*m/z* = 12,783) molecular ions are be observed. (B) Nanospray MS spectrum obtained from a solution containing 36 pmoles/μL of Ung dissolved in 50 % acetonitrile/water. The spectrum averages 40 spectra recorded on an ion trap mass spectrometer. A distribution of multiply protonated molecular ions is observed.

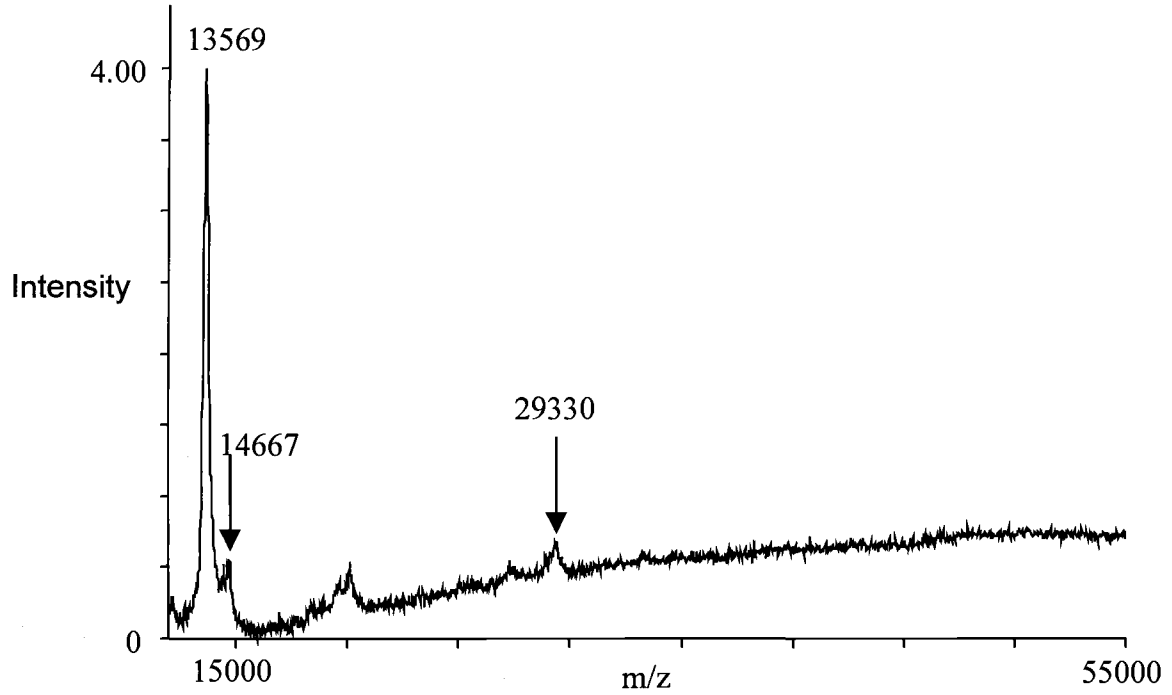


Figure 1.4. MALDI spectrum of hRPA. The purified protein was denatured and solubilized in 50 mM Tris-base buffer (pH=8) at a concentration of about 10 pmoles/μL. hRPA solution was mixed with a saturated solution of HCCA (in 50% acetonitrile) in a 1:2 ratio and analyzed by MALDI-TOF. The signal at *m/z* = 13569 corresponds to the singly protonated molecular ion of the unmodified subunit hRPA14. The signals at *m/z* = 29330 and *m/z* = 14667 are produced by the single protonated and doubly protonated molecular ions of the phosphorylated subunit hRPA32. No signal corresponding to the third subunit (~ 70 kDa) was detected in the corresponding mass range.

1.7. DNA SUBSTRATES

The DNA substrates used in our UV-crosslinking experiments were selected according to the DNA-binding site size of Ung and hRPA and according to the photocrosslinking reactivity of nucleic acids bases [17]. Based on these considerations, oligodeoxythymidine dT_{20} was chosen as our DNA model for Ung, while dT_{30} was preferred for hRPA.

Both oligonucleotides were synthesized and HPLC purified by the Biopolymer Core Facility, at the University of Maryland in Baltimore, and their purity was assessed by MALDI MS. Figure 1.5 shows the spectra recorded for dT_{20} (1.5A.) and for dT_{30} (1.5B.).

1.8. RESEARCH OBJECTIVES

The studies of Barofsky and co-workers demonstrated the feasibility of using mass spectrometry to investigate protein-DNA interactions [1, 3, 11-14, 57]. The analytical methodology associated with analyzing nucleoprotein complexes was established using *E. coli* uracil-DNA glycosylase (Ung) and dT_{20} as a model system [11-13]. One of the objectives of this thesis was to continue to investigate this system and identify crosslinked amino acids from a peptide region (T_{18}) that appeared to have more than one amino acid interacting with the DNA substrate.

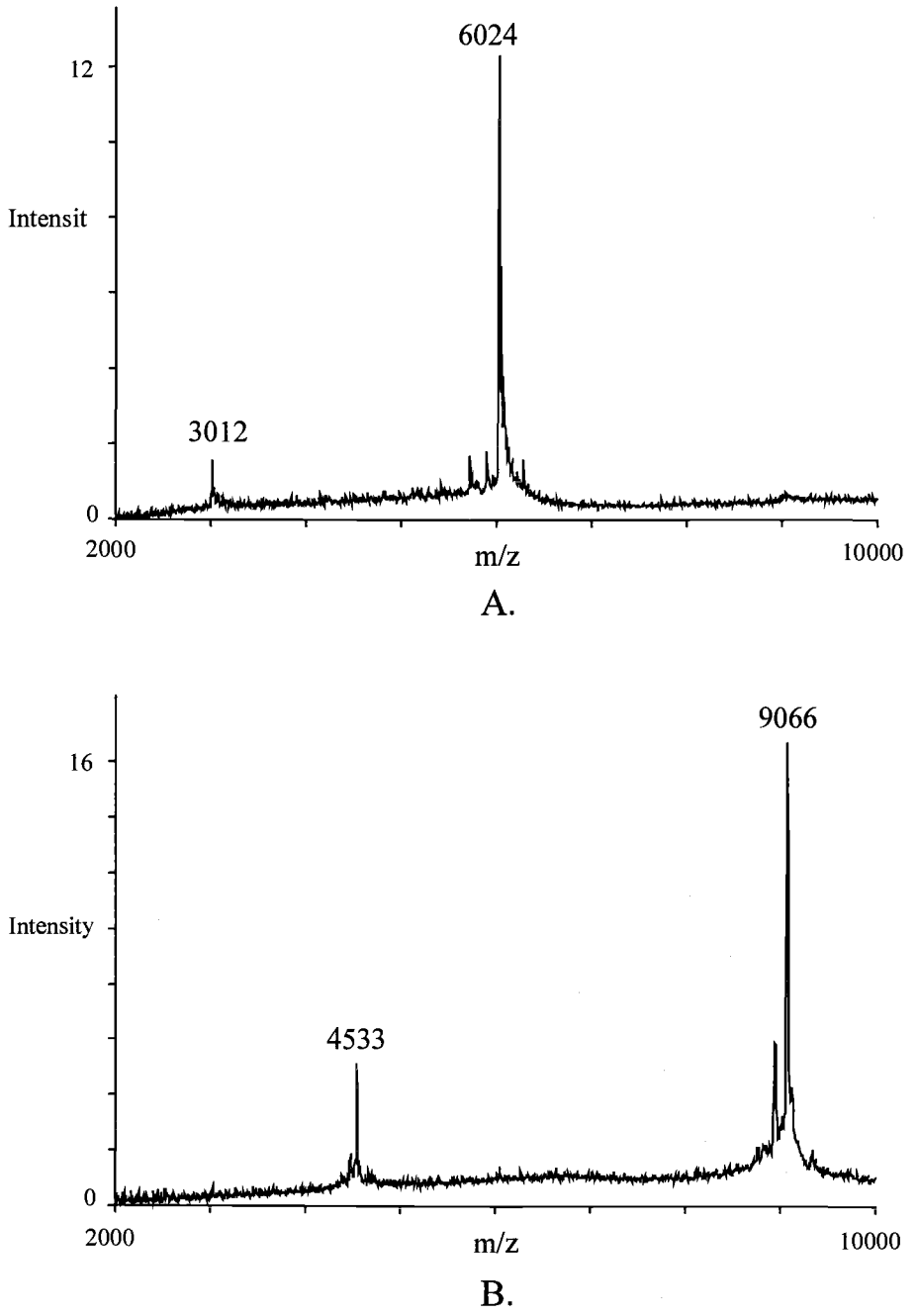


Figure 1.5. MALDI spectra of dT_{20} and dT_{30}. (A) A solution of 108 pmoles/μL of dT_{20} was mixed with a saturated matrix solution of 2,4,6-trihydroxy-acetophenone (THAP) in a 1:10 ratio and analyzed by MALDI. (B) A solution of 45 pmoles/μL of dT_{30} was mixed with a saturated matrix solution of THAP in a 1:5 ratio and analyzed by MALDI. Singly protonated and doubly protonated molecular ions are observed for both oligonucleotides

The results obtained for Ung are presented in Chapter 3 on this thesis.

Another objective was to test the general applicability of our approach for characterizing the DNA-binding domains of other proteins. Chapter 4 describes a study performed in order to investigate the DNA-binding domain of human replication protein A (hRPA).

One of the critical components in achieving success in studying nucleoprotein complexes by mass spectrometry is to employ highly sensitive, high-resolution nanoscale chromatographic techniques coupled to MALDI or electrospray MS. Since no commercial nano-LC systems were available in 1998, at the beginning of this project, I had to design and construct such a system [53]. This task became an intermediate objective for my thesis. The results obtained for the Ung project (Chapter 3) demonstrated that a high-resolution chromatographic system was indeed required to separate three nucleopeptide isomers. The details involved in designing such a system are presented in the second chapter of this thesis.

2. DESIGN OF THE NANO-LC SYSTEM

2.1. INTRODUCTION

The development of highly sensitive matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) mass spectrometric techniques for the analysis of peptides, proteins, and other biomolecules during the last decade generated the concomitant development of miniaturized separation techniques. Liquid chromatography at flow rates under 1 μL/min using packed fused-silica columns with inner diameters less than 100 μm was introduced in the early eighties by McGuffin and Novotny [47] and independently studied over the next dozen or so years by Karlsson and Novotny [58] and by Kennedy and Jorgenson [59,60]. In principle, such minimization should significantly increase mass sensitivity and separation efficiency. By 1991, it had been demonstrated that chromatography on this scale could be directly coupled to a mass spectrometer using a coaxial, continuous-flow, fast atom bombardment interface [61], and only two years later, it was shown that the same could be done with a micro-ESI interface [48,62]. The

term nano-LC was introduced to describe chromatographic separations performed at flow rates in the range of 10-1000 nL/min on packed fused silica columns with inner diameters between 10 and 150 μm [50,61].

In order to exploit the advantages of nano-LC, it is necessary to have a micropumping system that can provide gradient elution at flow rates below 1 μL/min while maintaining accurate proportioning, homogeneous mixing with minimal delay, and reproducible delivery of the mobile phase. Commercial HPLC systems, e.g. the Agilent 1100, that can do this directly have became available very recently, but the most common way to achieve gradient elution in the nano-LC domain at present is to insert a flow-splitter between a conventional HPLC solvent delivery system and a nanocolumn. The inlet ports of most microinjectors designed for capillary LC applications operating at flow rates of 5-10 μL/min have dead volumes between 5-15 μL. In nano-LC applications, dead volumes of this magnitude act as mixing chambers that introduce delays in establishing the gradient and alter the delivery system's gradient profile through solvent-mixing. Rather than taking inordinate measures to minimize the unavoidable dead volumes between the splitting point and the microinjector, one can use simpler, less expensive pumps to deliver solvents and take advantage of the exponential dilution [52] that occurs in the dead volumes to produce useful chromatographic gradients at flow rates less than 1 μL/min.

This alternative was pursued in our present work by constructing an exponential gradient nano-LC system from inexpensive components and examining its potential

for being used in combination with MALDI or Nano-ESI mass spectrometry to separate and analyze the peptides in an enzymatic digest of *Escherichia coli* uracil-DNA glycosylase (Ung). Uracil-DNA glycosylase is one of five proteins that participate in the uracil-excision DNA repair process, the major cellular defense mechanism against spontaneous DNA damage. The enzyme begins the repair process by catalyzing the cleavage of the N-glycosidic bond that joins uracil to the deoxyribose phosphate backbone of DNA [54]. Native Ung, which was the first enzyme of its class to be purified to homogeneity, is a single polypeptide consisting of 228 amino acids (M_r = 25,563). The results of this work demonstrate that an uncomplicated approach to nano-LC can be applied in both on- and off-line applications of protein analysis to produce useful data when more sophisticated but expensive commercial instrumentation is unavailable.

2.2. EXPERIMENTAL

2.2.1. Materials

Peptide standards, acetic acid (AA) and trifluoroacetic acid (TFA) were obtained from Sigma Chemical Co. (St Louis, MO). HPLC grade acetonitrile was purchased

from Fisher Scientific (Pittsburgh, PA). Water was generated with a Mili-Q water purification system (Milipore Corp., Bedford, MA).

2.2.2. Sample Preparation

Uracil-DNA glycosylase expressed in *E. coli* (Ung) was purified as described by Bennett and Mosbaugh [63]. Denatured protein was analyzed by 12.5% polyacrylamide SDS slab-gel electrophoresis according to a previously published method [3]. Protein bands, identified by staining the gel with 300 mM copper chloride for 5 min, were cut out (~100 pmoles protein/gel slice), placed in 1.5 mL Eppendorf tubes, and destained by rinsing 4 times on a vortex mixer with destaining solution (Bio-Rad, Hercules, CA). Each rinsing step was carried out for 15 min. In-gel protein digestion was performed according to a modified procedure introduced by Shevcenko et al. [64]. Gel slices were prepared for in-gel proteolytic digestion by washing them with 100 mM ammonium bicarbonate, dehydrating with 100% acetonitrile, rehydrating with bicarbonate solution, and dehydrating again with 100% acetonitrile. Each gel preparation step was carried out for 15 min with vortexing. After removing the liquid phase and drying under vacuum for 3 hours, the gel-slices were placed in 0.6 mL Eppendorf tubes and covered with 200 μL of digestion buffer (50 mM ammonium bicarbonate, 1 mM calcium chloride) containing 36 ng/μL trypsin (Boehringer, Mannheim, Germany). The tubes were

placed in an ice bath for one hour after which the supernatant was removed and replaced with 50 µl of digestion buffer without enzyme. After allowing enzymatic cleavage to proceed overnight at 37 °C, the volume of the digestion buffer collected from 6 digested gel pieces was reduced to about 50 µL under vacuum to yield a final peptide concentration of 5-10 pmoles/µL. This preparation provided a sufficient peptide sample for several mass spectrometric experiments.

2.2.3. Column Preparation

The nano-LC columns were prepared by packing 20 cm lengths of 360 µm o.d. x 75 µm i.d. fused-silica (Polymicro Technology (Phoenix, AZ) with 5 µm, 300 Å pore, Luna C_{18} silica gel particles (Phenomenex, Torrance, CA) using a pressurized bomb method described by Kennedy and Jorgensen [59]. In short, the end of the silica capillary that is to become the inlet is slipped through a seal into the bomb containing a slurry of isopropanol and the packing material (25 mg/mL); the outlet of the column is inserted into a Valco (Houston, TX) microbore end-fitting containing a 2 µm Valco metallic screen that acts as a temporary outlet frit. The packing material is forced into the capillary by pressurizing the bomb to 1,500 psi; about two hours are required to pack a 20 cm long column. After the gas pressure is slowly released from the bomb, the inlet of the column is connected to an HPLC pump and flushed with acetonitrile for 2 hours and Milli-Q water for another 2

hours. A fused-silica splicer (Fujikura, Japan) is used to prepare the final outlet frit while the column is still wet. The column is left without an inlet frit because this end must be occasionally trimmed to eliminate plugging by very small gel pieces from the in-gel protein digests.

2.2.4. Nano-LC System

Figure 2.1 shows a schematic of the nano-LC system. For LC/UV chromatography, solvents A and B were respectively 0.04 % TFA in water and 0.04 % TFA in 40 % acetonitrile, and for LC/MS chromatography, solvents A and B were respectively 0.1 % AA plus 0.01 % TFA in water and 0.1 % AA plus 0.01 % TFA in 40 % acetonitrile. Since an elution gradient is produced by introducing solvents A and B serially into the two mixing chambers, it suffices to use only one pump at a time to isocratically deliver the respective solvents. For convenience, two conventional HPLC pumps (Kratos Analytical Spectroflow 400, Ramsey, NJ) were used for this purpose. The solvents were each delivered at 30 – 50 μL/min. From the pump, solvent passes through an inline filter (Kratos Analytical, Ramsey, NJ) into a flow-splitter. The filter helps avoid clogging of the system downstream. The flow-splitter comprises a PEEK T-piece (Upchurch Scientific, Murrieta, CA), a 30 μm i.d x 50 cm fused-silica waste-line that functions as a flow resistor, and a 60 μm i.d x 5 cm PEEK connecting line to the first mixing chamber. The split-ratio of this

arrangement is between 1:50 and 1:100 so the flow through the analytical column is in the range 300-1000 nL/min. A ZDV PEEK union (Upchurch Scientific, Murrieta, CA) drilled out to an internal volume of 3 μL was used as the first mixing chamber of the exponential gradient elution system, and the inlet port of the injection valve served as the second mixing chamber.

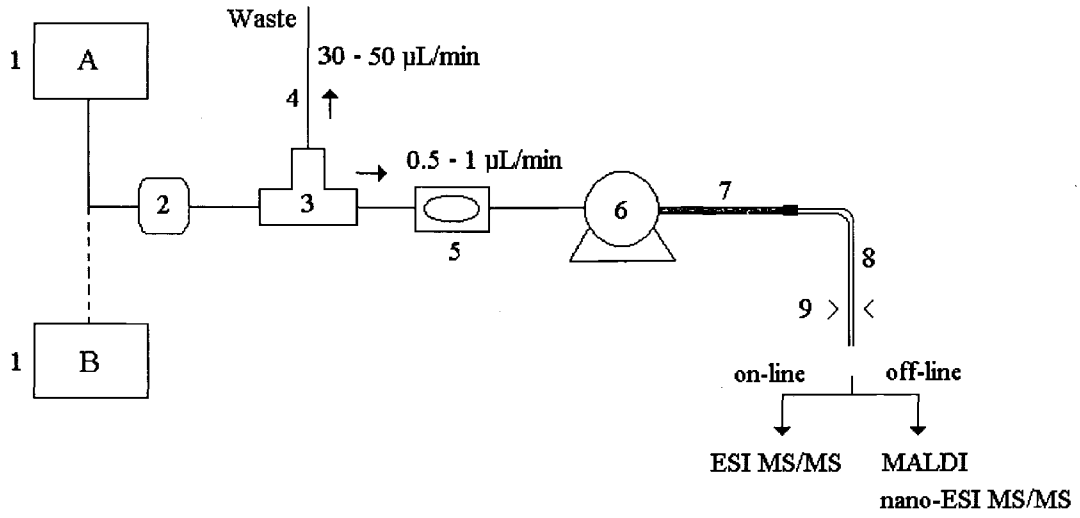


Figure 2.1. Schematic of the nano-LC system: (1) solvent delivery pumps, (2) inline filter, (3) flow splitter, (4) restriction capillary, (5) 3 μL mixing chamber; (6) loop injector (12 μL mixing chamber), (7) nanocolumn, (8) transfer capillary, and (9) UV detector.

A micro-injection valve with a port-volume in the range of 8-12 μL (Rheodyne Model 7520, Rohnert Park, CA) was used with a 1 μL internal sample injection loop for on-line nano-LC/MS whereas an injector with an internal volume of about 10 μL (Rheodyne Model 8125, Rohnert Park, CA) was used with a sample injection loop of 5 μL for off-line nano-LC/MS. PEEK tubing with 60 μm i.d. was used to connect the pumps to the filter, splitter, mixing chamber, and injector. The inlet of the 75 μm i.d. x 20 cm nanocolumn was attached directly into the injector's outlet port. The outlet of the nanocolumn was connected via a 30 μm i.d. x 20 cm fused-silica transfer line to a capillary electrophoresis (CE) unit (Beckmann P/ACE System 2210, Fullerton, CA) to provide UV detection and data acquisition capabilities for the system. The nanocolumn was attached to the transfer capillary with a short piece of Teflon tubing; the length of the transfer capillary from the outlet of the nanocolumn up to the detector was 13 cm. A small diameter hole was drilled in the upper part the CE unit's cartridge (the housing that holds the capillary in place) to allow the packed capillary column to be connected to the transfer line inside the cartridge. A small hole, aligned with the hole in the cartridge, was also drilled in the upper right part of the cartridge interface of the CE unit.

Samples were loaded from the injection valve onto the nanocolumn by connecting pump A to the system and raising the inlet pressure to 200 bar so the sample was displaced from the injection loop by solvent A onto the column at a flow rate of 1.2 μL/min. After allowing sufficient time to completely displace the sample from

the injection loop (~10 min), pump A was stopped and, after allowing the pressure to drop to 1 bar, disconnected from the system. Pump B was then connected to the system, and the inlet pressure was raised to 100 bar to deliver solvent B at 0.5 μL/min.

2.2.5. Mass Spectrometry

MALDI-TOF MS analyses were performed on a custom built instrument [1] equipped with delayed extraction and an ion mirror. For analysis of a protein-digest, 1 μL of a 1:3 (v:v) mixture of peptide solution (0.5 μL) and a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma, St. Louis, MO) in 50 % acetonitrile) was spotted on the MALDI probe. The instrument was operated in the positive ion mode with an accelerating potential of 20 kV and an extraction delay of 500 ns. A spectrum was produced by using *MoverZ* software (ProteoMetrics, LLC, *http://www.proteometrics.com/ software/mz.htm*) to collect and average data generated from 30 laser pulses.

Nano-ESI MS analyses were performed on two different instruments: an API III triple-quadrupole (PE-Sciex, Ontario, Canada) and an LC-Q ion trap (Finnigan, San Jose, CA). A custom built nanospray assembly [40] was designed so that, with minor adjustments, it can be operated on either instrument. For analysis of a protein-digest, 1 μL from each chromatographic fraction was loaded in a nanospray

capillary purchased from Protana (Odense, Denmark). The API-III was operated with the needle set at 750 V, the interface at 100 V, the orifice electrode at 60 V, and the curtain gas flow at 0.4 L/min. In the MS-mode, Q1 was scanned in steps of 0.1 Da with a 1.5 ms dwell time. For operation in the MS/MS-mode, Q1 was set to transmit a 2 Da-mass window centered on the precursor ion, and product ion spectra were recorded by scanning Q3 in steps of 0.2 Da with a 3 ms dwell time. The LC-Q was operated in the MS-mode with the needle potential set to 900 V, the temperature of the heated inlet-capillary at 180 °C, the capillary potential at 46 V, the tube-lens offset-potential at 30 V, and the maximum injection time was 50 ms. For operation in the MS/MS-mode, the maximum injection time was increased to 500 ms, the isolation width was set to 1 Da, and the collision energy was tuned individually for each peptide to obtained the best possible MS/MS spectra. MS/MS spectra from both instruments were interpreted with the aid of MS-Tag software (University of California at San Francisco, Mass Spectrometry Facility, *http://prospector.ucsf.edu/*).

On-line nano-LC/ESI MS was performed using a custom built micro-electrospray source installed on the LC-Q instrument. A short piece of Teflon tubing was used to couple the packed fused-silica column to a New Objective (Cambridge, MA) fused-silica, coated needle (360 μm o.d. x 50 μm i.d. x 5 cm, 30 μm tip). During operation, the needle potential was set at 2.6 kV.

2.3. RESULTS AND DISCUSSION

2.3.1. On-line Nano-LC/MS

A continuous exponential gradient from 0 to 100% solvent B will emerge from a chamber initially filled with solvent A if the later is displaced from the chamber by an equal volume of solvent B delivered at a constant flow rate [52]. More specifically, the concentration $B_{single}(t)$ of solvent B in the mobile phase exiting a single mixing chamber of volume V_1 is given at any time t, after solvent B begins to enter the chamber, by the exponential expression

$$B_{single}(t) = B_0 (1 - e^{-t/\tau_1}) \qquad (2.1.)$$

where B_0 is the concentration of solvent B in the mobile phase used to displace solvent A from the mixing chamber and $\tau_1 = V_1 / F$ is the time it takes to displace solvent A. When two mixing chambers with volumes V_1 and V_2 respectively are connected in series (as in Figure 2.2), $B_{single}(t)$ at the outlet of the first mixing chamber becomes the initial concentration of the mobile phase entering the second chamber. In this case, the differential equation for the rate of increase of B in the

mobile phase exiting the second chamber can be shown to reduce to the following simple formula:

$$\frac{dB_{double}(t)}{dt} = \frac{B_{single}(t) - B_{double}(t)}{\tau_2} \qquad (2.2.)$$

where $\tau_2 = V_2 / F$.

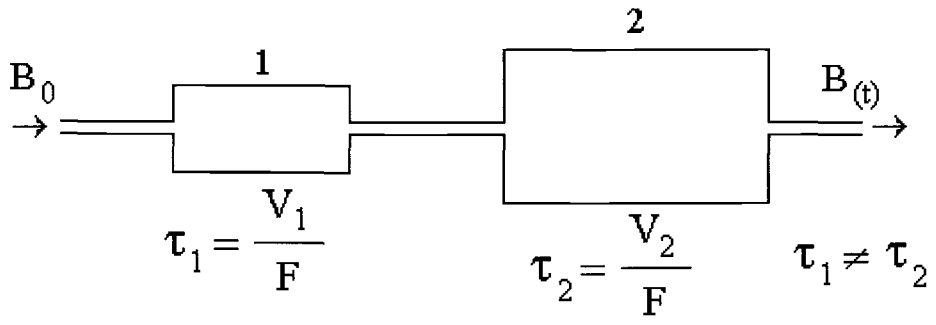


Figure 2.2. Two mixing chambers having different volumes V_1 and V_2 connected in series. τ_1 and τ_2 are the times it takes to displace volumes V_1 and V_2 of solvent A from chambers 1 and 2 respectively.

Solving equation (2.2.) yields the following formula for the time dependent concentration $B_{double}(t)$ of solvent B exiting the second mixing chamber:

$$B_{double}(t) = B_0 \left[1 - \frac{1}{\tau_2 - \tau_1}\left(\tau_2 e^{-\frac{t}{\tau_2}} - \tau_1 e^{-\frac{t}{\tau_1}}\right)\right] \qquad (2.3.)$$

Figure 2.3A shows the theoretical shapes obtained for 4 different gradients produced at a flow rate of 0.5 μL/min by two mixing chambers. In all four cases, the volume of the first mixing chamber was 3 μL, but the volume of the second mixing chamber was respectively 5, 10, 12, and 15 μL. Figure 2.3B exhibits the corresponding experimental profile obtained by monitoring the increase in UV absorbance as solvent B spiked with 10% acetone (v/v) elutes from the nano-LC system pictured in Figure 2.1. The onset of the curve in Figure 2.3B indicates a delay of around 10 min in establishing the gradient. Comparison of Figure 2.3A and 2.3B, with allowance for the 10 min delay in time in the experimental curve, indicates a close match between the shape of the experimental gradient and the shape of the theoretical gradient with V_1 = 3 μL and V_2 = 10 μL. This degree of agreement signifies that mixing of the solvents in both chambers was nearly homogeneous.

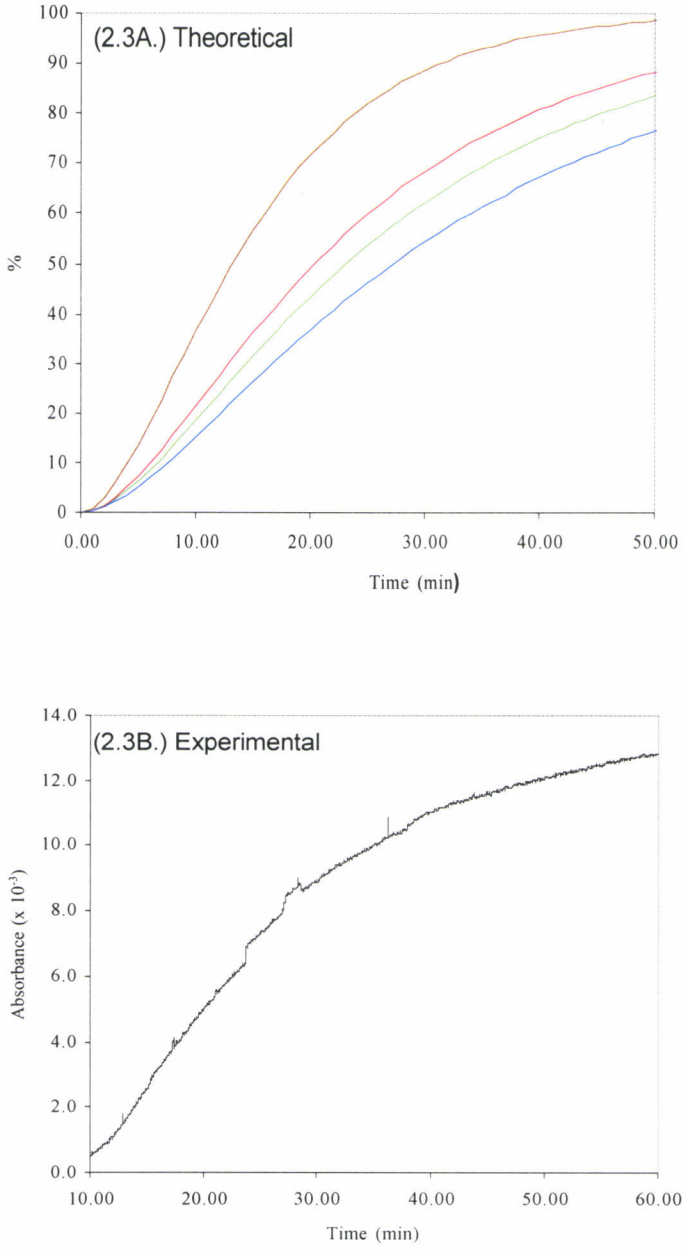


Figure 2.3. Mobile phase gradients formed by diluting Solvent A with Solvent B in the double mixing chamber system shown in Figure 2.1. (2.3A) Theoretical profiles generated by programming Microsoft Excel 6.00 to plot $B_{double}(t)$ as given by Equation 2.3 with $B_0 = 100\%$, F = 0.5 µL/min $V_1 = 3$ µL, and $V_2 = 5$ µL(brown trace), 10 µL (red trace), 12 µL (green trace), and 15 µL(blue trace). (2.3B) Experimental profile produced at 0.5 µL/min (pressure at inlet filter = 100 bar) after connecting solvent B:10% acetone (v/v) and monitoring the absorbance of the acetone in the eluant at 254 nm (0.014 AU = full scale absorbance for solvent B:10% acetone).

Gradient reproducibility was measured by three consecutive injections of a Ung tryptic digest sample. The nano-LC/MS base peak chromatograms obtained are shown in Figure 2.4.

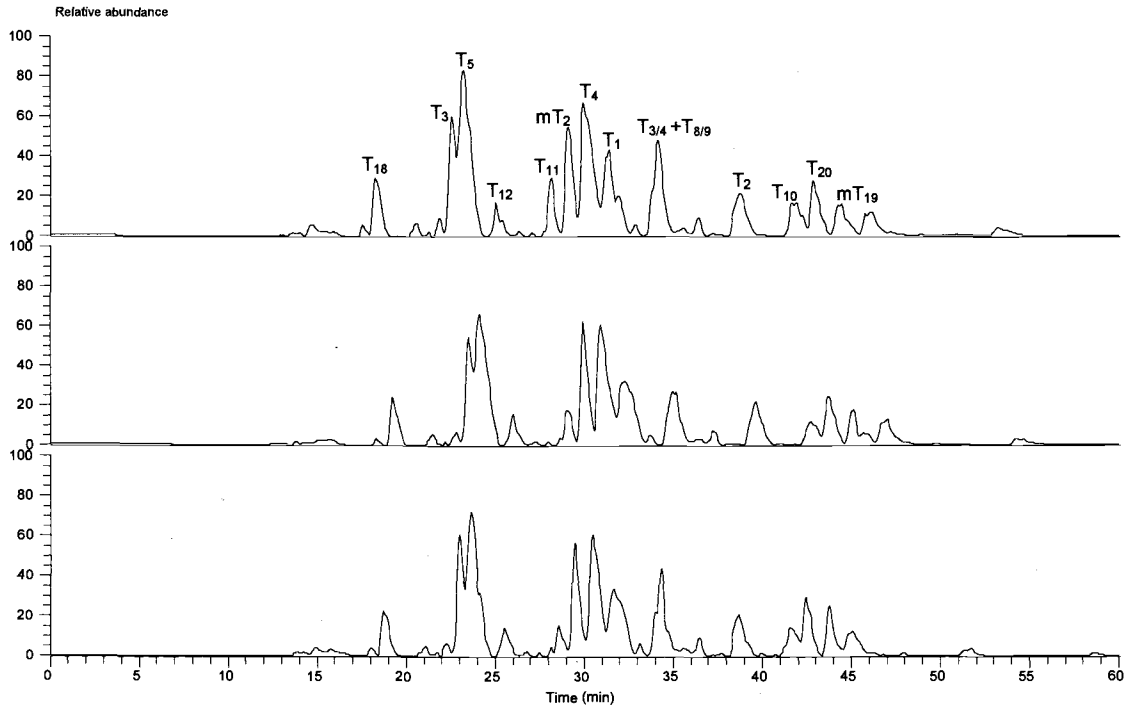


Figure 2.4. Nano-LC/ion trap base-peak ion-chromatograms of a tryptic digest of Ung obtained for three consecutive runs. Based on Ung's 20 theoretically possible tryptic fragments, the peptides are labeled T_n (n=1,20) from the N terminal; peptides containing missed tryptic cleavage sites are labelled $T_{n/n+1}$ (n=1,20). Each chromatogram was produced by injecting 1 µL of tryptic digest (5-10 pmoles/µL onto the column (75 µm i.d x 20 cm; 5 µm C_{18}) and eluting at 0.5 µL/min (pressure at inlet filter = 100 bar) with an exponentially produced gradient of 0% to ~40% Solvent B (0.1 % AA plus 0.01% TFA in 40 % acetonitrile). Ion detection was from *m/z* 300 to 2000.

Peptide samples serve as a sensitive monitor for gradient reproducibility because even minor changes in gradient composition result in detectable variations in retention time [65]. The adjusted retention time and the corresponding mean and relative standard deviation (RSD) for each of seven major peptides seen in the chromatograms are listed in Table 2.1. Although the RSD obviously increases systematically with increasing retention time, it remains below ± 0.3 % (n = 3) for adjusted retention times up to at least 20 minutes. Thus, the reproducibility of the gradient profile from run to run is analytically useful and indicates a high level of performance for the entire nano-LC system.

	Peptide						
Run	T3	T5	T11	mT2	T4	T8/9	T2
1	4.3	4.93	9.86	10.75	11.76	15.63	19.99
2	4.26	4.88	9.83	10.74	11.69	15.93	20.42
3	4.32	5	9.92	10.87	11.92	15.89	20.46
Mean (min)	4.29	4.94	9.87	10.79	11.79	15.82	20.29
RSD (%)	0.03	0.06	0.05	0.07	0.12	0.16	0.26

Table 2.1. Adjusted retention times in minutes calculated for seven major Ung tryptic peptides and the corresponding means and relative standard deviations (RSD). (Adjusted retention time = retention time for n^{th} peptide – retention time for T18 peptide.)

The nano-LC/UV chromatogram of the same sample under identical gradient conditions is exhibited in Figure 2.5. The sensitivity is not as great as would be obtained with a conventional UV-detector, because a very short path length (30 μm) is used for the nano-LC system. The baseline absorbance in this chromatogram gradually decreases with increasing time as acetonitrile from solvent B replaces water from solvent A.

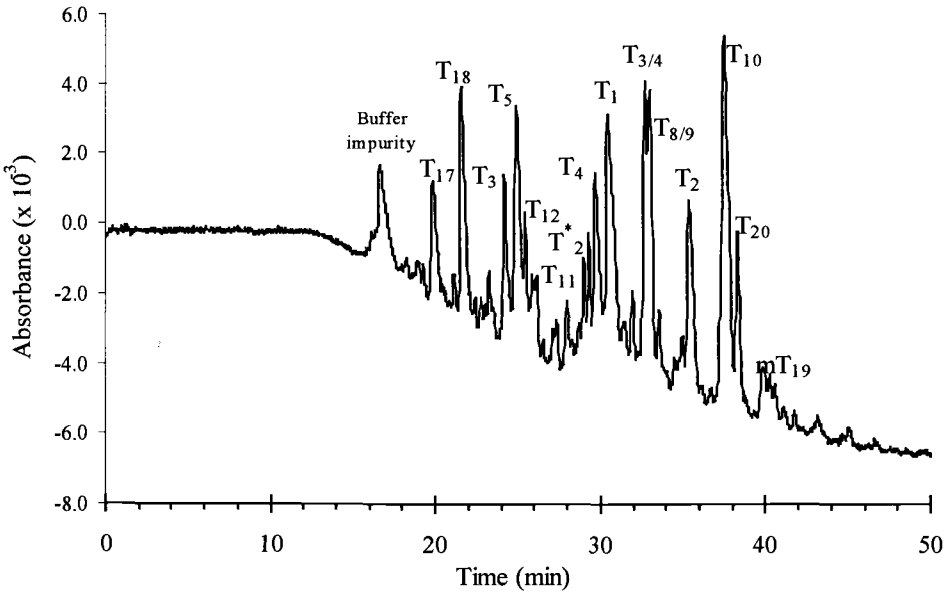


Figure 2.5. Nano-LC/UV chromatogram of a Ung tryptic digest. Based on Ung's 20 theoretically possible tryptic fragments, the peptides are labeled T_n (n=1,20) from the N terminal; peptides containing missed tryptic cleavage sites are labelled $T_{n/n+1}$ (n=1,20). The chromatogram was produced by injecting 1 μL of tryptic digest (5-10 pmoles/μL) onto the column (75 μm i.d x 20 cm; 5 μm C_{18}) and eluting at 0.5 μL/min (pressure at inlet filter = 100 bar) with an exponetially produced gradient of 0% to ~40% Solvent B (0.04% TFA in 40 % acetonitrile). UV detection was at 200 nm.

The peaks in the ion chromatograms are broader than the corresponding peaks in the UV chromatogram for two reasons. First, the level of TFA in the ion pairing agent employed in the LC/MS runs (0.1 % AA plus 0.01 % TFA) was lower than in the LC/UV runs (0.04 % TFA) because TFA suppresses peptide ion formation in electrospray ionization and, thus, degrades sensitivity. With less TFA present in the LC/MS mobile phase, ion pairing among the peptides is reduced, and consequently the peptide's chromatographic peaks are broader. Second, an additional dead volume is introduced into the LC/MS runs through the coupling between the nanocolumn and the spraying needle.

Of the 20 tryptic fragments theoretically possible for Ung, 15 were identified from the mass spectra that compose the ion chromatograms in Figure 2.4. As is typical for tryptic digests, the missing peptides are either very short (e.g., T_{14} is K and T_{16} is QR) or very long (e.g., the molecular weight of $T_{6/7}$ is over 3,000 Da). The small peptides cannot be efficiently trapped during injection onto the column, and the big peptides (> 3,000 Da) are not extracted into the digestion buffer. The incomplete digestion of $T_{3/4}$ is presumably due to steric hindrance whereas that of peptides $T_{6/7}$ and $T_{8/9}$ is due to the well known fact that trypsin does not efficiently cleave R-P bonds. The only cysteine in Ung, which appears in T_{19}, was acrylamide modified [3].

2.3.2. Off-line Nano-LC/MS

Numerous instances are still encountered in mass spectrometric analyses of proteins in which limited sample amounts or diluted sample solutions make it necessary to isolate or concentrate certain peptides contained in complex proteolytic mixtures in order to analyze them further by MALDI-TOF MS and nano-ESI MS/MS. In particular, MS/MS analysis of a peptide requires that the collision energy be tuned properly in order to obtain adequate sequencing information for identification. Generally in nano-LC/MS, the width of a peptide's elution profile at the outlet of the column is in the range of 7-15 s; this is too little time to allow for optimal adjustment of the MS/MS collision energy for a peptide as it elutes off the column. Consequently, an enzymatic digest of a protein is usually systematically analyzed at different collision energies in a progression of experiments in order to obtain sufficient sequence information from as many of the peptides as possible. This problem, which is associated with the narrowness of the chromatographic peaks, is compounded when peptides coelute. To overcome these problems, David and Lee introduced the concept of variable-flow nanoscale separation in which peptides are separated at 200 nL/min but MS/MS analysis is performed at a "peak parking" rate of 25 nL/min [51]. When a digest contains a particularly large number of peptides or a cluster of several peptides that only partially separate on the column, this approach can fail to produce the desired sequencing data. In such situations, nanospray MS/MS analysis can be more

practicable. The nanospray signal from a single sample of a protein digest typically lasts for periods of 20-30 min, which is generally ample time to determine the optimum collision energy and record the MS/MS spectrum for each of many peptides [39,40,66,67]. If a digest contains a larger than usual number of peptides however, even this method can be improved by fractionating the sample off-line on a nano-LC column and loading the individual fractions into nanospray tips.

Figure 2.6 shows a chromatogram produced by overloading the column with 10 μL of a Ung tryptic digest. The broad widths of the peaks are a clear indication of the excessive amount of sample on the column. In spite of this gross condition, peptidic separation was sufficient to permit collection of 15 well defined peptide fractions. Each of these fractions was analyzed by MALDI MS and nano-ESI MS to assess the degree of peptide-isolation and to tentatively assign peptide-identity. The MALDI and nanosprayed ion trap mass spectra of Fractions number 3,7,8, and 12 are displayed in Figure 2.7. Except for a general tendency to shift the distribution of charge states upward, the triple-quadrupole produced nanosprayed spectra (data not shown) that are similar in appearance to those recorded on the ion trap.

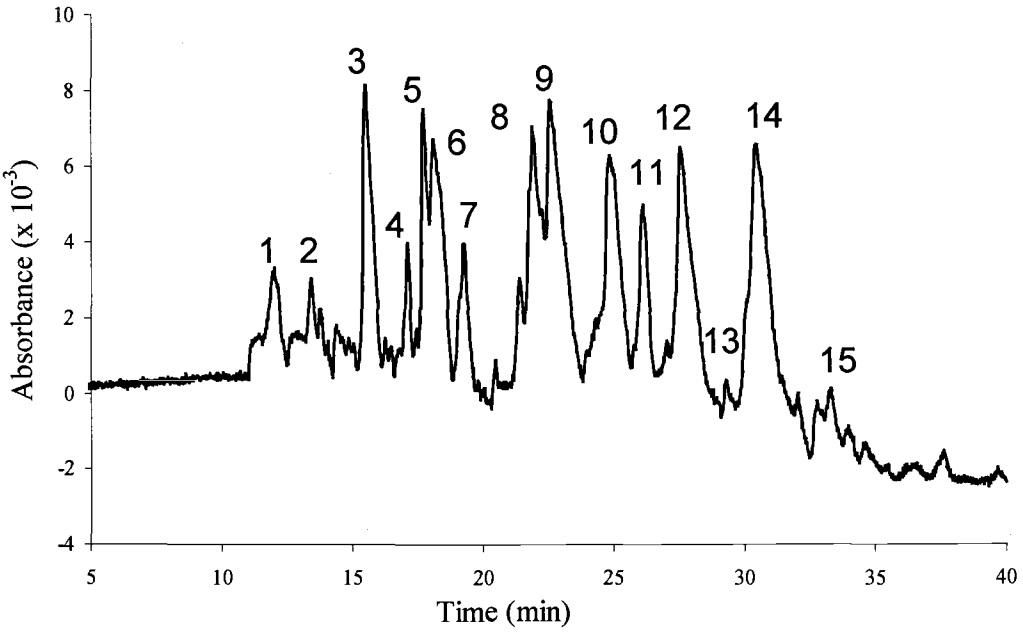


Figure 2.6. Nano-LC/UV chromatogram produced by loading the column with 10 μL of a tryptic digest of Ung in two, consecutive, 5-μL injections. The sample was eluted at 0.5 μL/min (pressure at inlet filter = 100 bar) with an exponentially produced gradient of 0% to ~40% Solvent B (0.04% TFA in 40 % acetonitrile). UV detection was at 200 nm. Fifteen fractions (labeled 1 through 15 in the order of collection) were collected manually; fraction volume varied from 2-3 μL depending on the collection period (2 to 4 min). 1 μL of solvent A (0.4 % TFA in water) was added to each collection vial before starting the manual collection.

Comparison of Figure 2.6 with Figure 2.5 suggests that Fractions 3 and 12, which were collected respectively from segments of the chromatogram corresponding to well isolated components, should contain essentially only a single peptide whereas

Fractions 7 and 8, which were taken from segments corresponding to crowded, incompletely separated components, should contain at least a couple of peptides. The mass spectra shown in Figure 2.7 are in accordance with this expectation. Specifically, the spectra indicate the presence of only Ung-T_{18} in Fraction 3 (Figure 2.7A) and, for most part, Ung-T_2 in Fraction 12 (Figure 2.7D) but of Ung-T_4 plus a non-tryptic component in Fraction 7 (Figure 2.7B) and Ung-T_1 plus Ung-T_{19} in Fraction 8 (Figure 2.7C). There were no traces of contamination in either the MALDI or nano-ESI mass spectra of any of the 15 fractions. The MALDI spectra shown in Figure 2.7 suggest that the nano-LC separation technique described in this paper might be particularly well suited to on-line coupling to a MALDI source through a vacuum deposition interface like that recently introduced by Preisler et al. [68,69].

Figure 2.7. MALDI and Nano-ESI mass spectra obtained from Fraction 3 (2.7A.), Fraction 7 (2.7B.), Fraction 8 (2.7C.), and Fraction 12 (2.7D.) respectively of the chromatographic collection indicated in Figure 2.6. The Nano-ESI spectra were recorded on the ion trap mass spectrometer.

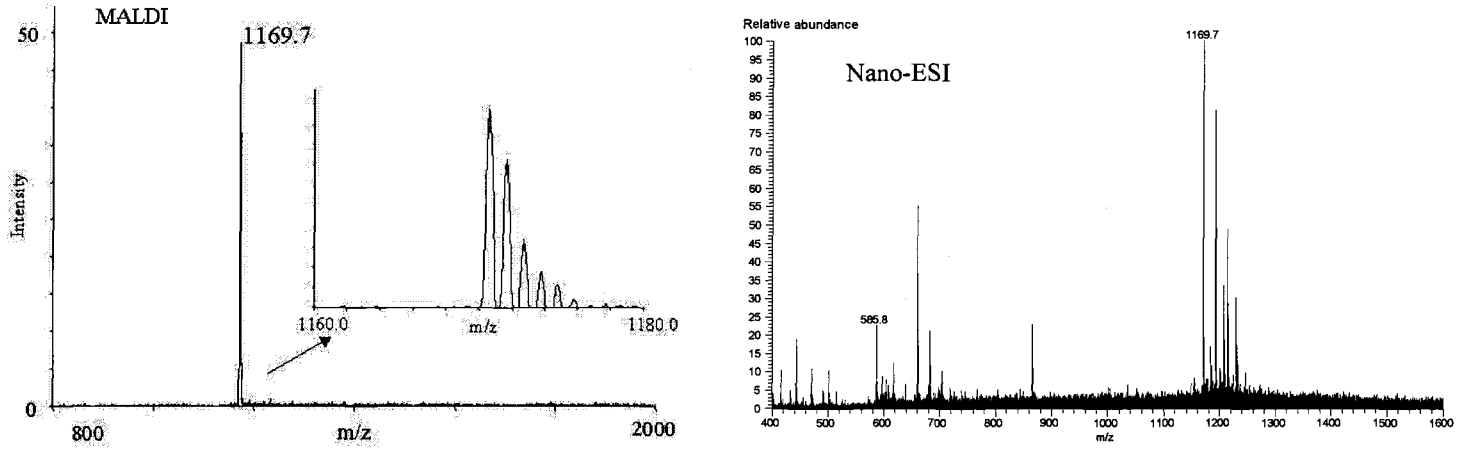


Figure 2.7A. MALDI and nano-ESI mass spectra obtained from Fraction 3.

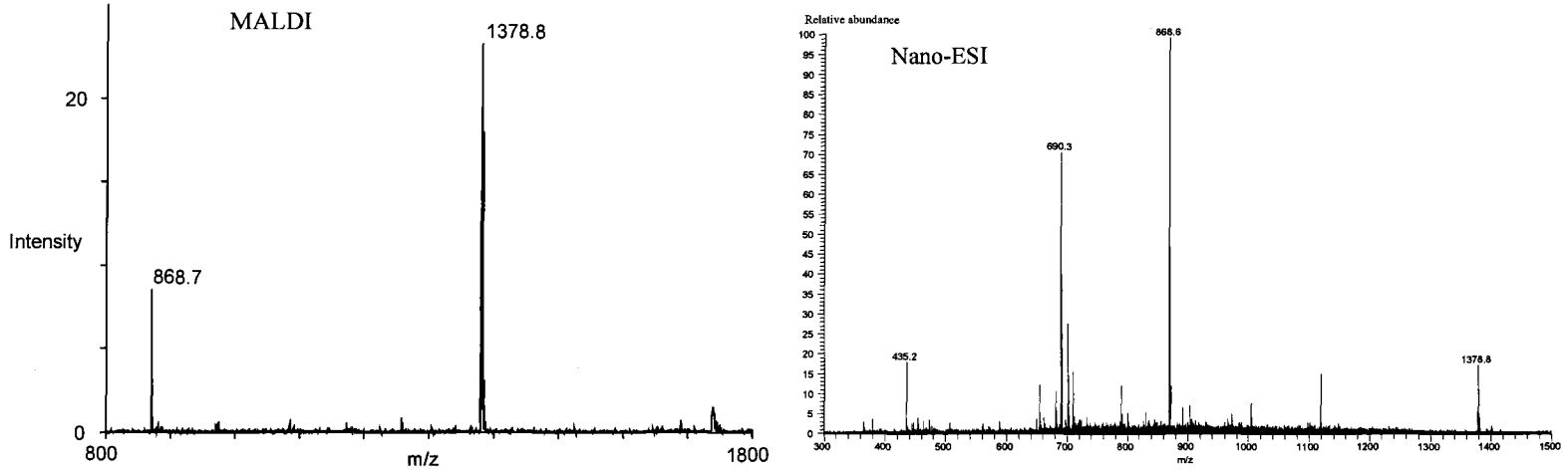


Figure 2.7B. MALDI and Nano-ESI mass spectra obtained from Fraction 7.

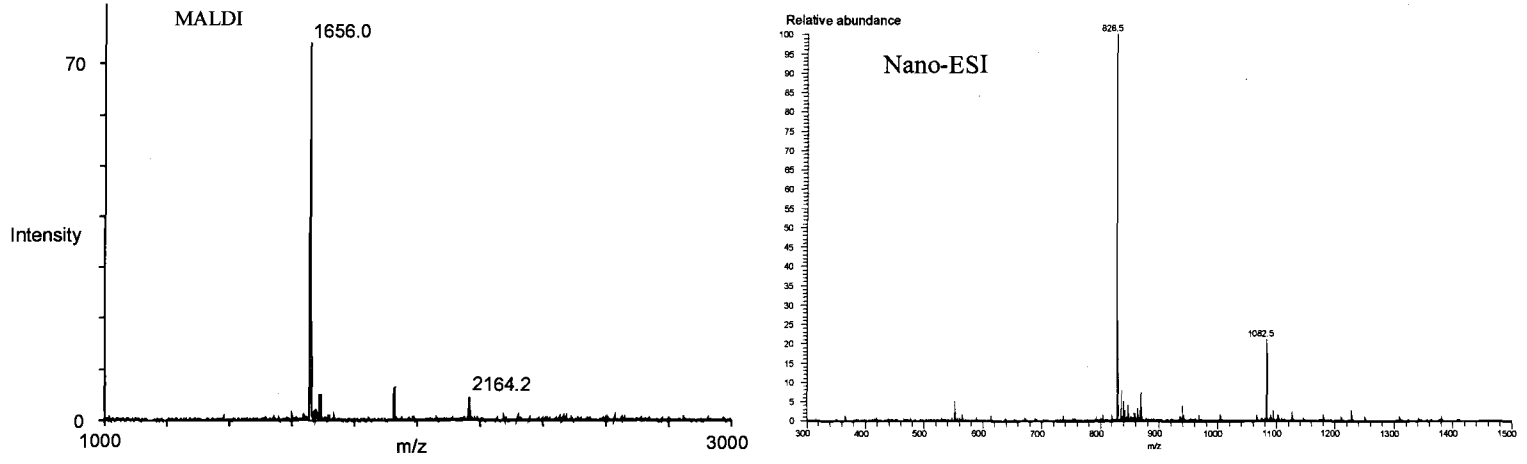


Figure 2.7C. MALDI and Nano-ESI mass spectra obtained from Fraction 8.

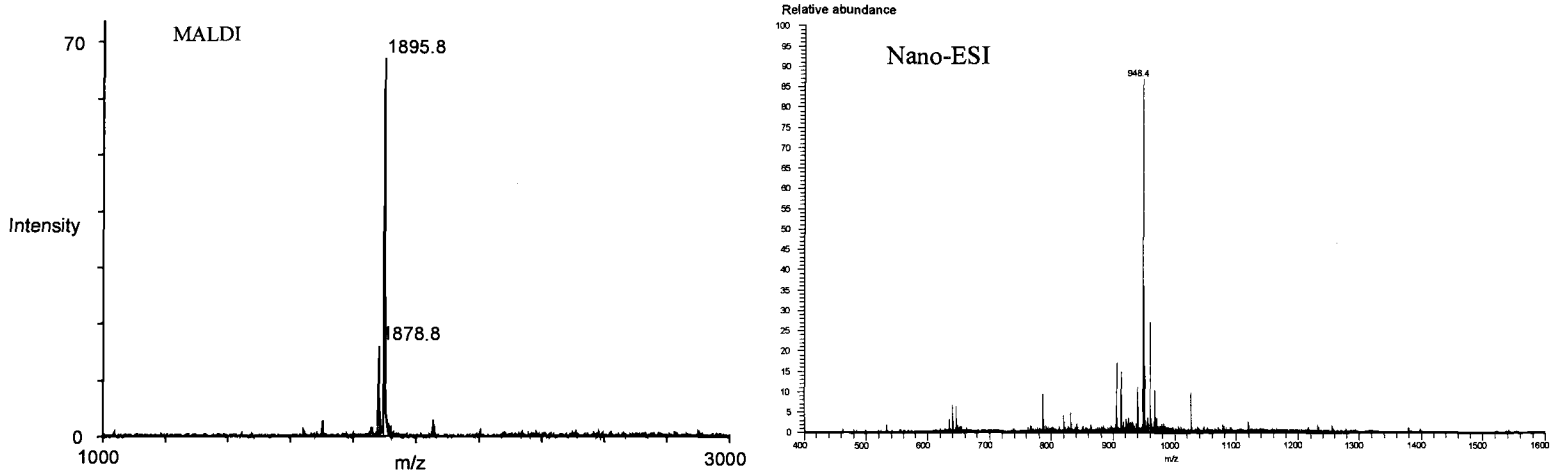


Figure 2.7D. MALDI and Nano-ESI mass spectra obtained from Fraction 12.

The ESI spectra in Figure 2.7 demonstrate that the nano-LC system fractionates complex mixtures into clean samples for nano-ESI MS even under conditions of excessive sample loading.

Nano-ESI MS/MS spectra of the components in all 15 of the nano-LC fractions of the Ung-tryptic digest were recorded on both ion trap and the triple-quadrupole instruments to confirm peptide-identity. Figure 2.8A displays the MS/MS spectra obtained off both instruments from the single component contained in Fraction 3. Although the ion trap spectrum was produced from the singly charged molecular ion (m/z = 1169.7) and the triple quadrupole spectrum was obtained from the doubly charged molecular ion (m/z = 585.4), both mass spectra exhibit predominantly *y*-ions. As can be seen from the fragmentation diagram atop the ion trap spectrum, the information generated from this particular peptide, despite the instrument's low-mass cutoff at ~m/z 300 [70], is sufficient to assign the sequence of all of its amino acid residues and identify it as T_{18} (APHPSPLSAHR). By contrast, the sequence information obtained off the triple-quadrupole in this case is incomplete. The later would, nonetheless, be sufficient to identify the peptide by searching a database. The non-tryptic peptide in Fraction 7 was one of several unknowns fortuitously provided by the Ung tryptic digest. Figure 2.8B shows the nanosprayed MS/MS spectra of the doubly charged ion (m/z = 690.3) of this molecule obtained respectively from the ion trap and the quadrupole instruments.

Figure 2.8. Ion trap and triple-quadrupole MS/MS spectra obtained from Fraction 3 (2.8A.) and Fraction 7 (2.8B.) respectively of the chromatographic collections indicated in Figure 2.6. The precursors for the ion trap spectrum (50 % collision energy) and triple quadrupole spectrum (20 eV collision energy) from Fraction 3 were respectively the singly charged ion at *m/z*=1169.7 and the doubly charged ion at *m/z*= 585.4. The precursor for the ion trap spectrum (20 % collision energy) and triple quadrupole spectrum (25 eV collision energy) from Fraction 7 was the doubly charged ion at *m/z*= 690.3. MS-Tag analyses of the sequencing data provided by the mass spectra are summarized in the fragmentation diagrams displayed at the top of each spectrum.

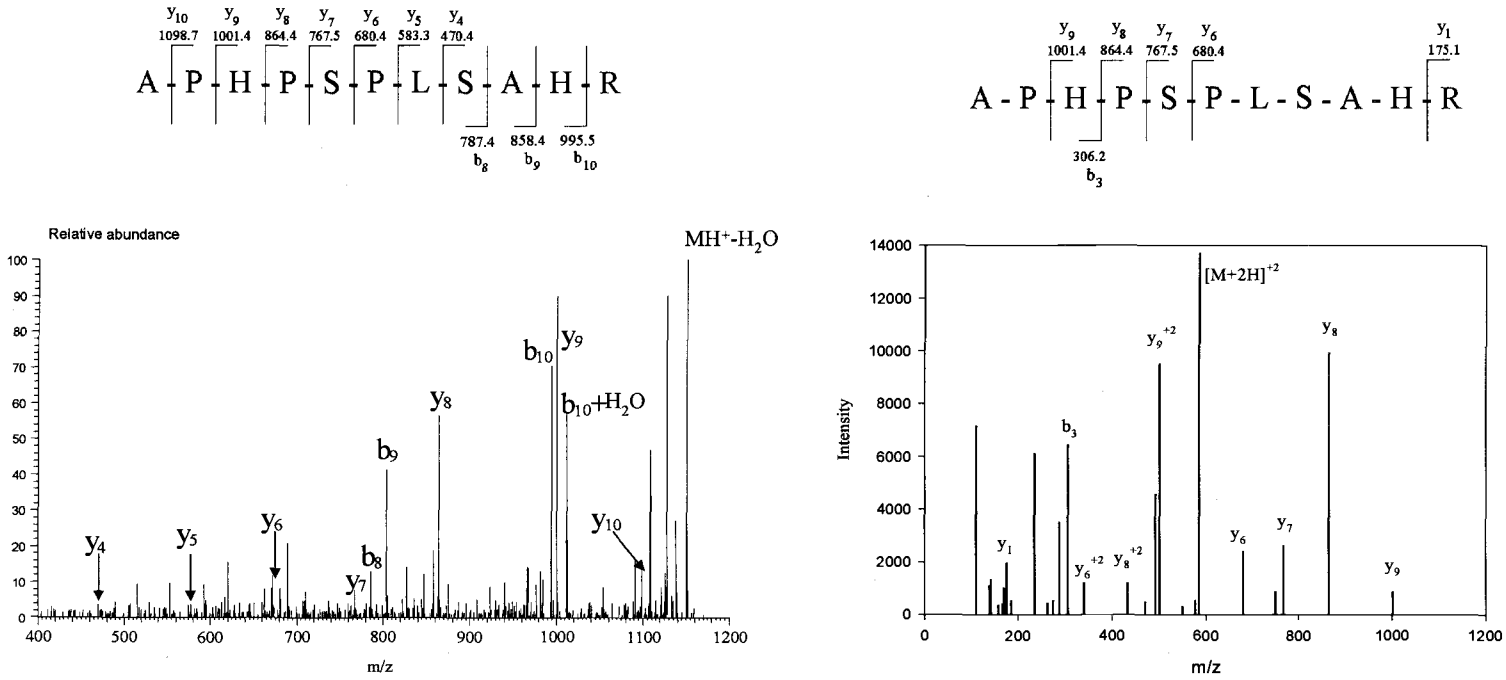


Figure 2.8A. Ion trap and triple-quadrupole MS/MS spectra obtained from Fraction 3.

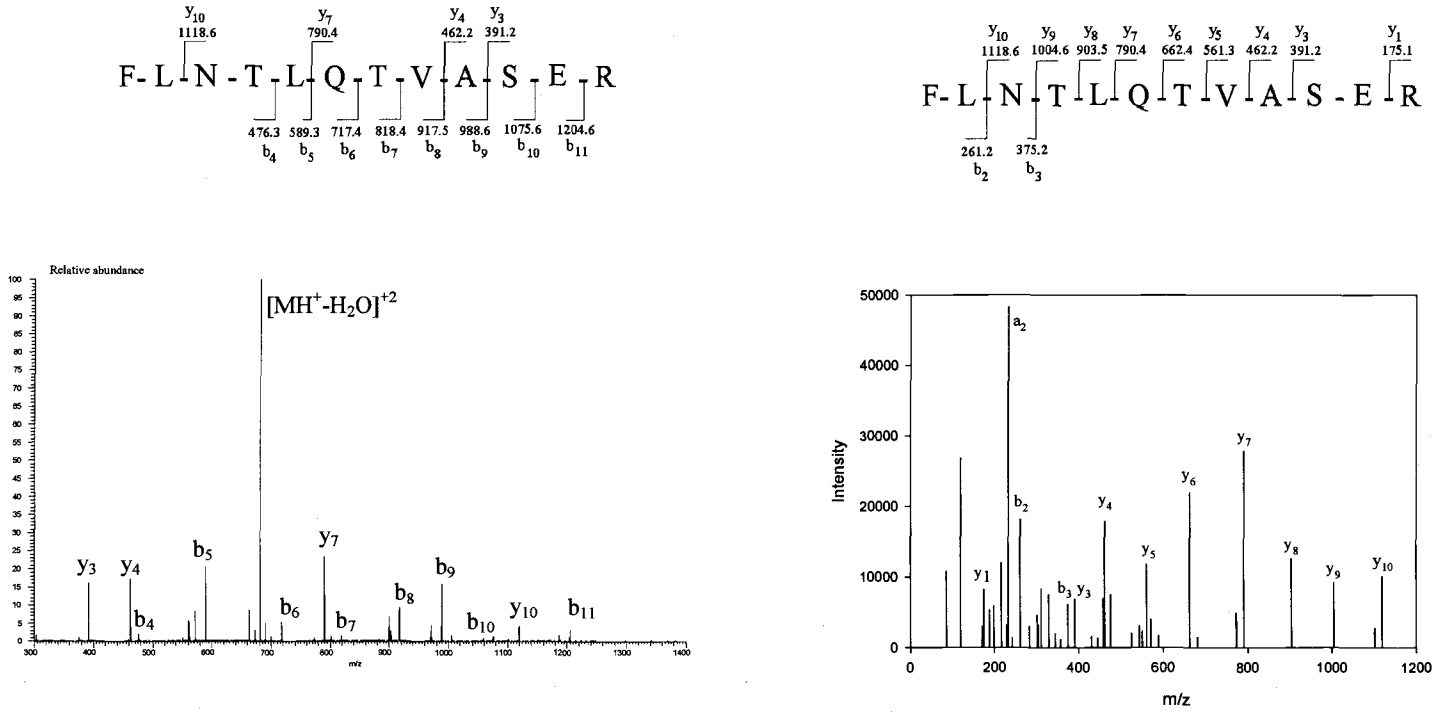


Figure 2.8B. Ion trap and triple-quadrupole MS/MS spectra obtained from Fraction 7.

The MS/MS spectrum produced by the ion trap contains almost exclusively *b*-ions while that recorded on the triple quadrupole exhibits mostly *y*-ions. As seen from the fragmentation diagrams for these two mass spectra, the sequence information from either spectrum was adequate for MS-Tag to identify the unknown peptide as the fragment (FLNTLQTVASER) that is formed by the loss of the first four amino acid residues (QQPY) from the N terminal of T_2. MS/MS analysis (data not shown) was used in a similar manner to verify that the relatively intense ion signal at m/z = 1878.8 in the MALDI mass spectrum of Fraction 12 (Figure 2.7D) corresponds to the pyroglutamic acid form of T_2 created by the elimination of ammonia from its N terminal glutamine.

An off-line batch technique cannot match an on-line alternative for convenience, but in those not infrequent instances where a protein defies routine, directly coupled HPLC/ESI MS/MS analysis, off-line nano-LC/nano-ESI MS/MS can provide a critical, complementary aid to identifying the protein. The two examples described in the preceding paragraph indicate how nano-LC might be used in situations where isolation of specific peptides from complex proteolytic digests is necessary in order to perform nano-ESI MS/MS or possibly MS^n under optimum collision conditions. ESI analyses in the MS and MS/MS modes typically requires mid-femtomole to low picomole amounts of sample on both ion trap and triple quadrupole instruments - lower levels are possible but not routine at present [70].

Fractionation by nano-LC does not measurably degrade this sensitivity because opportunity for sample contamination is minimal and additional concentrating steps are unnecessary.

3. IDENTIFICATION OF AMINO ACID RESIDUES IN THE DNA-BINDING DOMAIN OF URACIL-DNA GLYCOSYLASE FROM *E. COLI*

Reproduced with permission from Doneanu, C.E., Gafken, P.R., Bennett, S.E., and Barofsky, D.F, *Anal. Chem.*, submitted.

3.1. INTRODUCTION

A central aim of molecular biology is to elucidate the molecular interactions that occur between proteins and nucleic acids during replication, recombination, dynamic alteration and repair of DNA, and transcription and regulation of genes. Mass spectrometry in combination with photochemical/chemical crosslinking, along with X-ray diffraction and NMR spectroscopy, is considered a direct method that provides information about the structural relationship between proteins and nucleic acids. The strategy, which combines low intensity UV crosslinking and mass spectrometry, was initially demonstrated in 1994 [3]. The coupling of photocrosslinking to mass spectrometry for protein-DNA analysis has become increasingly popular in recent years [7-14,107]. Photochemically crosslinking nucleic acid binding proteins to their substrates is an established procedure for characterizing protein-nucleic acid interfaces [17-19]. This technique uses photons

as "zero-length" crosslinkers to induce the formation of covalent bonds between certain amino acids that are in close contact with nucleic acid bases. It is assumed that only amino acids located in the protein binding site are involved in the formation of these bonds. For the identification of amino acids that take part in the crosslinking, enzymatic digestion of the UV-irradiated nucleoprotein complex is used to produce free peptides and peptides crosslinked to the nucleic acid substrates (nucleopeptides). After separation from the non-crosslinked peptides, the crosslinked peptides are analyzed by MALDI-TOF in order to obtain their molecular weights. By subtracting the mass of the nucleic acid from the mass of a nucleopeptide and comparing the resulting mass to a peptide map of the non-crosslinked protein, one can identify a region in the protein containing the peptides involved in crosslinking. This straightforward strategy can be used only if the chemistry of the photochemical reaction is known. In the final stage of analysis, the nucleopeptides are sequenced by tandem mass spectrometry (MS/MS) in order to identify the amino acids attached to nucleic acid bases.

In this paper, photochemical crosslinking and mass spectrometry have been used to investigate the DNA-binding domain of *E. coli* uracil-DNA glycosylase (Ung), one of many enzymes involved in DNA repair processes. DNA repair is an essential process for maintaining the integrity of the genome. In the simplest type of DNA repair, the damaged base is repaired directly, e.g. dealkylated, by a one-step mechanism rather than being excised and replaced by a correct one. However, most lesions in DNA are repaired by the much more complex recombination repair or

excision repair systems. The latter include nucleotide excision repair (NEP), mismatch repair and base excision repair (BER). As the name implies, the initial step in BER is the removal of a base by DNA glycosylases rather than removal of an entire nucleotide. DNA glycosylases are relatively small monomeric proteins that do not require cofactors for their activity; therefore, they are good models for studying interactions between DNA and proteins. Uracil-DNA glycosylase is an enzyme involved in the uracil excision repair mechanism. It is widespread and highly conserved among biological species. A uracil base is introduced into DNA either when a deoxyuridine triphosphate (dUTP) is incorporated during DNA replication instead of a thymidine triphospahte (TTP), or when a cytosine is deaminated. About 10^4 uracil bases are incorporated into human DNA following S-phase DNA synthesis, and spontaneous cytosine deamination occurs at a rate of ~ 200 events per human cell per day [54]. Cytosine deamination is a major mutagenic force because it produces a CD to AT transition. The major cellular defense mechanism against this form of spontaneous DNA damage is a multistep uracil-excision DNA repair process. In the first step of this repair pathway, uracil-DNA glycosylase catalyzes the hydrolysis of uracil residues from single-stranded or double-stranded DNA, producing free uracil and abasic DNA. The abasic site is subsequently removed by a 5'-acting apurinic/apyrimidinic (AP) endonuclease and a deoxyribophosphodiesterase, leaving a gap that is filled by DNA polymerase and closed by DNA ligase.

Uracil-DNA glycosylase from *E. coli* (Ung) was the first DNA glycosylase discovered [71]. Subsequently, the *E. coli ung* gene was cloned and the protein was overexpressed [72]. The deduced nucleotide sequence indicates that the protein is a monomer containing 228 amino acid residues and has a relative molecular weight of 25,563 Da. It is believed that Ung's primary function is to recognize and remove uracil from U/G mispairs resulting from the deamination of cytosine [54,73,74].

The 3D structure of Ung has been investigated by X-ray crystallography. The first published crystal structure (obtained at a resolution of 3.2 Å) was for the enzyme co-crystallized with the uracil-DNA glycosylase inhibitor protein (Ugi) [75]. The experiment revealed that Ung is made up of a four-stranded parallel β-sheet and 11 separate α-helices. Shortly after, the crystal structure of the Ung-Ugi complex was solved for the native enzyme and a H187D mutant [76]. Finally, a high-resolution crystal structure of free Ung (at 1.6 Å resolution), and of Ung bound to uracil (1.5 Å resolution) became available [77]. In addition to crystallography, Raman spectroscopy [78] and NMR [79] have been used to investigate the mechanism of uracil removal by Ung. It has been shown that in the case of human uracil-DNA glycosylase, the protein undergoes a conformational change and "flips" the uracil base out of the major groove of the DNA helix into the active site where enzymatic interactions with the uracil base and deoxyribose can be made [80,81]. The same mechanism has been observed for Ung [77], and a comparison of the amino acid sequence alignment between *E. coli* and human uracil-DNA glycosylase has revealed that Ung possesses 12 of the same 13 residues attributed to DNA

interaction in the human enzyme [75,76]. These results indicate that Ung closely resembles its human counterpart and that it should be regarded as a prototypical enzyme. According to the crystal structure of Ung bound to uracil [77], a uracil-recognition pocket is formed at the active site of the enzyme by the following residues: Q63, D64, Y66, F77, N123 and H187. In addition to this active site, hydrogen bond interactions observed between several serine residues (S88, S189 and S192) and the phosphodiester groups located on a single strand of DNA, appear to contribute to the base flipping mechanism [82]. Three other amino acids (H134, L191 and R195) have been shown to interact with the uracil substrate [77].

In previous studies performed in our lab, we used photochemical crosslinking in combination with mass spectrometry to investigate the DNA-binding domain of Ung [1,3,11-13]. In the {Ung : dT_{20}} nucleoprotein complex, we identified three tryptic peptides (T_6, T_{11} and T_{18}) crosslinked to deoxythymidine dT_{20} [3,11,12]. Within T_6, LC/ESI-MS/MS definitely located Y66 and H67 as sites of crosslinking [11,12]. In the present paper, we report our research efforts to identify crosslinked amino acids within T_{18}.

3.2. EXPERIMENTAL

3.2.1. Chemicals

Acetic acid (AA), trifluoroacetic acid (TFA), ampicillin, bromophenol blue, calcium chloride, copper (II) chloride, ethylenediaminetetraacetic acid (EDTA), formic acid, HEPES, sodium chloride, streptomycin sulfate, Trizma (Tris-base) and Tris-HCl were purchased from Sigma Chemical Co. (St Louis, MO). Acrylamide (>99% pure), ammonium persulfate, 2-mercaptoethanol, bis N,N`-methyl-bis-acrylamide, and N,N,N`N`-tetramethylethylenediamine (TEMED) were obtained from Bio-Rad (Hercules, CA). Ammonium sulfate, dithiothreitol (DTT), glycine, isopropyl-β-δ-thiogalactopyranoside (IPTG), and sodium dodecyl sulfate (SDS) were bought from Life Technologies Inc. (Grand Island, NY). HPLC grade acetonitrile, ammonium acetate, ammonium bicarbonate, ammonium hydroxide (50%), glucose, hydrochloric acid, isopropanol, methanol, potassium chloride, sodium carbonate, and sodium hydroxide were supplied by Fisher Scientific (Pittsburgh, PA). Glycerol and potassium phosphate (dibasic and monobasic salts) were purchased from J.T. Baker (Phillipsburg, NJ). Agar, tryptone and yeast extract were supplied by Becton Dickinson (Franklin Lakes, NJ). Water was generated with a Milli-Q water purification system (Millipore Corp., Bedford, MA). Oligodeoxythymidine dT_{20}, was synthesized by the Biopolymer Core Facility, at the University of Maryland at Baltimore. To desalt the dT_{20}, the lyophilized sample

(approximately 5 μmole) was resuspended in 1 mL of 10 mM triethylammonium bicarbonate buffer (pH 7.2) and applied to a 1.79 cm^2 × 10 cm Sephadex G-25 column equilibrated in the same buffer. The column was operated at 20 mL/h and 1.5-ml fractions were collected. Fractions were monitored spectrophotometrically at 260 nm and the elution peak was pooled and quantified by using a molar extinction coefficient for dT_{20} of $\varepsilon_{260} = 1.7 \times 10^5$ L/mol•cm. The pooled fraction was analyzed by MALDI MS to check for incomplete synthesis products and then subdivided into 1 or 1.5-mL aliquots that were dried by vacuum centrifugation. Each aliquot was dissolved in DAB buffer (30 mM Tris-HCl (pH 7.4), 1 mM dithiothreitol, 1 mM EDTA, 5% (w/v) glycerol) to a final concentration of 242 pmoles/μL.

3.2.2. Purification of Uracil-DNA Glycosylase (Ung)

The purification of Ung was based on a procedure previously described [3]. *E. coli* JM105/pSB1051 were grown on LB plates (1% tryptone, 0.5% yeast extract, 1% NaCl, and 1.5% agar) containing 0.01% ampicillin at 37 °C. An isolated colony was selected and grown overnight in LB medium (1% tryptone, 0.5% yeast extract, and 1% NaCl) containing 0.01% ampicillin at 37 °C. The saturated culture was used to inoculate 6 liters (1:100) of LB medium containing 0.01% ampicillin. The culture was shaken at 37 °C until the cell density reached an OD_{600} of 0.8 at which

time IPTG was added to a final concentration of 1 mM. The induced culture was allowed to grow for an additional 3 h and the cells were collected by centrifugation in a Sorvall GSA rotor at 5,000 rpm for 15 min at 4 °C. The cell pellets were stored at –80 °C until the start of purification.

The following purification steps were carried out at 4 °C. Cell pellets were thawed and thoroughly resuspended in 240 mL of buffer TED (50 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 1 mM dithiothreitol). Cells were lysed by sonication, and cellular debris was pelleted by centrifugation in a Sorvall SA-600 rotor at 10,000 rpm for 15 min. Supernatant fractions were pooled, and an equal volume of 1.6% (w/v) streptomycin sulfate in TED was added dropwise with stirring over 30 min. Stirring was allowed to continue for an additional 30 min, after which time the solutions were cleared by centrifugation in a Sorvall SA-600 rotor at 10,000 rpm for 15 min. The supernatant was adjusted to 70% ammonium sulfate to precipitate the protein, and the mixture was allowed to stir for an additional 30 min to ensure complete precipitation. The precipitate was collected by centrifugation in a Sorvall SA-600 rotor at 10,000 rpm for 15 min. The precipitate was resuspended in 4 mL of UE buffer (10 mM Hepes-KOH (pH 7.4), 10 mM 2-mercaptoethanol, 1 mM EDTA, 1 M NaCl, 5% (w/v) glycerol) and dialyzed overnight against UE buffer. The dialyzed sample was applied to a Sephadex G-75 column (6 cm^2 × 88 cm) equilibrated in UE buffer, eluted with the same buffer, and collected into 7-mL fractions. Fractions containing uracil-DNA glycosylase activity were pooled, diafiltrated with 1.4 L of HA buffer (10 mM potassium phosphate (pH 7.4), 1 mM

dithiothreitol, and 200 mM KCl), and concentrated to ~30 mL with an Amicon stirred cell (Millipore) under 55 psi of N_2 gas using a YM10 membrane (Millipore). The sample was applied to a hydroxylapatite column (4.9 cm^2 × 15 cm), equilibrated in HA buffer, eluted with the same buffer, and collected into 5-mL fractions. Fractions containing uracil-DNA glycosylase activity were pooled and dialyzed overnight against DAB buffer. The sample was applied to a single-stranded DNA agarose column (19 cm^2 × 44 cm) equilibrated in DAB buffer, and 7-mL fractions were collected. After washing with 5 column volumes of DAB, the enzyme was eluted with DAB containing 150 mM NaCl. Fractions containing enzyme activity were analyzed by denaturing polyacrylamide gel electrophoresis to monitor the level of purity in each fraction, and those fractions considered to be greater than 90% pure for uracil-DNA glycosylase were pooled, diafiltrated against 1.4 L of buffer DAB, and concentrated to 2.5 mg/ml using an Amicon stirred cell under 55 psi of N_2 gas and a YM10 membrane. The protein concentration was determined spectrophotometrically by using a molar extinction coefficient of ε_{280} = 4.2×10^4 L/mol•cm for Ung. The purified protein was stored at –80 °C until used.

3.2.3. Preparative Crosslinking of Ung to dT_{20}

Purified Ung (36 nmol) was combined with a 3-fold molar excess of dT_{20}, and DAB buffer was used to bring the final volume to 1 mL. The mixture was placed in

a 4-mL, capped, quartz, cuvette (NSG Precision Cells, Farmingdale, NY) and equilibrated on ice for 15 min prior to crosslinking. The cuvette was placed lengthwise on a bed of ice, and irradiated with UV light (λ_{max}=254 nm) in a Stratalinker 1800 (Stratagene, La Jolla, CA) for 15 min. This solution was either used immediately or stored at -80 °C until future use.

3.2.4. Isolation of Tryptic Peptides Crosslinked to dT_{20}

Aliquots of the irradiated crosslinking mixture were loaded onto a 10% SDS polyacrylamide gel and subjected to gel electrophoresis. After removing the stacking gel, the lower 3-cm portion of the resolving gel corresponding to the region of free dT_{20} was removed, and the remaining portion of gel was copper stained [83]. Bands corresponding to Ung × dT_{20} were cut from the gel with a clean, sharp razorblade and placed into 1.5 mL Eppendorff tubes. Gel slices were in-gel digested with trypsin (36 ng/uL) following the procedure of Shevchenko et al. [64]. After 2 h of digestion at 37 °C, 400 µL of digestion buffer was added to the gel slices, and digestion was continued overnight. After overnight digestion, the aqueous digestion solution was removed and saved. The organic extraction mixture of FAPH (50% formic acid, 25% acetonitrile, 15% isopropyl alcohol, and 10% water) was added to each tube until the gel pieces were covered, and the tubes were vortexed at room temperature for 4 h. The extraction mixture was removed from each tube and saved. Gel slices were then crushed and dehydrated with 200 µL of

acetonitrile for 10 min. The acetonitrile was removed, with care not to remove any gel pieces, and added to the saved FAPH extraction solutions. The combined organic solutions were concentrated to dryness by vacuum centrifugation. The saved aqueous digestion solution was adjusted to 10 mM Tris-HCl (pH 7.2), 1 mM EDTA, and 100 mM sodium chloride (TE-100), and the dried organic extraction fractions were resuspended with 250 μL TE-100. Both the adjusted aqueous digestion solution and the resuspended organic extraction solution were applied to a NAC-52 anion exchange cartridge (Life Technologies Inc.) that had been activated with 3 mL TE-2000 (TE buffer containing 2000 mM sodium chloride) and equilibrated with 5 mL TE-100. After sample loading, the cartridge was washed with 5 mL TE-100 and the peptide × dT_{20} fragments (nucleopeptides) were eluted with 1 mL TE-1000 directly into a Centricon-3 (Millipore) desalting cartridge. The Centricon-3 was centrifuged at 6000 × g. The eluate was desalted by washing the Centricon-3 sequentially with 4.5 mL of water, 4.5 mL of 10 mM ammonium acetate, and 3 mL of water. The washed sample was concentrated to 200 μL and then transferred to a 750-μL Eppendorff tube and further concentrated to about 10 μL by vacuum centrifugation.

3.2.5. Nuclease P1 Digestion

Nuclease P1 (Amersham Biosciences, Piscataway, NJ) was diluted with 50 mM ammonium acetate pH=6.5 (0.04 activity units/μL). Digestion was conducted at 37 ^{o}C for 4 h, after mixing 1 μL of nuclease P1 solution with 1 μL of nucleopeptide sample.

3.2.6. Column Preparation

The nano-LC columns were prepared by packing 20 cm lengths of 360 μm o.d. x 75 μm i.d. fused-silica (Polymicro Technology (Phoenix, AZ) with 5 μm, 300 Å pore, Luna C_{18} silica gel particles (Phenomenex, Torrance, CA) using the pressurized bomb method described by Kennedy and Jorgensen [60]. In short, the end of the silica capillary that is to become the inlet is slipped through a seal into the bomb containing a slurry of isopropanol and the packing material (25 mg/mL); the outlet of the column is inserted into a Valco (Houston, TX) microbore end-fitting containing a 2 μm Valco metallic screen that acts as a temporary outlet frit. The packing material is forced into the capillary by pressurizing the bomb to 1,500 psi; about two hours are required to pack a 20 cm long column. After the gas pressure is slowly released from the bomb, the inlet of the column is connected to an HPLC pump and flushed with acetonitrile for 2 hours and Mili-Q water for

another 2 hours. A fused-silica splicer (Fujikura, Japan) is used to prepare the final outlet frit while the column is still wet.

For on-line nano-LC/ESI MS a fused-silica column with an integral frit - "PicoFrit" (360 μm o.d. x 75 μm i.d. x 40 cm, 15 μm tip) obtained from New Objective (Cambridge, MA) was packed using the same procedure described above. The column was connected to an injector (Rheodyne Model 8125, Rohnert Park, CA) placed in a plastic box, and the spraying voltage (2.5 kV) was applied to the injector after the sample was injected.

3.2.7. Nano-LC System

The nano-LC system employed uses the exponential dilution method to produce gradient separations. A detailed description of the operating principle and system setup is provided elsewhere [53]. Elution gradients are produced by exponential dilution in two mixing chambers. Two conventional HPLC pumps (Kratos Analytical Spectroflow 400, Ramsey, NJ) were used, one to deliver solvent A (0.04 % TFA in water for LC/UV runs or 0.1 % AA plus 0.01 % TFA for LC/MS runs) and the other to deliver solvent B (0.04 % TFA in 40 % acetonitrile for LC/UV runs or 0.1 % AA plus 0.01 % TFA for LC/MS runs). From the pump, solvent is passed through an inline filter (Kratos Analytical, Ramsey, NJ) into a flow-splitter. The filter helps avoid clogging of the system downstream. The flow-splitter

comprises a PEEK T-piece (Upchurch Scientific, Murrieta, CA), a 30 μm i.d x 50 cm fused-silica waste-line that functions as a flow resistor, and a 60 μm i.d x 5 cm PEEK connecting line to the first mixing chamber. The split-ratio of this arrangement is between 1:50 and 1:100 so the flow through the analytical column is in the range 300-1000 nL/min. A ZDV PEEK union (Upchurch Scientific, Murrieta, CA) drilled out to an internal volume of 3 μL was used as the first mixing chamber of the exponential gradient elution system, and the inlet port of the injection valve served as the second mixing chamber. A micro-injection valve with a port-volume of about 10 μL (Rheodyne Model 8125, Rohnert Park, CA) was used with a sample injection loop of 5 μL for off-line nano-LC/UV runs and on-line LC/MS runs. PEEK tubing with 60 μm i.d. was used to connect the pumps to the filter, splitter, mixing chamber, and injector. The inlet of the 75 mm i.d. x 20 cm nanocolumn was attached directly into the injector's outlet port. The outlet of the nanocolumn was connected via a 30 μm i.d. x 20 cm fused-silica transfer line to a capillary electrophoresis (CE) unit (Beckmann P/ACE System 2210, Fullerton, CA) in order to provide UV detection and data acquisition capabilities for the system. The nanocolumn was attached to the transfer capillary with a short piece of Teflon tubing; the length of the transfer capillary from the outlet of the nanocolumn up to the detector was 13 cm. A small diameter hole was drilled in the upper part the CE unit's cartridge (the housing that holds the capillary in place) to allow the packed capillary column to be connected to the transfer line inside the cartridge. A small

hole, aligned with the hole in the cartridge, was also drilled in the upper right part of the cartridge interface of the CE unit.

For LC/MS runs, the column was connected directly to the injector placed in a plastic box, and the spraying voltage was applied directly to the injector after loading the sample.

Samples were loaded from the injection valve onto the nanocolumn by connecting pump A to the system and raising the inlet pressure to 200 bar so the sample was displaced from the injection loop by solvent A onto the column at a flow rate of 1.2 μL/min. After allowing sufficient time to completely displace the sample from the injection loop (~10 min), pump A was stopped and, after allowing the pressure to drop to 1 bar, disconnected from the system. Pump B was then connected to the system, and the inlet pressure was raised to 100 bar to deliver solvent B at 0.5 μL/min.

3.2.8. Mass Spectrometry

MALDI-TOF MS analyses were performed on a custom built instrument [1] equipped with delayed extraction and an ion mirror. The MALDI probe was spotted with 1 μL of a 1:3 (v:v) mixture of nucleopeptide sample (0.5 μL) and a saturated solution of 2,4,6- trihydroxyacetophenone (THAP) (Sigma, St. Louis, MO) in 50 % acetonitrile). The instrument was operated in the positive ion mode with an

accelerating potential of 20 kV and an extraction delay of 500 ns. A spectrum was produced by using *MoverZ* software (ProteoMetrics, LLC, *http://www.proteometrics.com /software/mz.htm*) to collect and average data generated from 30 laser pulses.

On-line nano-LC ESI MS runs were performed on an LC-Q ion trap (Finnigan, San Jose, CA) mass spectrometer. The LC-Q was operated in the MS-mode with the spraying potential set to 2.5 kV (applied to the injector), the temperature of the heated inlet-capillary at 180 °C, the capillary potential at 46 V, the tube-lens offset-potential at 30 V, and a maximum injection time of 50 ms. For operation in the MS/MS-mode, the maximum injection time was increased to 500 ms, the isolation width was set to 1 Da, and the collision energy was set to 30 %. MS/MS spectra were recorded only for the doubly charged molecular ions of nucleopeptides.

3.3. RESULTS

3.3.1. MALDI Spectra of Nucleopeptides

A preparative crosslinking reaction mixture containing Ung and dT_{20} (1 : 3 molar ratio) was UV irradiated and separated by 12.5% SDS-PAGE. Two gel bands corresponding to the free protein (~ 25 kDa) and to the Ung x dT_{20} nucleoprotein complex (~ 30 kDa) were revealed after copper staining [83]. The bands

corresponding to the nucleoprotein complex were excised and in-gel digested with trypsin. Crosslinked tryptic peptides (nucleopeptides) were separated from non-crosslinked peptides using a procedure based on anion exchange chromatography developed in our lab. According to this procedure, peptides and nucleopeptides produced from the in-gel digestion were extracted from gels using an organic extraction mixture. This organic solution was concentrated to near dryness by vacuum centrifugation. The sample was redissolved in a low ionic strength buffer and loaded on a anion exchange column. Nucleopeptides were selectively retained on this column while tryptic peptides were washed away. Finally, nucleopeptides were eluted with a high ionic strength buffer, desalted, and concentrated. Samples prepared in this manner are suitable for mass spectrometric analysis.

The MALDI spectrum of the purified mixture of Ung × dT_{20} nucleopeptides (Figure 3.1A) showed the presence of four ion species at *m/z* 7193, 7810, 7967 and 8102. The average molecular weight of dT_{20} (6024 Da) was subtracted from the measured molecular weights resulting in the peptide masses of 1169 Da, 1786 Da, 1943 Da and 2078 Da. From previous work [11,12], these peptide masses are known to correspond to T_{18} (1169 Da), $T_{18/19}$ (1786 Da), T_{11} (1943) and $^{*}T_6$ (2078, T_6 minus the last 3 amino acids from the C-terminal). Based on this evidence, it was concluded that the following peptide regions from Ung are involved in binding DNA: T_6 (58-VVILGQDPYHGPGQAHGLAF-77), T_{11} (130-AGQAHSHASL GWETFTDK-147) and T_{18} (185-APHPSPLSAHR-195).

Previous crosslinking studies performed in our lab using poly-dT substrates [3,11,12] indicated that nucleopeptide masses were very close to the sum of the neutral oligonucleotide mass plus the neutral peptide mass. This finding is strong evidence for the formation of "zero-length" covalent bonds between DNA and proteins after low intensity UV-irradiation [17].

The nucleopeptide isolation procedure is essential in order to achieve reliable information by MALDI MS. Specifically, the nucleopeptide sample must be free of non-crosslinked peptides and DNA substrate. If both free peptides and dT_{20} are present, the MALDI process can form adducts between them, and these adducts can be mistaken for nucleopeptides. The MALDI spectrum exhibited in Figure 3.1B was produced after mixing a peptide fraction containing T_{17} and T_{18} with dT_{20}. The signals observed at *m/z* 6656 and *m/z* 7193 correspond to MALDI adducts produced by the addition of a dT_{20} molecule to each peptide. These adducts are artifacts of the MALDI ionization process and, as such, do not reflect the actual composition of the sample. The spectrum also shows the signal corresponding to the free T_{18} peptide (*m/z* 1169.6), but the signal corresponding to the T_{17} peptide (*m/z* 633) was obscured by the matrix ions normally present in MALDI spectra below 700 Da. Clearly, it is not possible to distinguish between real nucleopeptides present in the sample and MALDI artifacts unless the nucleopeptide sample is free of peptides or free of dT_{20}. Purification is the most challenging aspect in the MALDI analysis of nucleopeptides.

Figure 3.1. MALDI spectra of nucleopeptides. (A) Ung nucleopeptides were purified from in-gel digestions of 30 bands corresponding to the Ung x dT_{20} nucleoprotein complex. An aliquot (1 μL) of the purified nucleopeptide sample was mixed with THAP matrix in a 1:3 ratio and analyzed by MALDI MS. The spectrum shows the presence of four nucleopeptides (*m/z* 7193, 7810, 7967, 8102). (B) MALDI adducts produced by mixing Ung tryptic peptides with dT_{20} in a molar ratio of about 1:10. Tryptic peptides T_{17} and T_{18} were isolated in the same fraction by preparative scale nano-LC and mixed with dT_{20}. An aliquot (1 μL) of this sample was analyzed as above by MALDI MS. The signal observed at *m/z* 6656 corresponds to an adduct of dT_{20} with T_{17}, while the signal at *m/z* 7193 corresponds to an adduct of dT_{20} with T_{18}.

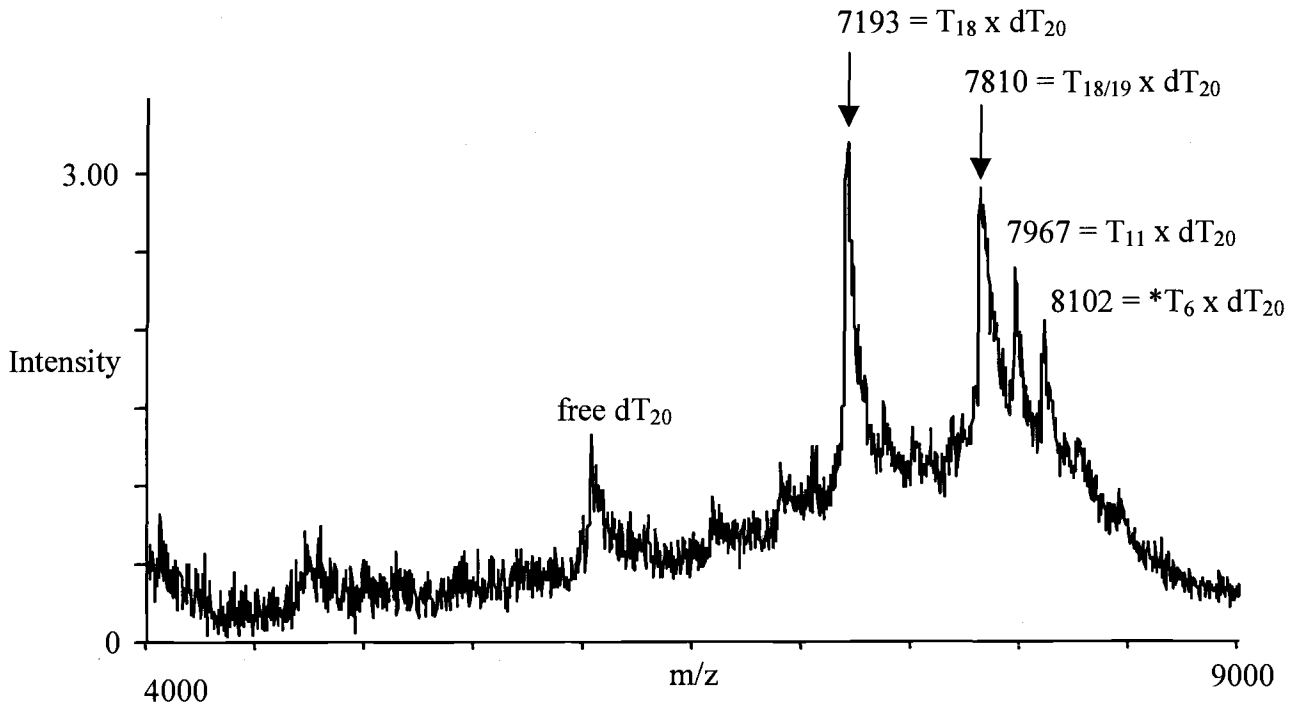


Figure 3.1A. MALDI spectrum of Ung peptides crosslinked to dT_{20}.

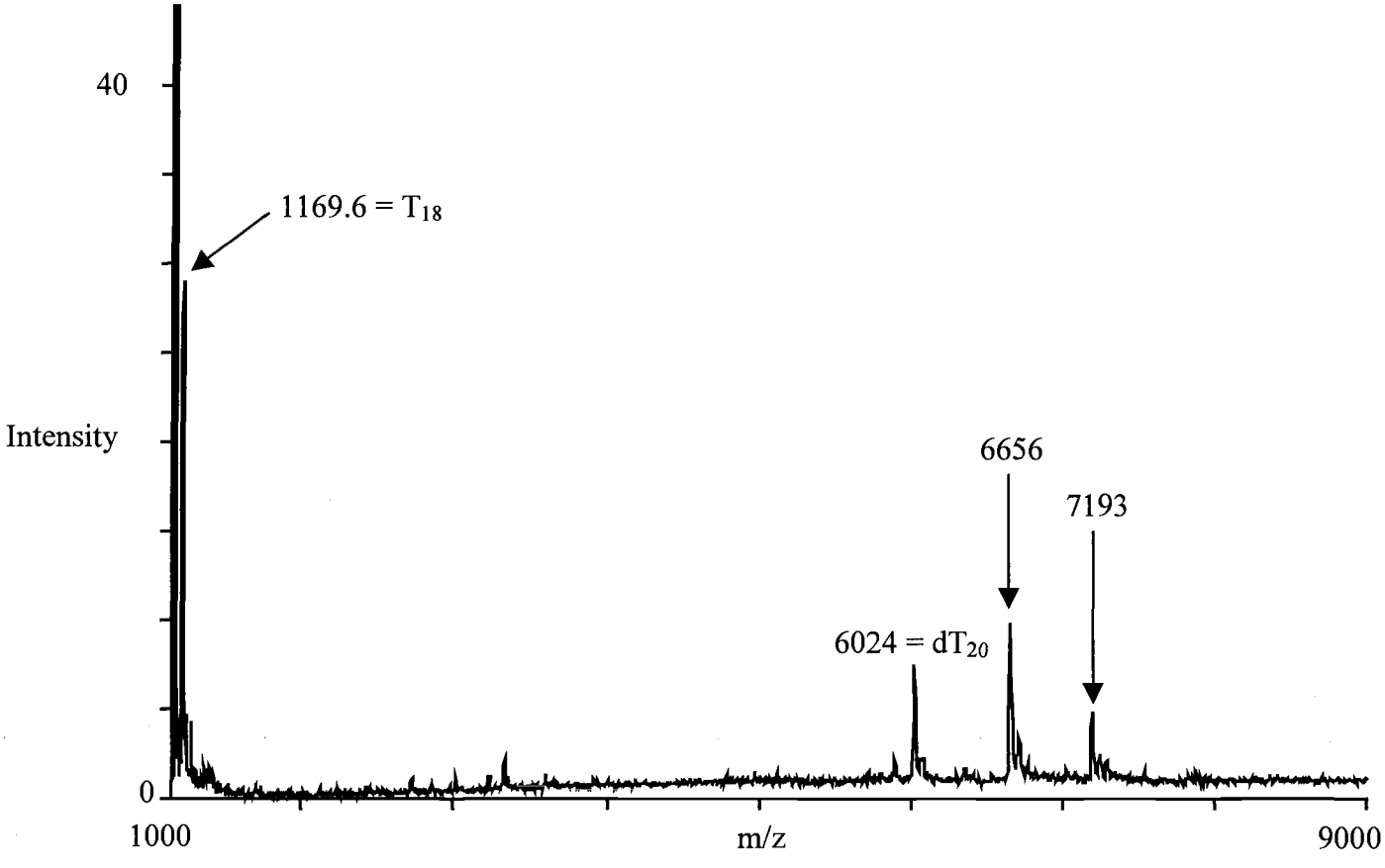


Figure 3.1B. MALDI spectrum showing the adducts formed after mixing a fraction containing T_{16} and T_{18} with dT_{20}.

3.3.2. Off-line Nano-LC/UV Followed by MALDI Analysis

Characterization of individual nucleopeptides was initially attempted by collecting corresponding fractions of eluate from nano-LV/UV runs followed by MALDI and nanospray analysis. The utility of peptide fractionation by nano-LC before MALDI or nano-ESI has been previously demonstrated [53]. A nucleopeptide sample was digested with nuclease P1, and the sample was injected into the nano-LC system. The nano-LC/UV chromatogram obtained is exhibited in Figure 3.2. The sensitivity is not as great as would be obtained with a conventional UV-detector because a very short path length (30 μm) is used for the nano-LC system. The baseline absorbance in this chromatogram gradually decreases with increasing time as acetonitrile from solvent B (40 % acetonitrile with 0.1 % TFA) replaces water from solvent A (0.1 % TFA in water). Two fractions with a volume of 3 μL each were collected manually. Figure 3.3A shows the MALDI spectrum obtained for Fraction 1, while Figure 3.3B shows the MALDI spectrum obtained for Fraction 2. Fraction 1 contained the T_{18} peptide crosslinked to two (MW=1796.4) or three (MW=2100.8) deoxythymidine bases and, Fraction 2 revealed a $^{*}T_{6\ (-5)}$ peptide fragment ($^{*}T_{6(-5)}$ is T_6 without the last five C-terminal amino acids) crosslinked to two thymidine bases (MW=2486.4).

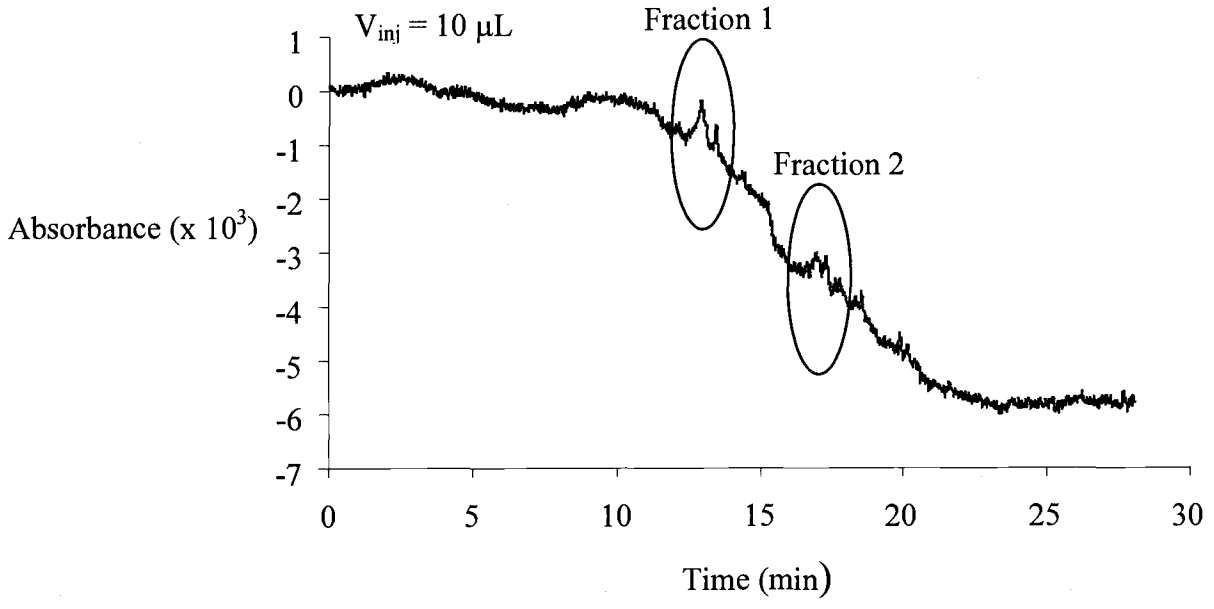


Figure 3.2. Nano-LC/UV chromatogram produced by loading the column with 10 µL of a nuclease P1 digest. The sample was eluted at 0.5 µL/min (pressure at inline filter = 100 bar) with an exponentially produced gradient of 0% to ~ 100% solvent B (0.1 % TFA in 40% acetonitrile). UV detection was at 200 nm. Two fractions were collected manually with a fraction volume of 2 µL (2 min collection period). One µL of Solvent A (0.1% TFA in water) was added to each vial before starting the manual collection.

Figure 3.3. MALDI spectra of dT_{2-3} isolated nucleopeptides. (A) MALDI spectrum of T_{18} x dT_2 (*m/z* 1796.4) and T_{18} x dT_3 (*m/z* 2100.8) nucleopeptides isolated in Fraction 1. The inset picture shows the isotopic distribution of T_{18} x dT_2. (B) MALDI spectrum of $^{*}T_{6\ (-5)}$ x dT_2 nucleopeptide isolated in Fraction 2. An aliquot (1 μL) of the purified nucleopeptide sample was mixed with THAP matrix in a 1:3 ratio and analyzed by MALDI MS.

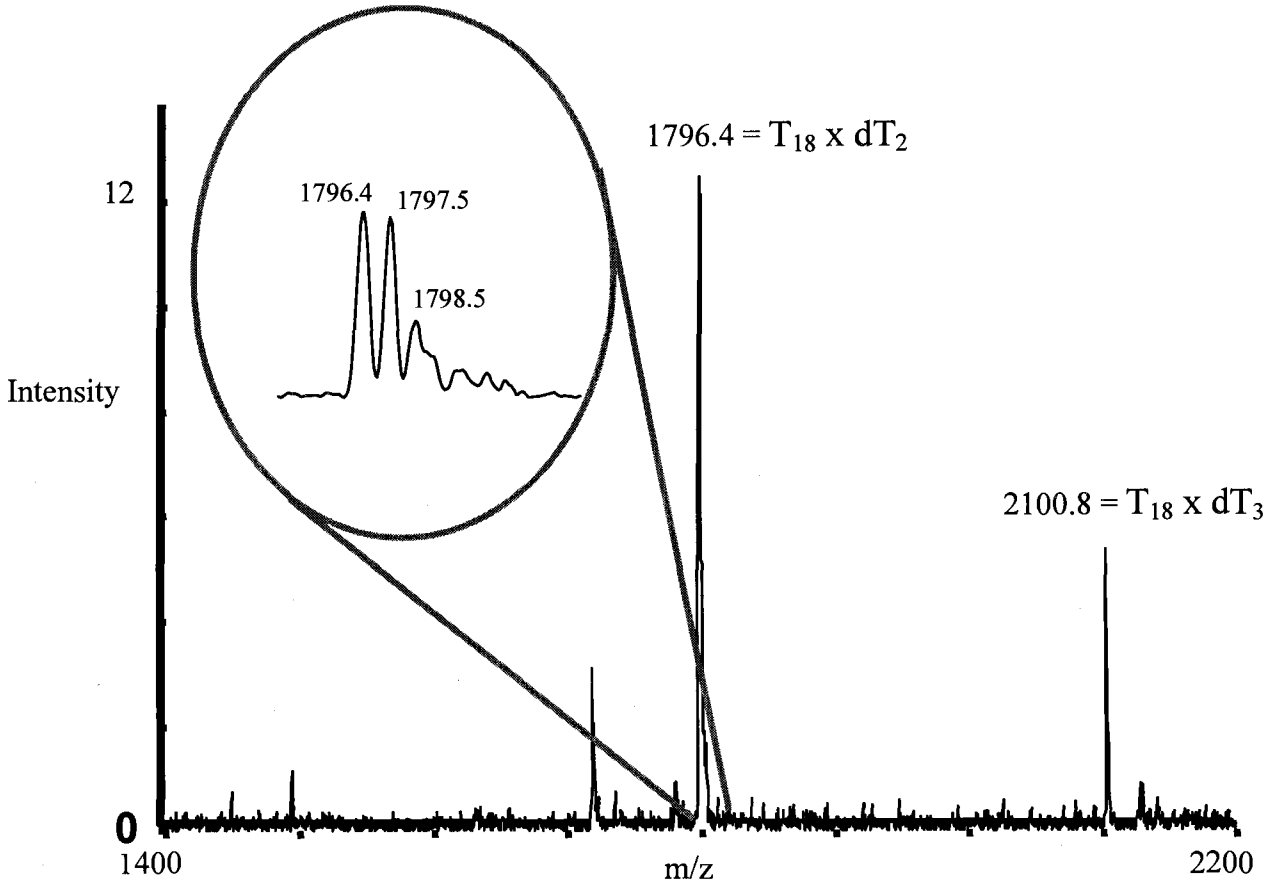


Figure 3.3A. MALDI spectrum of Fraction 1.

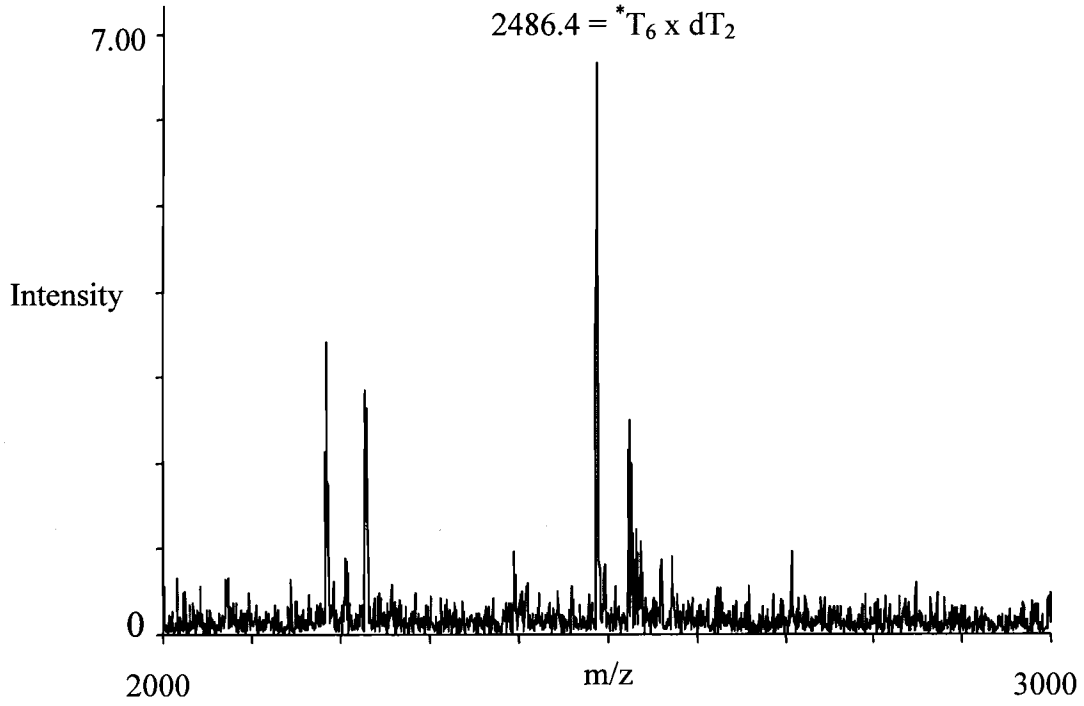


Figure 3.3B. MALDI spectrum of Fraction 2.

These results demonstrate clearly that some of the Ung amino acids involved in binding DNA are located within the T_6 and T_{18} tryptic peptides. However, nanospray MS/MS analysis performed on the Fraction 1 failed to provide the sequence information needed to assign the locations of amino acids involved in DNA interaction.

3.3.3. On-line Nano-LC/MS

In a second attempt to characterize individual nucleopeptides, an on-line nano-LC/ESI-MS/MS approach was used in order increase sensitivity. To preserve the chromatographic resolution provided by the nano-LC column and to avoid additional dead volumes introduced by the coupling between the nanocolumn and the spraying needle, a 40 cm long column was packed into a fused-silica capillary with an integral frit. The spraying voltage (2.5 kV) was applied at the injector since the tapered fused silica needle is not an electrically conducting material. Figure 3.4 shows the nano-LC/ESI-MS extracted ion chromatogram obtained for a nuclease P1 digested nucleopeptide sample. The MS/MS spectra were recorded for the doubly charged molecular ions of T_{18}, T_{11} $^{*}T_{6(-3)}$, $^{*}T_{6(-5)}$ and $^{*}T_{6(-7)}$ nucleopeptides. Three different nucleopeptide species derived from Ung peptide T_6 with 3, 5, or 7 aminoacids lost from the C-terminal were observed. The fragmentation patterns

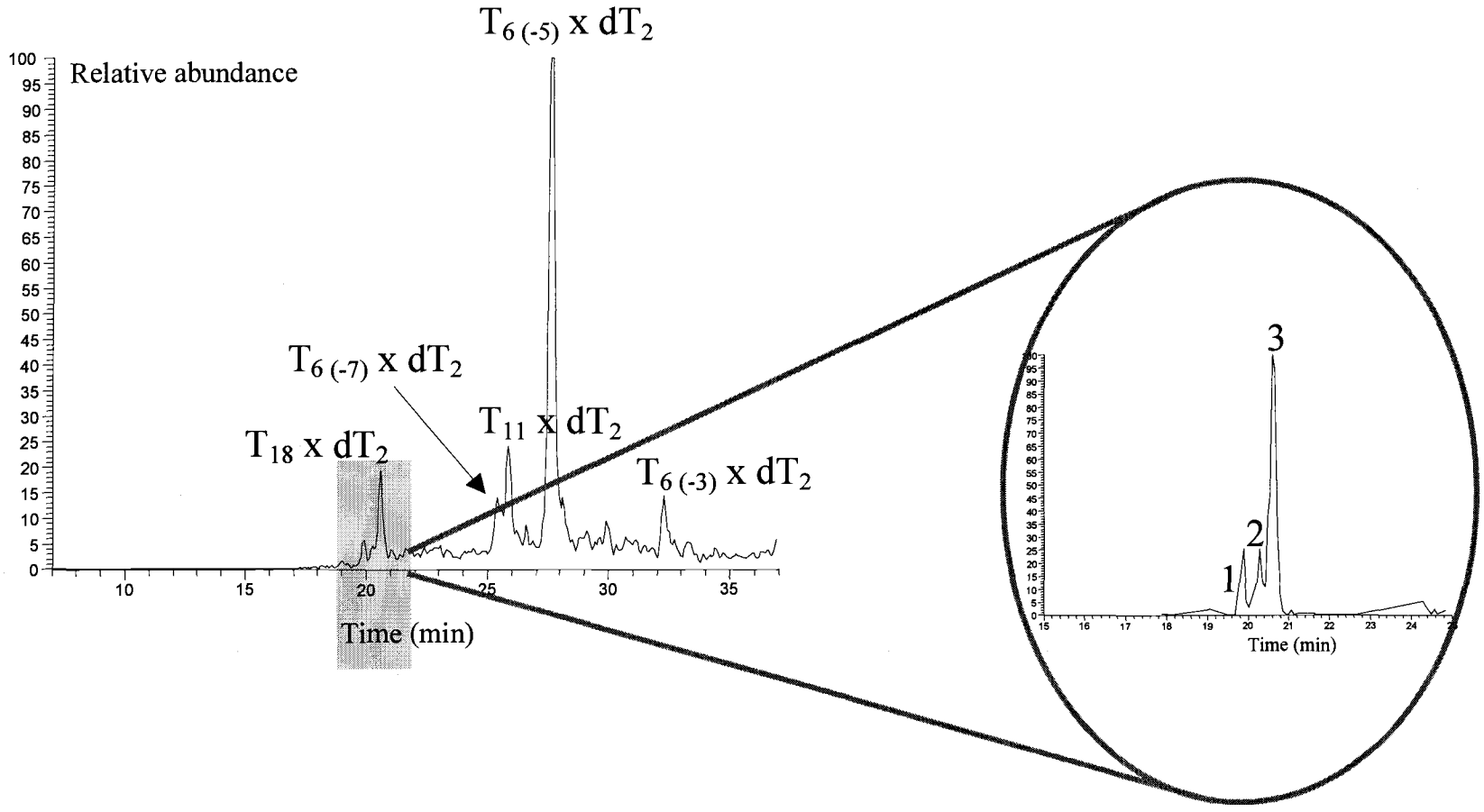


Figure 3.4. Nano-LC/ESI-MS extracted ion-chromatogram of P1 digested nucleopeptides. For all peaks the doubly charged molecular ions were selected. The inset picture shows the LC/MS chromatogram produced for the T_{18} isomers. The nucleopeptide sample (5 μL) was injected into the column (75 μm i.d. x 40 cm long; 5 μm C_{18}) and eluted at 0.5 μL/min (pressure at inlet filter = 140 bar) with an exponentially produced gradient of 0% to ~ 100% solvent B (0.1% AA plus 0.01% TFA in 40% acetonitrile). Ion detection was from m/z 300 to 2000.

deduced from the MS/MS spectra of all three T_6 derived peptides identified two aminoacids (Y66 and H67) as being crosslinked, confirming the previous findings for this peptide [11,12]. The MS/MS spectrum obtained for the T_{11} x dT_2 nucleopeptide failed to reveal a definite location for any crosslinked amino acid. Unexpectedly clustered around an elution time previously found to correspond to the T_{18} x dT_2 nucleopeptide, three different peaks were resolved in the LC/MS chromatogram. These three peaks are displayed in the inset picture in Figure 3.4. It was determined from the MS/MS spectra of these three compounds that they have the same peptide backbone (T_{18} peptide) but with dinucleotides attached to different aminoacids. The tandem mass spectra of these T_{18} x dT_2 isomers are presented in Figure 3.5A-C. The fragmentation patterns observed in the case of each isomer are summarized in the fragmentation diagrams displayed at the top of each spectrum. There are a few common ions present in all three spectra: two of them produced by the fragmentation of the dinucleotide "tag" (a_1 at *m/z* 737.4 and c_1 at *m/z* 777.4); two others produced by fragmentation of the peptide backbone ($^{*}y_9^{+2}$ at *m/z* 814.5 and $[y_9+\text{Thymine}]^{+2}$ at *m/z* 564.4); and one other doubly charged species (*m/z* 648.4) corresponding to the T_{18} peptide crosslinked to thymine. Crosslinked amino acids can be recognized from peptide backbone fragmentation since some of the *b* and *y* ions will retain the dinucleotide "tag" (626.4). These ions are labeled *b and *y in the spectra shown in Figure 3.5A-C. Based on the MS/MS fragmentation pattern observed for the first T_{18} x dT_2 isomer eluted off the column (Figure 3.5A) it was concluded that either P188 or S189 is crosslinked.

Figure 3.5. Tandem mass spectra of T_{18} x dT_2 nucleopeptide isomers apearing in the insert shown in Figure 3.4. (A) MS/MS spectrum of the nucleopeptide isomer corresponding to Peak 1 (B) MS/MS spectrum of the nucleopeptide isomer corresponding to Peak 2; and (C) MS/MS spectrum of the nucleopeptide isomer corresponding to Peak 3. The precursor selected for fragmentation (30% collision energy) was the doubly charged molecular ion at 898.5 *m/z*. The fragmentation patterns observed in the case of each isomer are summarized in the fragmentation diagrams displayed at the top of each spectrum.

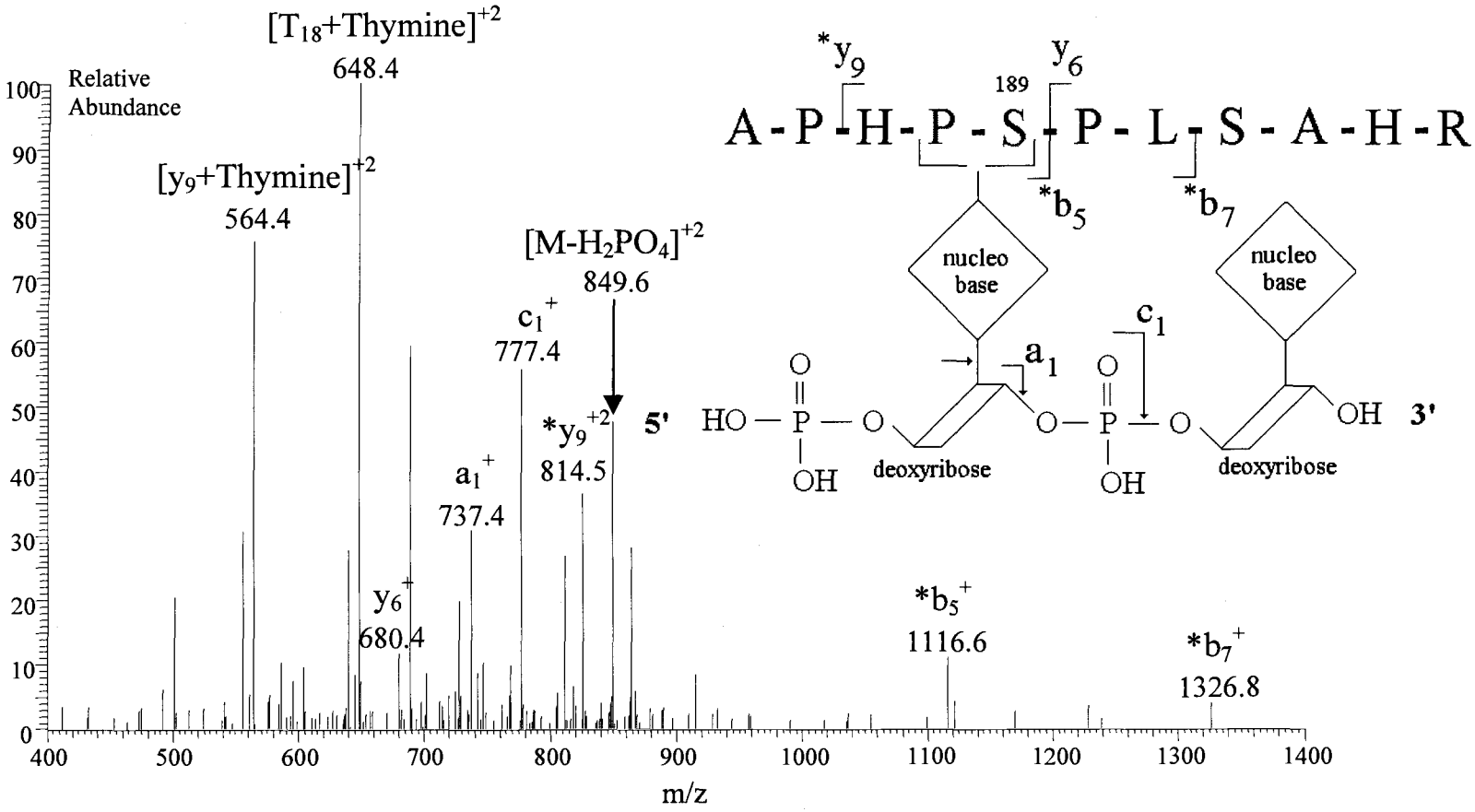


Figure 3.5A. MS/MS spectrum of the T_{18} x dT_2 nucleopeptide isomer corresponding to Peak 1 from Figure 3.4.

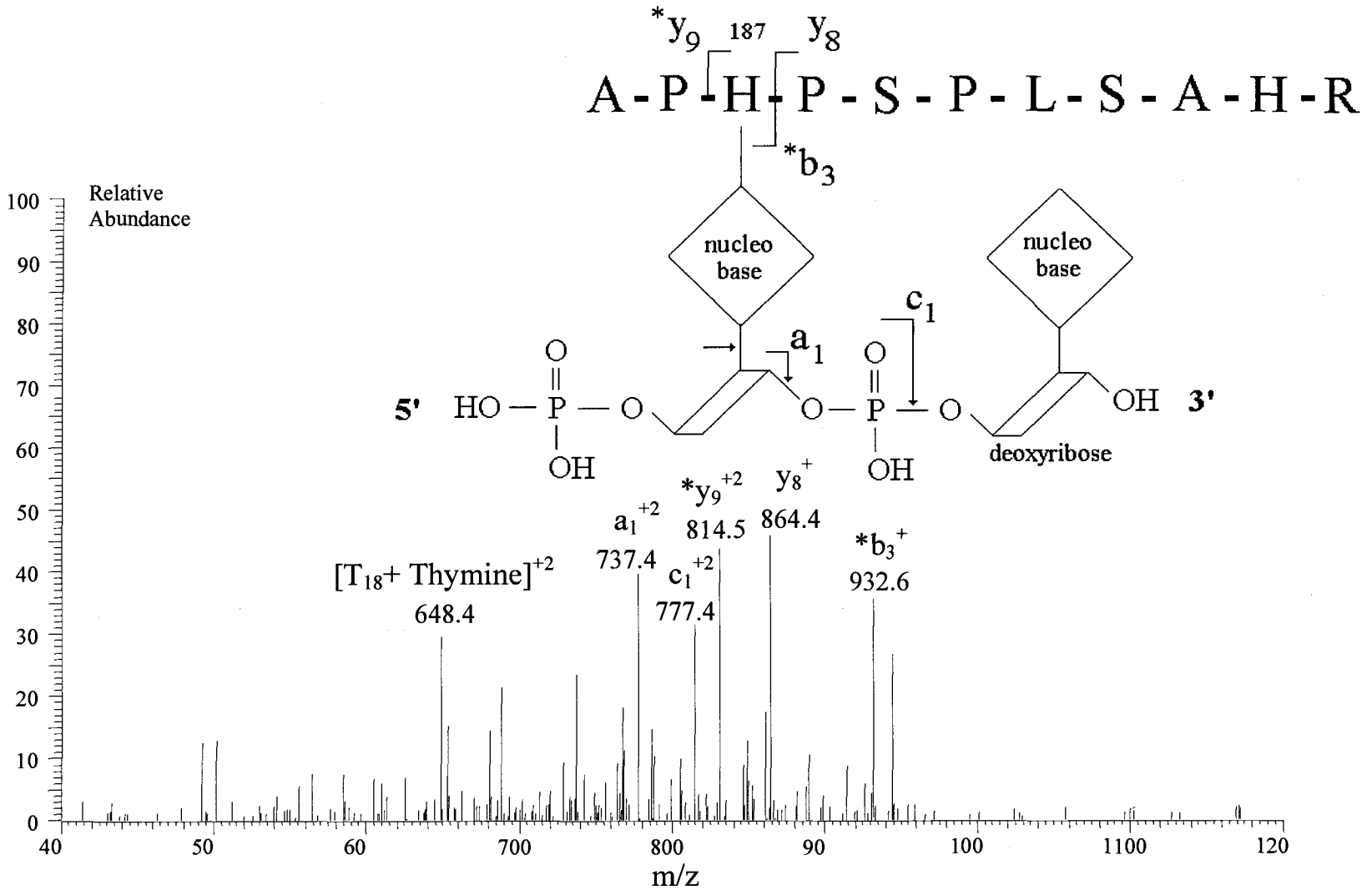


Figure 3.5B. MS/MS spectrum of the T_{18} x dT_2 nucleopeptide isomer corresponding to Peak 2 from Figure 3.4.

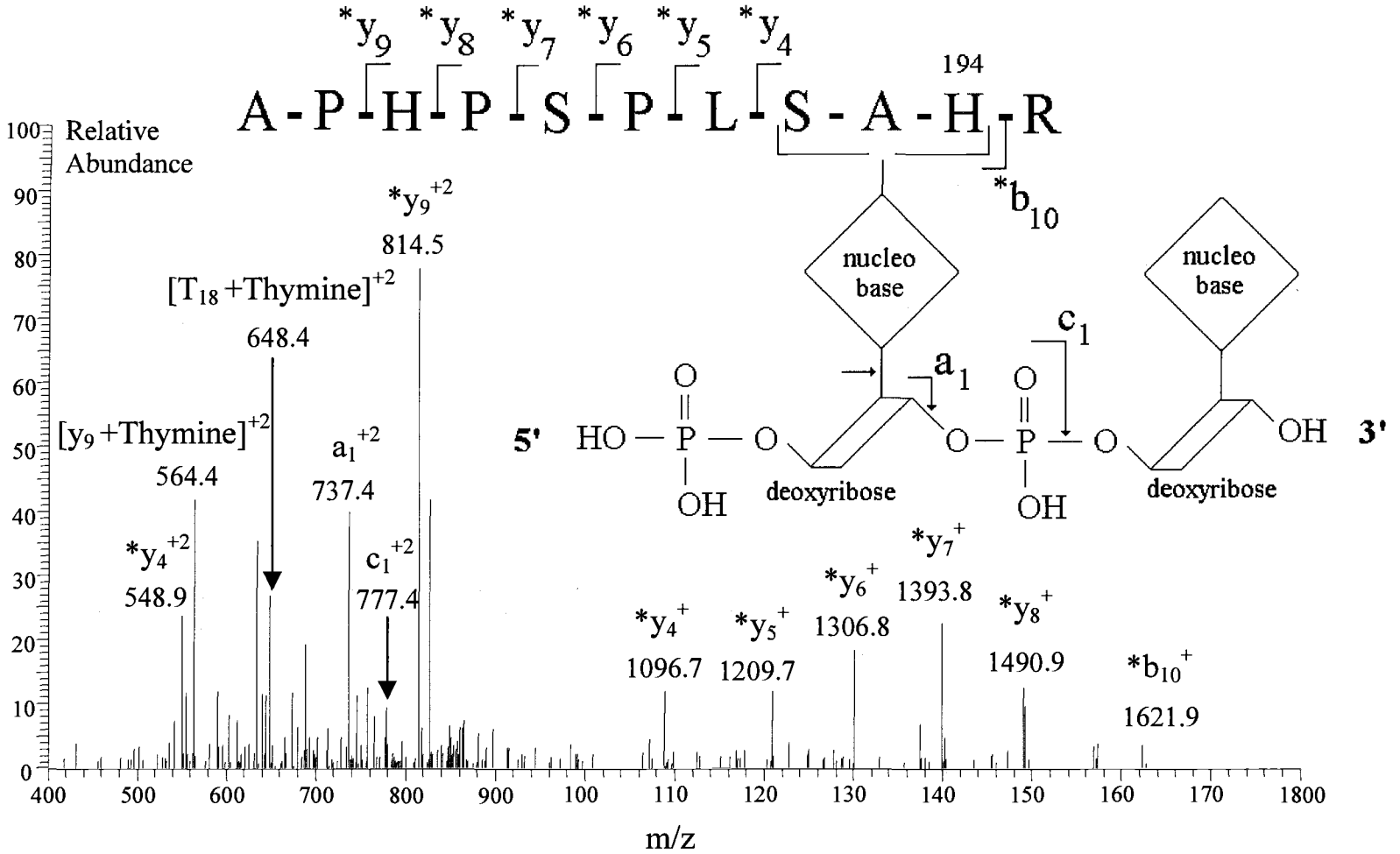


Figure 3.5C. MS/MS spectrum of the T_{18} x dT_2 nucleopeptide isomer corresponding to Peak 3 from Figure 3.4.

In the case of the second nucleopeptide isomer (Peak 2), the crosslinked amino acid was unambiguously identified as H187 based on the presence of three intense ions: $^{*}y_9^{+2}$, $^{*}b_3^{+}$ and y_8^{+} observed in its spectrum (Figure 3.5B). The fragmentation pattern observed for the last T_{18} x dT_2 isomer (Peak 3) indicates that at least one aminoacid from the sequence 192-SAH-194 is crosslinked (see Figure 3.5C).

3.4. DISCUSSION

The ultimate outcome of using mass spectrometry to characterize peptide-nucleic acid complexes is the identification of specific amino acids that are crosslinked to the oligonucleotide. In pursuing this objective, tandem mass spectrometry was used to investigate the fragmentation patterns of dinucleopeptides isolated from the {Ung x dT_{20}} nucleoprotein complex. Dinucleopeptides were produced by nucleolytic digestion of a nucleopeptide sample isolated according to a three step purification protocol developed in the authors' laboratory. As previously reported [7,11,12], steric interference between the peptide portion of the nucleopeptide and the nuclease prevents nucleolysis from proceeding to the mononucleotide level. Two nucleopetides were identified by off line nano-LC MALDI-TOF approach shown in Figure 3.2 and seven dinucleopetides were identified by online nano-

LC/ESI-MS/MS. The nanospray MS/MS spectrum obtained for Fraction 1 (data not shown) contained contradictory information regarding the fragmentation pattern of T_{18} x dT_2. Subsequently, on-line nano-LC/ESI-MS (Figure 3.4) clearly revealed that T_{18} x dT_2 comprised three, barely resolvable isomers. This finding made it immediately clear that it would have been impossible to distinguish the individual fragmentation pattern of any of these isomers from the other two in the spectrum of Fraction 1 because all three were present in this off-line fraction. The chromatographic separation of the T_{18} x dT_2 isomers thus was decisive in resolving the nature of the photochemical crosslinking to dT_{20} that occurs in the DNA-binding domain of Ung. The MS/MS spectra of the three isomers have some common features that are easily attributable to the fragmentation of the dT_2 – tag that each possesses. As has been previously observed [10-12], collision induced fragmentation along the backbone of the peptide portion of a nucleopeptide is usually weak. Nonetheless, the information obtained from these spectra of the three T_{18} x dT_2 isomers was sufficient to unambiguously show that at least three amino acids in T_{18} are crosslinked to a dT_2. This new finding suggests that the mass spectrometric protocol used to obtain it is capable of revealing biologically relevant information about DNA-binding domains in proteins.

Definitive identification of H187 as one of the crosslinked amino acids (Figure 3.5B) is in good agreement with previous information obtained using site-directed mutagenesis [84], crystallography [77,79] and NMR spectroscopy [79,85]. Specifically, Ung exhibited a 55,000-fold lower specific activity and a shift in pH

optimum from 8.0 to 7.0 when H187 was mutated into aspartic acid [84]. Surprisingly, H187 is not involved in a stacking interaction with the uracil base as are most aromatic amino acids. NMR spectroscopy [79,85] showed the formation of a short hydrogen bond between H187 and an oxygen atom from uracil. It seems plausible that a similar interaction could occur between H187 and one of the thymidines from dT_{20}.

Furthermore, S189 is known to be involved in DNA interactions with the phosphate backbone [82]. Although the fragmentation data obtained from the spectrum shown in Figure 3.5A indicated that either P188 or S189 will sometimes crosslink to dT_{20} under UV irradiation, photochemical crossslinking studies conducted by others indicate that a serine is more likely to be involved in crosslinking than a proline [17].

It can only be deduced from the fragmentation pattern exhibited in Figure 3.5C that at least one of the three residues in the sequence 192-SAH-194 forms covalent bonds to dT_{20} during the photochemical crosslinking experiment. The known nature of chemical interactions between proteins and nucleic acids and the fact that H194 is the only aromatic aminoacid in this group of residues, suggests that H194, rather than S192 or A193 is the residue in the third T_{18} x dT_2 isomer (Peak 3, Figure 3.4) that crosslinks to dT_{20}.

UV photocrosslinking can only reveal amino acid interactions with the nucleobases that make up a DNA substrate; electrostatic interactions with the phosphate backbone cannot be observed via this technique. Thus, base stackings between

aromatic amino acids and nucleobases are the most likely form of protein-nucleic acid interaction to be observed in a UV crosslinking experiment. The mass spectrometric data obtained from T_6 in this study as well as an earlier study [11,12] are consistent with these mechanisms. Specifically, base stacking interactions were observed in the T_6 peptide for Y66 and H67, but the electrostatic interaction between D64 and the phosphate backbone was not observed despite the fact that it was previously reported in several X-ray diffraction papers [75,77]. Given that the third T_{18} x dT_2 isomer is the most intense peak in the inset of the LC/MS chromatogram from Figure 3.4 and that the aromatic H194 is the only residue on the C-terminal of T18 that can participate in a base stacking interaction with dT20, it is very probable that H194 is the dT_2 - tagged residue in this nucleopeptide isomer.

All of the information regarding the DNA-binding domain of Ung obtained from this study by UV crosslinking and mass spectrometry is summarized in the 3D-structure showed in Figure 3.6. This 3D-model corresponds to the crystal structure of *E.coli* uracil-DNA glycosylase reported by Xiao et al [77]. Three peptide regions involved in crosslinking (T_6, T_{11} and T_{18}) are colored in blue, and five crosslinked amino acids (Y66, H67, H187, S189 and H194) are colored in red. All these amino acids belong to the DNA-binding domain of Ung, and their interactions with the DNA substrate are consistent with the formation of a stable {Ung x dT_{20}} nucleoprotein complex.

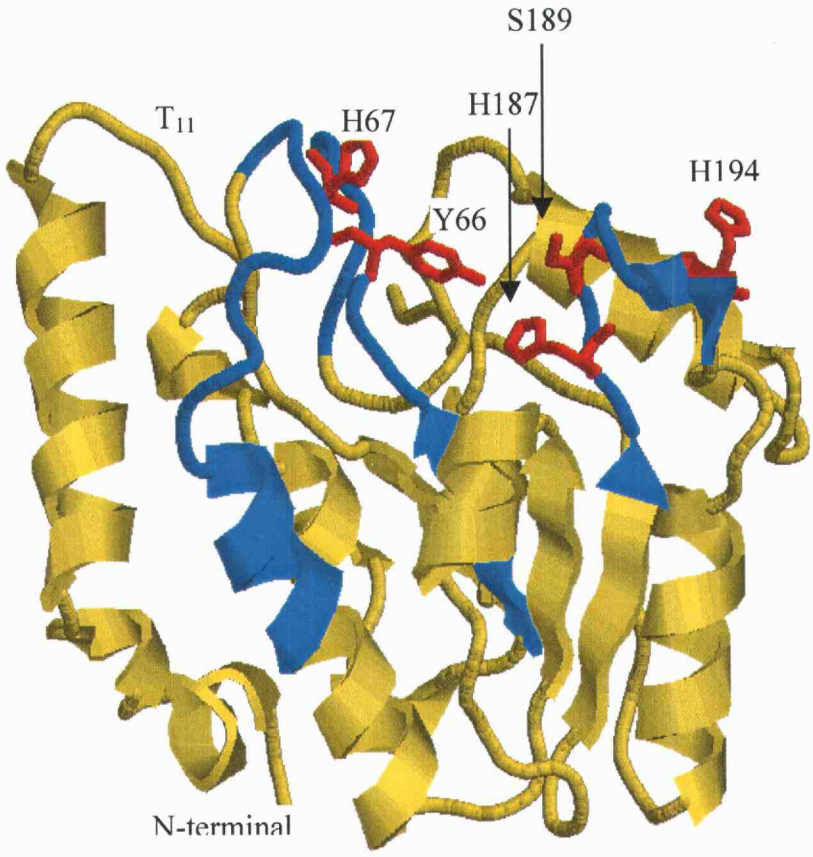


Figure 3.6. 3D structure of Ung showing the DNA-binding region of the protein obtained from crystallography data [77]. Each of the three distinct peptide regions colored in blue have at least one crosslinked amino acid. Several amino acids, along with a serine residue, were identified by UV-photocrosslinking and mass spectrometry as being crosslinked to oligonucleotide dT_{20}. All crosslinked amino acids are colored in red. The drawing was generated using the RasMol program [86].

In addition to these five residues, the isolation of a T_{18} x dT_2 nucleopeptide in this study and an earlier study [11,12] suggests there should be at least one other amino acid (from T_{11}) interacting with the DNA substrate. More experiments will be required to identify the crosslinked amino acids in T_{11}.

The results of this study have demonstrated that chromatographic resolution of nucleopeptide isomers coupled with tandem mass spectrometry enable in depth structural analysis of a nucleoprotein complex.

4. IDENTIFICATION OF AMINO ACID RESIDUES IN THE SINGLE-STRANDED DNA BINDING DOMAIN OF HUMAN REPLICATION PROTEIN A

Reproduced with permission from Doneanu, C.E., Gafken, P.R., Bennett, S.E., and Barofsky, D.F, *J. Biol. Chem.*, submitted.

4.1. INTRODUCTION

Protein-nucleic acid interactions are involved in many cellular processes including transcription, translation, replication, recombination and genomic damage repair. Mass spectrometry in combination with photochemical/chemical crosslinking, along with X-ray diffraction and NMR spectroscopy, is considered to be a method that provides information directly about the structural relationship between proteins and nucleic acids. The strategy, which combines low intensity UV crosslinking and mass spectrometry, was initially demonstrated in 1994 [3]. Subsequently, laser-induced crosslinking [4,5] and chemical crosslinking [6] have been used in combination with mass spectrometry to investigate protein-DNA interactions. The coupling of photocrosslinking to mass spectrometry for protein-DNA analysis has seen more frequent usage of late [7-14,107].

Photochemically crosslinking nucleic acid binding proteins to their substrates is an established procedure for characterizing protein-nucleic acid interfaces [17-19].

This technique exploits photons to induce the formation of "zero-length" covalent bonds between certain amino acids that are in close contact with nucleic acid bases [17]. It is assumed that only amino acids located in the protein's binding site are involved in the formation of these bonds. For mass spectrometric identification of amino acids that take part in the crosslinking, a general methodology has evolved. This methodology uses enzymatic digestion of the UV-irradiated nucleoprotein complex to produce free peptides and peptides crosslinked to the nucleic acid substrates (nucleopeptides). After separation from the non-crosslinked peptides, the nucleopeptides are analyzed by MALDI-TOF to determine their molecular weights. By comparing the mass of a nucleopeptide minus the mass of the attached oligonucleotide to a peptide map of the native protein, one can identify regions comprising peptides that are involved in crosslinking. This straightforward strategy can be used only if the chemistry of the photochemical reaction is known. Finally, the nucleopeptides are sequenced by tandem mass spectrometry (MS/MS) in order to identify the particular amino acids attached to nucleic acid bases. In a few instances, a combination of exhaustive enzymatic digestion and MALDI-TOF MS can provide enough information to locate the crosslinked amino acids within a nucleopeptide [8].

In this paper, we report the use of photochemical crosslinking and mass spectrometry to investigate the DNA-binding domain of human Replication Protein A (hRPA). hRPA is a heterotrimeric single-stranded DNA-binding protein that is required for multiple processes in cellular DNA metabolism, including DNA

replication, repair, and recombination [87]. This protein was originally purified from human cell extracts and identified as a protein required for simian virus 40 (SV40) DNA replication [88,89]. Analysis of hRPA indicated that it is a stable complex of three subunits with molecular masses of 70, 32 and 14 kDa (hRPA70, hRPA32, and hRPA14, respectively) [55,90]. The protein binds single-stranded DNA (ssDNA) with specific polarity [91] and high affinity. It forms a stable complex with DNA that has an occluded binding site size of 30 nucleotides [92,93], although two other binding modes involving 8-10 nucleotides and 13-14 nucleotides were reported [94,95]. Deletion and mutational analyses [87,96] have identified four domains in hRPA70: an N-terminal domain (residues 1-168) that participates in hRPA-protein interactions, a central DNA binding domain (residues 169-441), a putative zinc-finger motif (residues 481-503), and a C-terminal domain (residues 503-616). Recent data [97,98] indicate that the C-terminal domain of hRPA70 is also a DNA binding domain (domain C).

The structure of the central DNA-binding domain has been previously investigated by X-ray diffraction [56,99] and site-directed mutagenesis [98,100]. The crystal structure revealed a central DNA-binding domain comprising two, three-dimensionally identical copies of a single-stranded DNA binding subdomain. These subdomains (identified as A and B) resemble an OB-fold motif that has been previously found in various oligonucleotide/oligosaccharide-binding proteins [101]. Subdomains A (residues 181-290) and B (residues 300-426) are oriented in tandem, and a DNA binding channel extends from one subdomain to the other. Each

subdomain is in direct contact with three nucleotides while two nucleotides fill the space between the subdomains. Two primary types of interaction were observed to occur between the protein and DNA: stacking interactions between several aromatic aminoacids (F238, F269, W361 and F386) and individual nucleic acid bases, and hydrogen bonding between amino acid side chains and either the phosphate backbone or individual nucleic acid bases. The X-ray data does not make clear what the individual contributions of these interactions to the overall DNA-binding affinity of hRPA. Furthermore, the importance of aromatic amino acids interacting with single-stranded DNA has been questioned since site directed mutants of hRPA70, in which F238, F269, W361 and F386 were changed to alanine, resulted in only slight changes in the protein's binding constant. A double mutant, in which F238 and W361 were changed to alanine, resulted in a decrease of the hRPA binding constant by 3 orders of magnitude [100]; however, the structural identity of the double mutant was determined by proteolytic sensitivity experiments to be disrupted. Structural perturbation makes it difficult to draw conclusions about the double mutant's phenotype, but the results obtained with the single aromatic residue mutants suggest that no single aromatic residue is critical for the formation of the hRPA-DNA complex [100].

Very recent studies [102-104] suggest a dynamic mechanism for hRPA binding to ss-DNA, involving all three domains from hRPA70 (A, B and C) and a fourth domain from hRPA32 (residues 43-171). This mechanism explains different binding modes observed for hRPA.

The present study focuses on the identification of some of the amino acid residues located in the single-stranded DNA-binding domain of hRPA that are critical for the formation of a stable protein-DNA complex.

4.2. EXPERIMENTAL

4.2.1. Chemicals

Acetic acid (AA), trifluoroacetic acid (TFA), ampicillin, bromophenol blue, calcium chloride, copper (II) chloride, chloramphenicol, ethylenediaminetetraacetic acid (EDTA), ferric chloride hexahydrate, formaldehyde, formic acid, iodoacetamide, inositol, leupeptin, HEPES, Nonidet P-40, phenylmethylsulfonyl fluoride (PMSF), potassium ferricyanide, sodium chloride, sodium thiocyanate and Trizma (Tris-base), and 2,4,6-trixydroxyacetophenone (THAP) were purchased from Sigma Chemical Co. (St Louis, MO). Acrylamide (>99% pure), ammonium persulfate, Coomassie Brilliant Blue G-250, 2-mercaptoethanol, bis N,N`-methyl-bis-acrylamide, and N,N,N`N`-tetramethyl-ethylenediamine (TEMED) were obtained from Bio-Rad (Hercules, CA). Ammonium sulfate, dithiothreitol (DTT), glycine, isopropyl-β-δ-thiogalactopyranoside (IPTG), and sodium dodecyl sulfate (SDS) were acquired from Life Technologies Inc. (Grand Island, NY). HPLC grade acetonitrile, ammonium acetate, ammonium bicarbonate, ammonium hydroxide

(50%), glucose, hydrochloric acid, isopropanol, methanol, potassium chloride, sodium carbonate, sodium hydroxide, sodium nitrate and sodium thiosulfate were bought from Fisher Scientific (Pittsburgh, PA). Glycerol and potassium phosphate (dibasic and monobasic salts) were purchased from J.T. Baker (Phillipsburg, NJ). Agar, tryptone and yeast extract were supplied by Becton Dickinson (Franklin Lakes, NJ). Water was generated with a Milli-Q water purification system (Millipore Corp., Bedford, MA). Oligodeoxythymidine dT_{30}, was synthesized by the Biopolymer Core Facility, at the University of Maryland at Baltimore and further purified on a 15% native polyacrylamide gel.

4.2.2. Protein Expression and Purification

Recombinant hRPA protein was expressed in *E. coli* cells by using the T7 expression system [105]. The plasmid p11d-tRPA [106], along with the plasmid pRIL (Stratagene, La Jolla, CA), were electroporated into *E. coli* strain BL21(DE3) and plated on LB plates (1% tryptone, 0.5% yeast extract, 1% NaCl, 1.5% agar) containing 1% glucose, 100 μg/ml ampicillin, and 34 μg/ml chloramphenicol. A single recombinant colony was selected to inoculate one liter of TB medium (1.25 tryptone, 2.4% yeast extract, 0.4% (w/v) glycerol, 0.017 M KH_2PO_4, and 0.07 M K_2HPO_4) containing 1% glucose, 100 μg/ml ampicillin, and 34 μg/ml chloramphenicol. The culture was incubated overnight at 30 °C without shaking

and then placed on an orbital shaker and shaken at 37 °C. When the absorbance at 600 nm was between 0.6 and 0.8, the culture was induced by adding IPTG to 0.3 mM. After 2 h of induction, the cells were harvested by centrifugation at 4,000 × g and frozen at -80 °C until used.

All purification steps were carried out at 4 °C. Frozen cells (from 1 L of culture) were resuspended in 50 mL of HI buffer (30 mM HEPES (pH 7.8), 1 mM DTT, 0.25 mM EDTA, 0.25% (w/v) inositol and 0.01% (v/v) Nonidet P-40) containing 50 mM KCl, 1 mM PMSF, and 0.1 μg/ml leupeptin. The cells were disrupted by sonication, and the resulting lysate was cleared by centrifugation at 30,000 × g for 30 min. Soluble protein was applied to a 20-mL Affi-Gel Blue Gel column (12 cm × 1.77 cm^2, BioRad) equilibrated in HI buffer containing 50 mM KCl. The column was washed sequentially with 3 column volumes each of HI buffer containing respectively 0.05 M KCl, 0.8 M KCl, 0.5 M NaSCN, 1.0 M NaSCN, and 1.5 M NaSCN. The peak of elution (represented by those fractions that were enriched for the 70 kDa, 32 kDa, and 14 kDa subunits of hRPA) was determined by denaturing polyacrylamide gel electrophoresis and visualizing the protein bands with Coomassie stain. Fractions containing the elution peak (present in the column washings with HI buffer containing 1.5 M NaSCN) were pooled into 12,000-14,000 molecular weight cutoff dialysis tubing (SpectraPor) and dialyzed against HI buffer containing 500 mM NaCl. It is important to note that dialysis into HI buffer containing 250 mM NaCl resulted in extensive protein precipitation. After 6 h, the dialysis solution was replaced with fresh buffer and dialysis was continued

for an additional 14 h. The dialyzed sample was loaded onto a 23-mL single-stranded DNA agarose column (4.5 cm × 4.91 cm^2) equilibrated in HI buffer containing 500 mM NaCl. The column was washed with 3 column volumes of HI equilibration buffer and then eluted with a 100-mL linear gradient of equilibration buffer ranging from 500 mM NaCl to 3 M NaCl. The peak of elution (present from the 0.7 M to 1.7 M NaCl region of the gradient), also monitored by denaturing polyacrylamide gel electrophoresis, was pooled and concentrated by using a Centriprep-30 cartridge (Millipore) and the buffer was exchanged to HI buffer containing 200 mM KCl. Typically, about 1.0 mg of protein was obtained per liter of cultured cells, and this material was estimated by densitometry experiments to be greater than 90% pure hRPA (data not shown). Electrophoretic mobility shift analysis of the purified protein, using [^{32}P]dT_{30} as the DNA substrate, showed that virtually all of the purified protein was capable of binding this substrate (data not shown).

4.2.3. Photochemical Crosslinking Reaction

A 1:1 mixture of purified hRPA (800 pmol) and dT_{30} (800 pmole) in HI buffer containing 200 mM KCl was placed in a 4 mL, capped quartz cuvette (NSG Precision Cells, Farmingdale, NY) and equilibrated at room temperature for 30 min. The cuvette was placed lengthwise on a bed of ice and irradiated with UV

light (λ_{max}=254 nm) for 1 min in a Stratalinker 1800 crosslinker (Stratagene, La Jolla, CA).

4.2.4. Isolation of Tryptic Peptides Crosslinked to dT_{30}

Ten aliquots of the irradiated crosslinking mixture were loaded onto a 10% SDS polyacrylamide gel and subjected to gel electrophoresis. After removing the stacking gel, the lower 3-cm portion of the resolving gel corresponding to the region of free dT_{30} was removed, and the remaining portion of gel was silver stained [64]. Bands corresponding to hRPA70 × dT_{30} (XL-1 band) were cut from the gel with a clean, sharp razorblade and placed into 1.5 mL Eppendorff tubes. Gel slices were in-gel digested with trypsin (36 ng/uL) following the procedure of Shevchenko et al. [64]. After 2 h of digestion at 37 °C, 400 μL of digestion buffer was added to the gel slices, and digestion was continued overnight. Upon completion of digestion, the aqueous solution was removed and saved. An organic extraction mixture of FAPH (50% formic acid, 25% acetonitrile, 15% isopropyl alcohol, and 10% water) was added to each tube until the gel pieces were covered, and the tubes were vortexed at room temperature for 4 h. The extraction mixture was removed from each tube and saved. Gel slices were then crushed and dehydrated with 200 μL of acetonitrile for 10 min. The acetonitrile was removed, with care not to remove any gel pieces, and added to the saved FAPH extraction

solutions. The combined organic solutions were concentrated to near dryness by vacuum centrifugation. The saved aqueous digestion solution was adjusted to 10 mM Tris-HCl (pH 7.2), 1 mM EDTA, and 100 mM sodium chloride (TE-100), and the dried organic extraction fractions were resuspended with 250 μL TE-100. Both the adjusted aqueous digestion solution and the resuspended organic extraction solution were applied to a NAC-52 anion exchange cartridge (Life Technologies Inc.) that had been activated with 3 mL TE-2000 (TE buffer containing 2000 mM sodium chloride) and equilibrated with 5 mL TE-100. After sample loading, the cartridge was washed with 5 mL TE-100 and the peptide × dT_{30} fragments (nucleopeptides) were eluted with 1 mL TE-1000 directly into a Centricon-3 (Millipore) desalting cartridge. The Centricon-3 was centrifuged at 6000 × g. The eluate was desalted by washing the Centricon-3 sequentially with 4.5 mL of water, 4.5 mL of 10 mM ammonium acetate, and 3 mL of water. The washed sample was concentrated to 200 μL, after which, it was transferred to a 750-μL Eppendorff tube and further concentrated to about 10 μL by vacuum centrifugation.

4.2.5. Nuclease P1 Digestion

Nuclease P1 (ICN Pharmaceuticals, Costa Mesa, CA) was diluted with 50 mM ammonium acetate pH=6.5 (0.04 activity units/μL). After mixing 1 μL of nuclease

P1 solution with 1 μL of nucleopeptide sample, digestion was allowed to proceed for 4 h at 37 °C.

4.2.6. Carboxypeptidase Y Digestion

Carboxypeptidase Y (Sigma) was diluted with 25 mM ammonium acetate pH=6 (0.5 activity units/μL). Extensive digestion was conducted for 48 h at 37 °C after mixing 1 μL of CPY solution with 1 μL of nucleopeptide sample.

4.2.7. ZipTip Clean-up of Digests

The nucleopeptide samples were cleaned up on C_{18} ZipTips (Millipore) following the manufacturer's directions for peptides. Metal chelate ZipTips (Millipore) were used to clean-up a CPY digested nucleopeptide sample before recording its MALDI spectrum. Briefly, 4 μl of 0.1% AA were added to 1 μl of nucleopeptide sample. The ZipTip was preconditioned twice with 10 μL of 0.1% TFA in 50% acetonitrile, then rinsed 10 times with 10 μL of 200 mM ferric chloride hexahydrate in 10 mM HCl, 10 μL of water (Millipore) and 10 μL of 1% AA in 10% acetonitrile. The sample was loaded on the tip by pipetting the solution 20 times in and out of the

tip. The tip was also rinsed 5 times with 10 μL of 0.1% AA in 50% acetonitrile and 10 μL of water. Nucleopetides were eluted with 1 μL of 2% ammonium hydroxide.

4.2.8. Nano-LC Column Preparation

For on-line nano-LC/ESI MS, a fused-silica column with an integral frit - "PicoFrit" (360 μm o.d. x 75 μm i.d. x 40 cm, 15 μm tip) obtained from New Objective (Cambridge, MA) was packed using the pressurized bomb method described by Kennedy and Jorgensen [60]. Briefly, a 40 cm long "PicoFrit" column was packed with 5 μm, 300 Å pore, Luna C_{18} silica gel particles (Phenomenex, Torrance, CA). The packing material suspended in isopropanol (25 mg/mL) was forced into the capillary by pressurizing the bomb to 1,500 psi (about 6 hours are required to pack a 40 cm long column). After the gas pressure was slowly released from the bomb, the inlet of the column was connected to an HPLC pump and flushed with acetonitrile for 4 hours and Milli-Q water for another 4 hours.

4.2.9. Nano-LC Solvent Delivery System

The nano-LC system employed uses an exponential dilution method to produce gradient separations. A detailed description of the system's operating principle and

setup is provided elsewhere [53]. Briefly, elution gradients are produced by exponential dilution in two mixing chambers. Two conventional HPLC pumps (Kratos Analytical Spectroflow 400, Ramsey, NJ) were used, one to deliver solvent A (0.1 % AA plus 0.01 % TFA) and the other to deliver solvent B (0.1 % AA plus 0.01 % TFA in 50% acetonitrile). From the pump, solvent is passed through an inline filter (Kratos Analytical, Ramsey, NJ) into a flow-splitter. The filter helps avoid clogging of the system downstream. The flow-splitter comprises a PEEK Y-piece (Valco, Houston, TX), a 30 μm i.d x 50 cm fused-silica waste-line that functions as a flow resistor, and a 60 μm i.d x 5 cm PEEK connecting line to the first mixing chamber. The split-ratio of this arrangement is between 1:50 and 1:100 so the flow through the analytical column is in the range 300-800 nL/min. A ZDV PEEK union (Upchurch Scientific, Murrieta, CA) drilled out to an internal volume of 3 μL was used as the first mixing chamber of the exponential gradient elution system, and the inlet port of the injection valve served as the second mixing chamber. A micro-injection valve with a port-volume of about 10 μL (Rheodyne Model 8125, Rohnert Park, CA) and a sample injection loop of 5 μL was used. PEEK tubing with 60 μm i.d. was used to connect the pumps to the filter, splitter, mixing chamber, and injector. The nano-LC column was connected directly to the injector, which was placed in a plastic box, and the spraying voltage (2.5 kV) was applied directly to the injector after loading the sample since the fused silica needle was not electrically conductive.

Samples were loaded from the injection valve onto the nanocolumn by connecting pump A to the system and raising the inlet pressure to 200 bar so the sample was displaced from the injection loop by solvent A onto the column at a flow rate of 0.8 μL/min. After allowing sufficient time to completely displace the sample from the injection loop (~15 min), pump A was stopped and, after allowing the pressure to drop to 1 bar, disconnected from the system. Pump B was then connected to the system, and the inlet pressure was raised to 120 bar to deliver solvent B at 0.3 μL/min.

4.2.10. Mass Spectrometry

MALDI-TOF MS analyses were performed on a custom built instrument [1] equipped with delayed extraction and an ion mirror. The MALDI probe was spotted with 1 μL of a 1:3 (v/v) mixture of sample and a saturated solution of THAP in 50% acetonitrile. The instrument was operated in the positive ion mode with an accelerating potential of 20 kV and an extraction delay of 500 ns. A spectrum was produced by using *MoverZ* software to collect and average data generated from 30 laser pulses (ProteoMetrics, LLC, *http://www.proteometrics.com/software.html*).

On-line nano-LC ESI MS runs were performed on an LC-Q ion trap (Finnigan, San Jose, CA) mass spectrometer. The LC-Q was operated in the MS-mode with the spraying potential set to 2.5 kV (applied to the injector), the temperature of the

heated inlet-capillary at 180 °C, the capillary potential at 46 V, the tube-lens offset-potential at 30 V, and a maximum injection time of 50 ms. The instrument was set to acquire a full MS scan between 400 and 2000 *m/z* followed by an MS/MS scan. For operation in the MS/MS-mode, the maximum injection time was increased to 500 ms, the isolation width was set to 2 Da, and the relative collision energy was set to 35 % with a 30-ms activation time. Using a data dependent algorithm, the first most intense ion from a list of parent ions was selected for MS/MS providing the ion's MS signal was higher than 2×10^4 counts.

4.3. RESULTS

Previous studies employing a combination of photocrosslinking and mass spectrometry to investigate protein-DNA interactions almost exclusively used photoaffinity labeling either with low intensity UV-irradiation [8,10] or with laser-induced crosslinking [6,7,107]. In these early experiments, photosensitive nucleobase analogs (containing 5-bromouracil or 5-iodouracil in most cases) were used instead of natural nucleobases in order to increase the crosslinking efficiency. Typically, these analogs are irradiated at wavelengths longer than 300 nm, far beyond the absorbance bands of nucleic acids and proteins, to decrease protein or DNA photodegradation. In the opinion of the present authors, these non-natural nucleobases can structurally perturb the native nucleoprotein complex or

photochemically bind to amino acids that are not biologically involved in DNA-binding. For this reason, photoaffinity labeled DNA substrates were not used in this work.

The size of the dT_{30} substrate used in the present experiments was selected to correspond to the reported size of hRPA's DNA-binding size [92,93], and its polydeoxythymidine composition was chosen because the thymine's photocrosslinking reactivity is much greater than the other nucleic acid bases [17]. Several small scale electrophoretic mobility shift assay experiments were conducted in order to optimize the ratio of dT_{30} to hRPA needed to form a stable protein-DNA complex (data not shown). Small scale UV-crosslinking experiments were also performed to optimize the irradiation time. These crosslinking experiments (data not shown) clearly indicated hRAP70 as the major subunit involved in crosslinking to dT_{30}, a finding in agreement with results obtained by others [106].

Preparative scale crosslinking reactions were performed to obtain sufficient quantities of crosslinked material for characterization of the oligonucleotide binding domain of hRPA70. Aliquots were removed from the crosslinking mixture both before (Figure 4.1, Lane 1) and after UV-irradiation (Figure 4.1, Lane 2) and analyzed by SDS-PAGE. The three bands present in the non-irradiated sample (Lane 1) correspond respectively from top to bottom to hRPA70, a proteolytic fragment of hRPA70, and hRPA32. No nucleoproteins were present in Lane 1 because the noncovalent complexes present in the reaction mixture were disrupted

by the denaturing conditions of SDS-PAGE. These three bands were also present after UV-irradiation (Lane 2) along with the two new bands labeled XL-1 and XL-2. Given the appearance of these additional bands after UV-irradiation and the fact that their position in the gel was more retarded than hRPA70, these bands were assumed to contain crosslinked nucleoprotein complexes involving hRPA70. The XL-1 band, at a position corresponding to a relative molecular weight of about 80 kDa, was thought to corresponds to a hRPA70 x dT_{30} nucleoprotein complex, while the XL-2 band, at a position corresponding roughly to 90 kDa was presumed to correspond to a hRPA x $(dT_{30})_2$ nucleoprotein. This study focused only on the analysis of the XL-1 band leaving the more difficult multinucleotide complex for future study.

Observation of additional, higher molecular weight bands in Lane 2 of the gel shown in Figure 4.1 seems to be a clear indication that photochemical crosslinking produces stable complexes. The bands corresponding to XL-1 from all ten aliquots of the preparative scale reaction mixture were cut out and individually subjected to in-gel digestion with trypsin; the resulting ten digests were then pooled. Crosslinked tryptic peptides (nucleopeptides) were separated from non-crosslinked peptides using a procedure based on anion exchange chromatography developed in the authors' laboratory (see Experimental section). MALDI mass spectra of samples prepared in this manner consistently showed seven prominent ion species. A typical spectrum is shown in Figure 4.2.

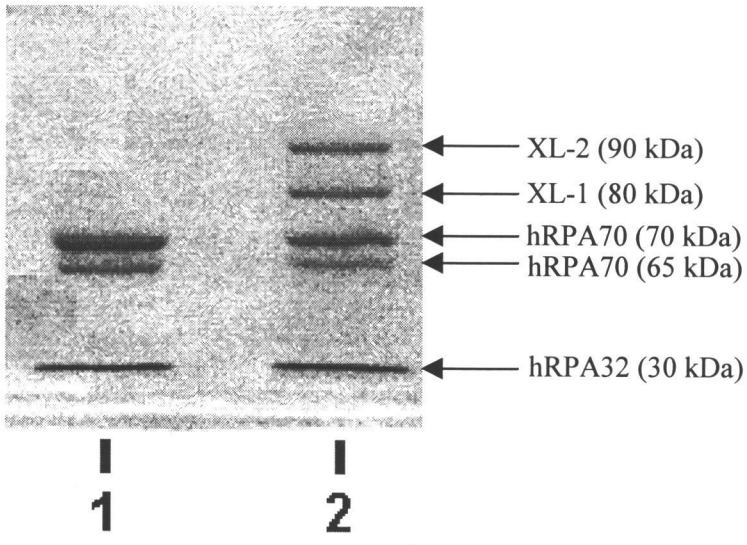


Figure 4.1. Analysis of a preparative scale crosslinking reaction of hRPA to dT_{30}. Mixtures containing hRPA and dT_{30} were irradiated at 254 nm for 1 min and separated by SDS-PAGE for subsequent isolation of the nucleopeptides. Portions were removed from a crosslinking mixture both before (Lane 1) and after irradiation (Lane 2), and dispersed in 8% SDS-PAGE. The protein bands were visualized by Coomassie stain. The positions of bands corresponding to hRPA32, a proteolytic product of hRPA70, hRPA70, and the putative crosslinks hRPA70 x dT_{30} and hRPA70 x $(dT_{30})_2$ (labeled respectively XL-1 and XL-2) are indicated by arrows to the right of the gel.

Previous mass spectrometric crosslinking studies performed in the authors' laboratory using poly-dT substrates [3,11-13] established that nucleopeptide masses are equal to within a mass unit to the sum of the neutral oligonucleotide mass plus the neutral peptide mass. This experimental finding is consistent with the prediction that "zero-length" covalent bonds form between DNA and proteins after low intensity UV-irradiation [17].

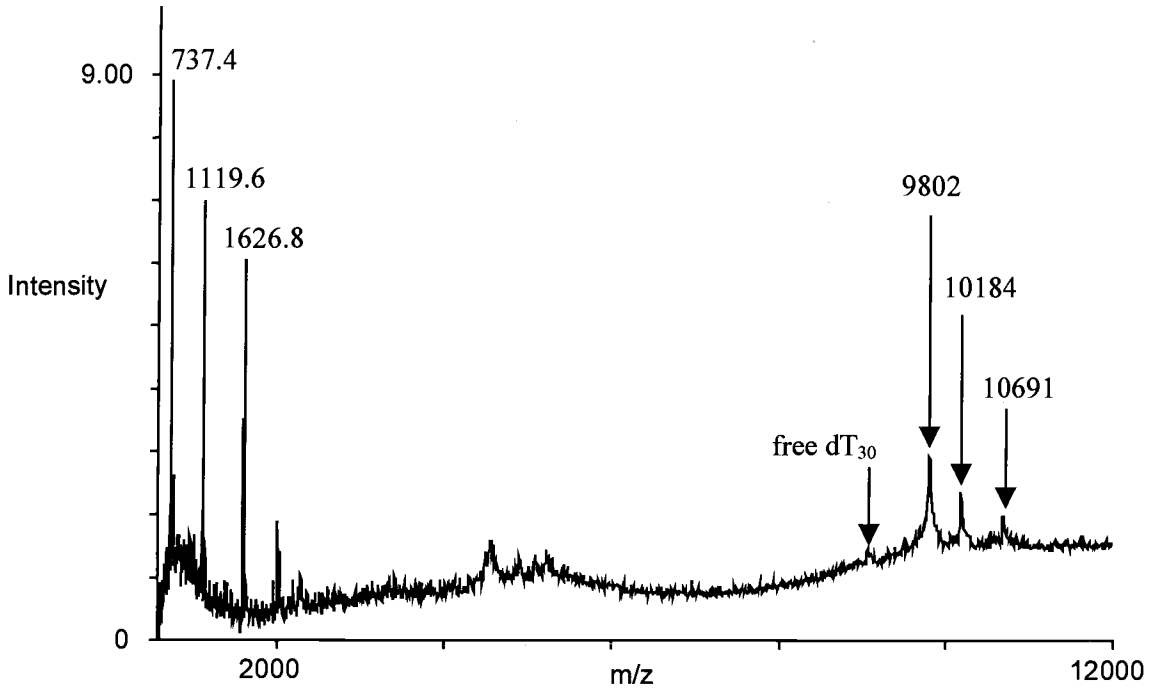


Figure 4.2. MALDI spectrum of hRPA70 peptides crosslinked to dT_{30}. The mass spectrum was produced from a pooled sample of purified in-gel digests of ten bands corresponding to an 80 kDa crosslink (XL-1 in Figure 4.1). An aliquot (1 μL) of the purified nucleopeptide sample was mixed with THAP matrix in a 1:3 ratio for the MALDI analysis. The spectrum shows the presence of both crosslinked (*m/z* 9082, 10184, 10691) and non-crosslinked hRPA70 peptides (*m/z* 737.4, 1119.6, 1626.8).

Subtracting the average molecular weight of dT_{30} (9064 Da) from the experimentally determined molecular weights of the three putative nucleopeptides evident in Figure 4.2 yields the masses of 738 Da, 1120 Da, and 1627 Da for the three corresponding protonated peptides. Ion signals corresponding to non-crosslinked peptides are also present in the nucleopeptide spectrum at *m/z* values of 737.4, 1119.6, and 1626.8.

The nucleopeptide sample was digested with nuclease P1 in order to reduce the oligonucleotide portion of the molecules. MALDI MS characterization of the nuclease digestion products (Figure 4.3) revealed three nucleopeptides (*m/z* 1363.8, 1746.0, and 2253.2) and the same free peptide (*m/z* 1626.8) observed in the MALDI mass spectrum from Figure 4.2. Subtracting the average molecular weight of dT_2 (626.4) from the signals observed at *m/z* 1363.8, 1746.0, and 2253.2 resulted in the molecular weights of 737.4 Da, 1119.6 Da and 1626.8 Da, for the corresponding protonated peptides, respectively. These same peptide masses were identified from the MALDI mass spectrum of the nucleopeptides prior to nuclease digestion (Figure 4.2).

The sample digested with nuclease P1 was analyzed on a nanoscale liquid chromatography electrospray ionization mass spectrometry (nanoLC-ESI/MS) system, specially constructed to preserve the chromatographic resolution provided by the nano-LC column and to avoid the additional dead volumes that usually accompany the coupling between a nanocolumn and the spraying needle.

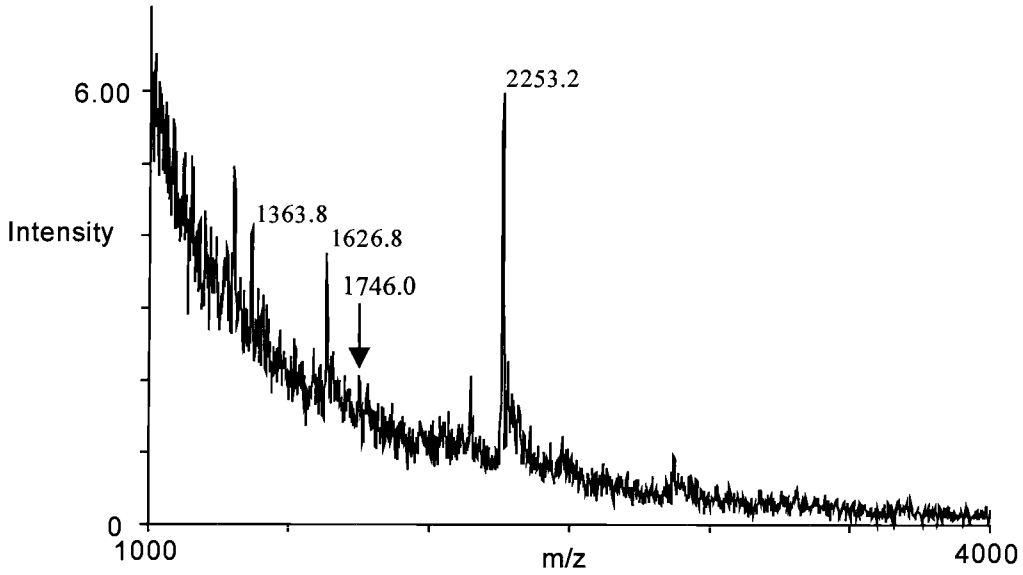


Figure 4.3. MALDI spectrum of hRPA70 nucleopeptides digested with nuclease P1. The mass spectrum was produced from a nuclease P1 digest of the purified nucleopeptide mixture whose MALDI spectrum is displayed in Figure 4.2. An aliquot (1 μL) of the P1-digested sample was mixed with THAP matrix in a 1:3 ratio for the MALDI analysis. Three of the observed signals correspond to products of the nucleolytic digestion of nucleopeptides (*m/z* 1363.8, 1746.0, and 2253.2). The signal at *m/z* 1626.8 is due to a non-crosslinked peptide.

The reconstructed ion chromatogram displayed in Figure 4.4 indicates the presence of three free peptides in the nuclease P1 digest.

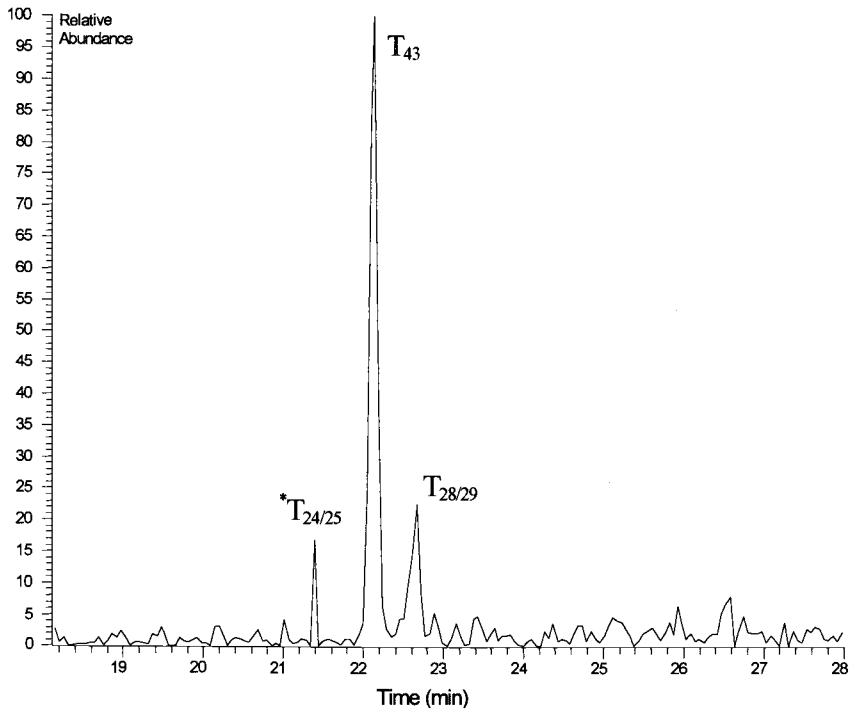


Figure 4.4. Nano-LC/ESI-MS reconstructed ion chromatogram of the nuclease P1 digest. A portion (2 uL) of nuclease P1 digested sample (Figure 4.3) was mixed with solvent A (4 uL), injected onto the 40 cm nano-LC column, and analyzed by nanoLC/ESI-MS. Ions were detected from *m/z* 400 to 2000. The chromatogram was reconstructed from the doubly protonated molecular ion signals produced by the three peptides found in the MALDI spectrum from Figure 4.2. The tryptic peptides are labeled starting from the N-terminus of hRPA70.

The ESI MS/MS spectra obtained from the peptides that give rise to the three chromatographic peaks seen in Figure 4.4 are displayed in Figure 4.5A-C.

Figure 4.5. MS/MS spectra of the hRPA70 tryptic peptides corresponding to the chromatographic peaks seen in the ion chromatogram of the nuclease P1 digest (Figure 4.4). Tandem mass spectra were obtained for the (**A**) singly charged ion at *m/z* 737.4, (**B**) the doubly charged ion at *m/z* 560.4, and (**C**) the doubly charged ion at *m/z* 813.9. The sequence information provided by these mass spectra revealed these peptides respectively to be T_{43}, $T_{28/29}$ and $T_{24/25}$ minus its five C-terminal residues.

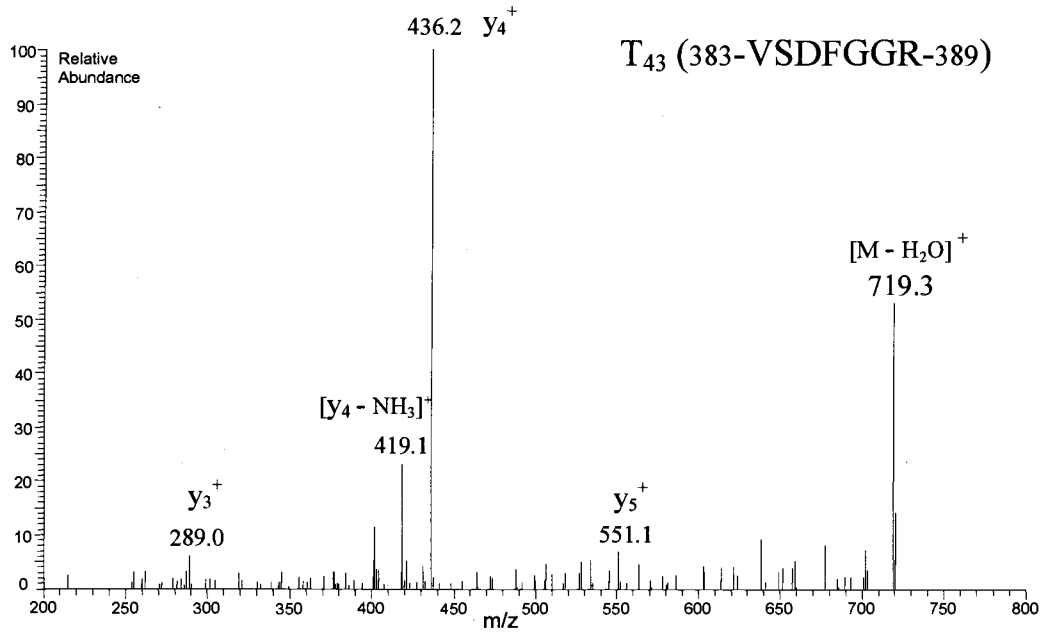


Figure 4.5A. MS/MS spectrum of *m/z* 737.4.

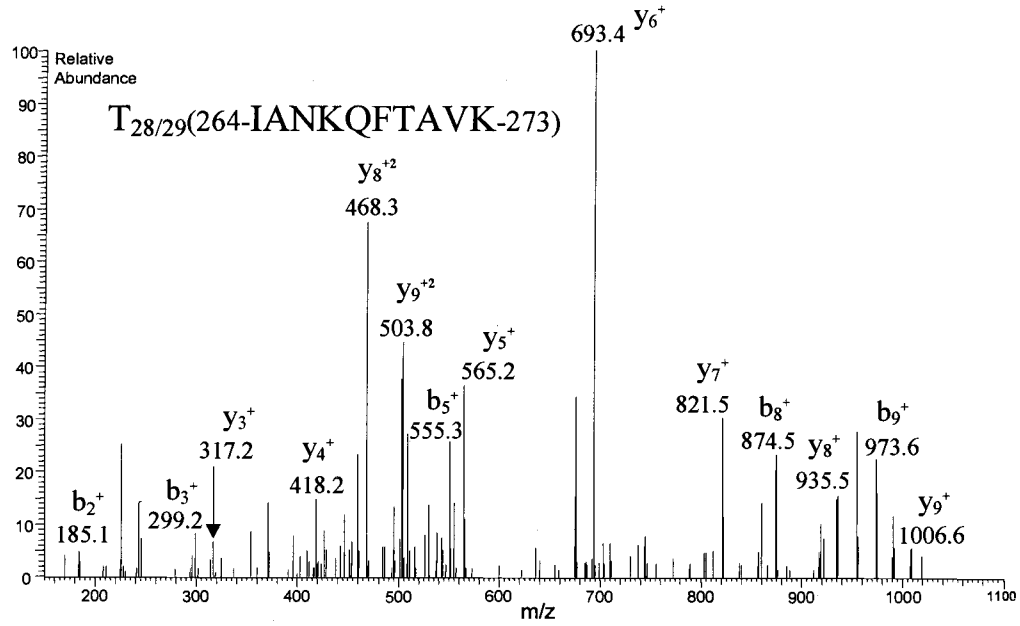


Figure 4.5B. MS/MS spectrum of *m/z* 560.4.

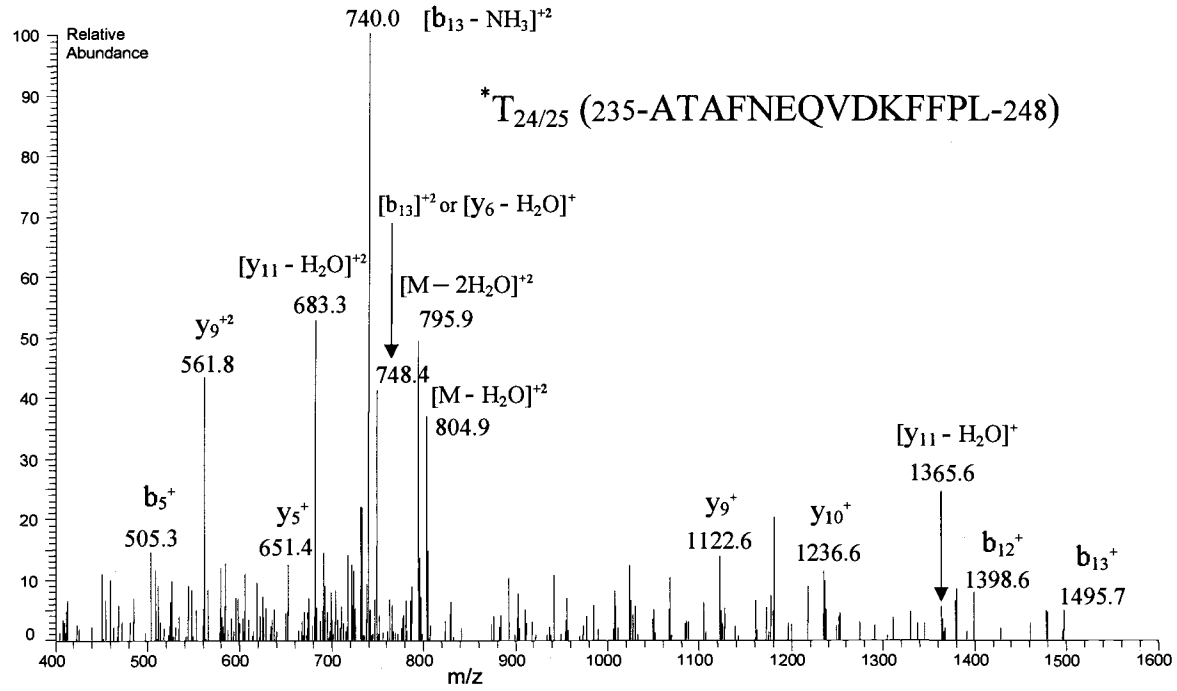


Figure 4.5C. MS/MS spectrum of *m/z* 813.9.

ESI-MS/MS was attempted on the nucleopeptides present in the nuclease P1 digest, but data could only be obtained for the nucleopeptide $T_{28/29} \times dT_2$ (*m/z* 1746.0), even though this nucleopeptide produced the weakest signal of the three nucleopeptides observed in the MALDI spectrum of the tryptic digest (Figure 4.3). The nanoLC ESI-MS/MS spectrum of doubly protonated $T_{28/29} \times dT_2$ (Figure 4.6) shown that the peptides's backbone fragments mainly into *b*-ions and that two of these ions ($^{*}b_8$ and $^{*}b_9$) have a dinucleotide tag (626.4 Da) attached to them.

The hRPA nucleopeptide sample was extensively digested with Carboxypeptidase Y (CPY) for 48 h then cleaned-up by immobilized metal affinity chromatography (IMAC).

The MALDI spectrum of this CPY digest is exhibited in Figure 4.7. The two strongest signals in the range *m/z* 9,000 – 12,000 of this spectrum correspond to free dT_{30} (*m/z* 9065) and T_{43} x dT_{30} (*m/z* 9802). Five other weaker signals in this envelope of ions presumably correspond to C-terminal truncated nucleopeptides. Prior to CPY digestion (Figure 4.2), several signals were observed in the low-mass region (*m/z* 500-2000), but no signals were observed in this region after CPY digestion. It is presumed that the free $^{*}T_{24/25}$, $T_{28/29}$ and T_{43} peptides were completely digested by CPY.

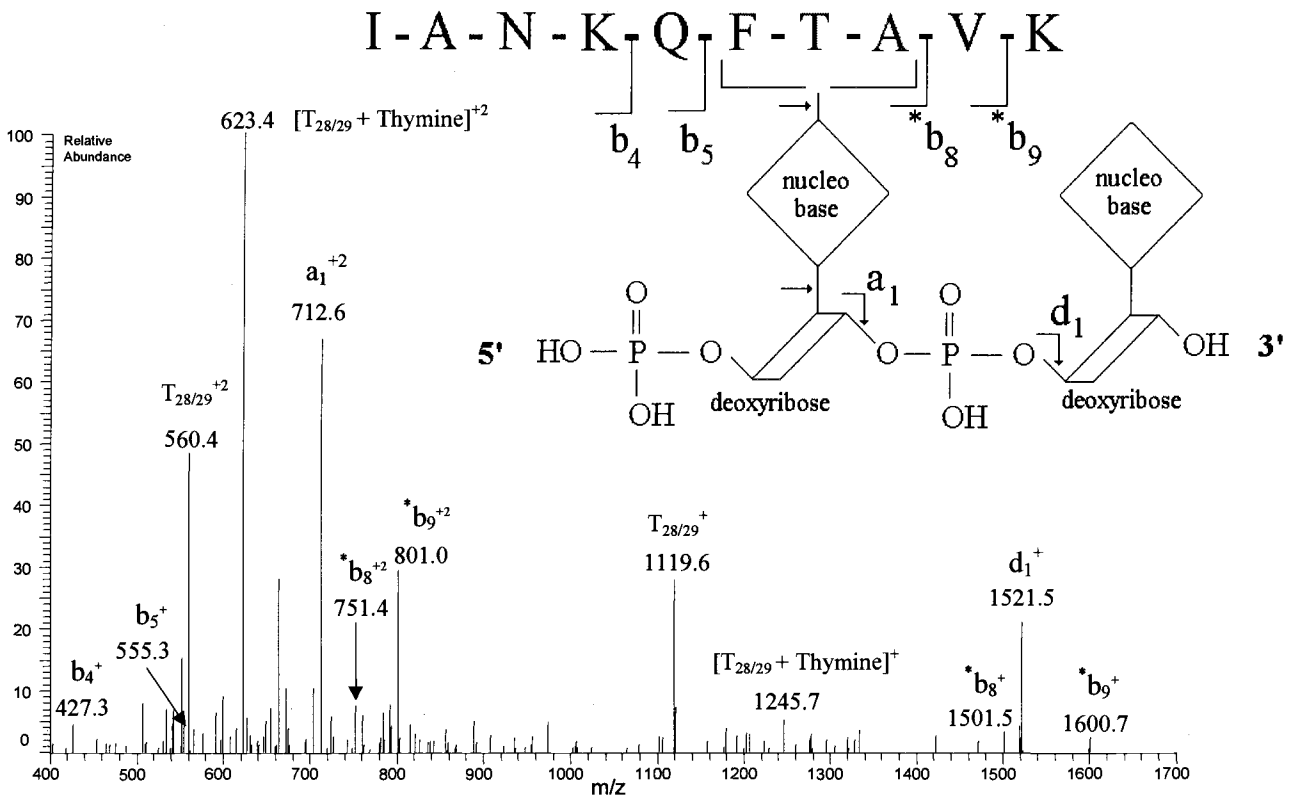


Figure 4.6. MS/MS spectrum of the tryptic peptide $T_{28/29}$ crosslinked to dT_2 nucleotide ($T_{28/29}$ x dT_2). The doubly protonated ion (*m/z* 873.4) was fragmented. The fragmentation pattern observed for this nucleopeptide is summarized in the sequence shown in the inset. The b ions labeled as $^{*}b_n$ ions have a dinucleotide tag (626.4 Da) attached.

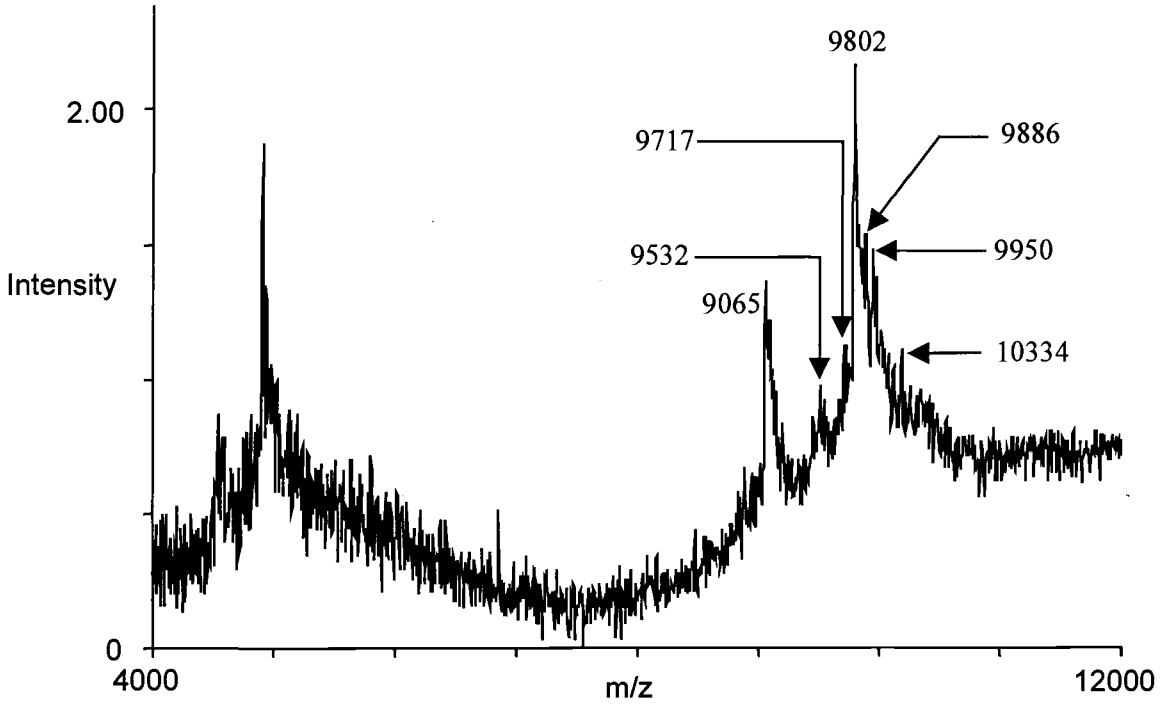


Figure 4.7. MALDI mass spectrum of a CPY digested nucleopeptide sample. 1 μL of digested sample was mixed with 3 μL of THAP matrix solution and analyzed. The signals observed at *m/z* 9532, 9717, 9886, 9950 and 10334 were assigned to several C-terminal truncated nucleopeptides (see Table 4.1 for identification).

4.4. DISCUSSION

The information contained in the MALDI mass spectra shown in Figure 4.2 and 4.3 was sufficient to tentatively identify two of the three putatively crosslinked peptides of hRPA70. The ion signal at *m/z* 737.4 is consistent with the T_{43} tryptic peptide (383-VSDFGGR-389), and the ion signal at *m/z* 1119.6 corresponds to the tryptic fragment $T_{28/29}$ (264-IANKQFTAVK-273), which contains a missed cleavage site. No hRPA tryptic peptide's mass matches the *m/z* value 1626.8 that is prominent in Figures 4.2 and 4.3.

The appearance of signals corresponding to noncrosslinked peptides in the MALDI spectra in Figures 4.2 and 4.3 suggests that nucleopeptide decomposition might have taken place. This possibility is further indicated by the observation of a weak signal corresponding to free dT_{30} in the spectrum from Figure 4.2.

Sequence data obtained from the MS/MS spectra of the peptides that appear at 22.1 min and 22.6 min in the ion chromatogram (Figure 4.5A-B) confirmed the identities of T_{43} and $T_{28/29}$ that were tentatively made from the MALDI mass spectrum from Figure 4.2. The ESI mass spectrum of the species appearing at 21.4 min in the ion chromatogram (data not shown) exhibited a protonated ion signal at *m/z* 1626.8, which almost exactly equals the *m/z* value determined for one of the peptides that shows up in the MALDI mass spectra in a putative crosslinked form (*m/z* 10,691, Figure 4.2) and a noncrosslinked form (Figure 4.2 and 4.3). None of the predicted tryptic peptides of hRPA70 correspond to a protonated mass of

1626.8; however, the data in the ESI MS/MS spectrum of the doubly protonated form of the unknown peptide (Figure 4.5C) is conform to the sequence 235-ATAFNEQVDKFFPL-248, which is that of the tryptic peptide $T_{24/25}$ minus its last five C-terminal amino acids ($^{*}T_{24/25}$).

The appearance of ion signals corresponding to free peptides in the MALDI mass spectra of the nucleopeptide sample (Figure 4.2), the MALDI spectrum of the nuclease P1 digest (Figure 4.3), and in the nanoLC/ESI mass spectra of the nuclease P1 digest (Figure 4.4) implies an inherent instability in the bands formed by UV-crosslinking. Whatever its cause, this instability reduces the nucleopeptide concentration. On the one hand, this adds to the difficulty of performing tandem mass spectrometry on the nucleopeptides themselves, but on the other hand it simplifies identification of the peptide components of the nucleopeptides.

MS/MS data could not be obtained for the T_{43} x dT_2 and the $^{*}T_{24/25}$ x dT_2 conjugates, but MS/MS data from $T_{28/29}$ x dT_2 was forthcoming. The appearance of b_4 and b_5 fragments in the latter's MS/MS spectrum (Figure 4.6) indicated that no dT_2 is attached to the N-terminal segment 264-IANKQ-268 whereas the appearance of $^{*}b_8$ and $^{*}b_9$ fragments implies that a dT_2 is attached to a residue in the segment 264-IANKQFTA-271. Thus, at least one of the residues in the sequence 269-FTA-271 must be a site of crosslinking. The intense signals produced by the singly protonated (*m/z* 1119.6) and doubly protonated (*m/z* 560.4) forms of the free $T_{28/29}$ (Figure 4.6) suggest that the dinucleotide tag is easily lost from this nucleopeptide. This might account for only two $^{*}b$ fragment ions that appear in the spectrum. The

ion signals at *m/z* 1245.7 and *m/z* 623.4 correspond respectively to singly and doubly protonated $T_{28/29}$ (1118.6 Da) crosslinked to thymine (126.1 Da). This is additional evidence for the occurrence of "zero-length" photocrosslinking reactions between the nucleic acid base thymine and amino acids.

In some instances a combination of exhaustive proteolytic digestion and MALDI-TOF can provide information about the sites of crosslinked amino acids in nucleopeptides [8]. The enzyme is used to remove non-crosslinked residues from the N-terminal (aminopeptidase M) or from the C-terminal (carboxypeptidase) end of the nucleopeptide. Since the proteolysis stops when the enzyme encounters a crosslinked amino acid, the result is a truncated nucleopeptides whose molecular weight can be determined by MALDI-TOF. This approach was attempted with carboxypeptidase Y (CPY) in this study because carboxypeptidases have been successfully used in combination with MALDI-TOF for C-terminal sequencing [108-110] and CPY, in particular, cleaves nonspecifically from the C-terminus all residues, including proline. In using this method, one must be aware that the kinetics of enzymatic release for different amino acids varies considerably [108] and, furthermore, that it is quite possible for steric interference from the oligonucleotide to limit the extent of the carboxypeptidase digestion.

It was necessary to use IMAC to clean up the CPY digested nucleopeptide sample (Figure 4.7) in order to achieve a MALDI mass accuracy of about 0.05% in the mass range of 9,000 – 12,000 Da. The separation principle of IMAC is based on electronic interactions between phosphate groups and metal ions (Fe^{3+} in this case)

immobilized (chelated) to a stationary phase. This technique was recently employed for enrichment of phosphorylated peptides from crude peptide mixtures [111], as well as for the isolation of nucleopeptides [10]. Instead of using the packed column format, it was found preferable in this study to use metal chelate (MC) ZipTips (Milipore) designed for low volume (1-5 μL) samples.

Table 4.1 displays all possible average molecular weights of truncated hRPA70 nucleopeptides generated by a C-terminal cleavage.

$^{*}T_{24/25}$ nucleopeptide	Average Mass (Da)	$T_{28/29}$ nucleopeptide	Average Mass (Da)	T_{43} nucleopeptide	Average Mass (Da)
$^{*}T_{24/25}$ x dT_{30}	10691	$T_{28/29}$ x dT_{30}	10184	T_{43} x dT_{30}	9802
ATAFNEQVDKFFP x dT_{30}	10578	IANKQFTAV x dT_{30}	10056	VSDFGG x dT_{30}	9646
ATAFNEQVDKFF x dT_{30}	10481	**IANKQFTA x dT_{30}**	**9957**	VSDFG x dT_{30}	9589
ATAFNEQVDKF x dT_{30}	**10334**	**IANKQFT x dT_{30}**	**9886**	**VSDF x dT_{30}**	**9532**
ATAFNEQVDK x dT_{30}	10187	IANKQF x dT_{30}	9785	VSD x dT_{30}	9385
ATAFNEQVD x dT_{30}	10059	IANKQ x dT_{30}	9638	VS x dT_{30}	9270
ATAFNEQV x dT_{30}	**9944**	IANK x dT_{30}	9510	V x dT_{30}	9183
ATAFNEQ x dT_{30}	9845	IAN x dT_{30}	9382		
ATAFNE x dT_{30}	**9717**	IA x dT_{30}	9268		
ATAFN x dT_{30}	9588	I x dT_{30}	9197		
ATAF x dT_{30}	9474				
ATA x dT_{30}	9327				
AT x dT_{30}	9256				
A x dT_{30}	9155				

Table 4.1. Protonated average molecular weights of all possible truncated hRPA70 nucleopeptides generated by a C-terminal cleavage. The molecular weights in bold letters correspond to MALDI signals observed in Figure 4.7.

By comparing the masses in Table 4.1 with the masses determined experimentally from the CPY-digest spectrum shown in Figure 4.7, it was determined that the signal at *m/z* 9532 corresponds to the T_{43} derived nucleopeptide VSDF x dT_{30}, the signal at *m/z* 9886 matches the $T_{28/29}$ derived nucleopeptide IANKQFT x dT_{30}, and the signals at *m/z* 9717 and *m/z* 10334 Da conform respectively to the $^{*}T_{24/25}$ derived nucleopeptides ATAFNE x dT_{30} and ATAFNEQVDF x dT_{30}. Although the mass is too inaccurate to be conclusive, the signal observed at *m/z* 9950 Da could have been produced either by IANKQFTA x dT_{30} or by ATAFNEQV x dT_{30}. The experimental evidence obtained from the CPY digestion suggests that crosslinking occurred to at least one amino acid from the sequence 383-VSDF-386 located in T_{43}, to least one residue in the segment 269-FT-270 in $T_{28/29}$, and to at least one amino acid in the N-terminus segment 235-ATAFNE-240 of $^{*}T_{24/25}$.

Using the data obtained from the carboxypeptidase digestion in conjunction with the tandem mass spectral data it is possible to deduce a map of the DNA binding domain of hRPA70. These data are in good agreement with previous information obtained from X-ray crystallography [56,99] and site-directed mutagenesis [98,100]. Each of the crosslinked peptides from hRPA70 contains one phenylalanine residue previously identified as being involved in base stacking interactions: F238 belongs to $^{*}T_{24/25}$ (235-ATA**F**NEQVDKFFPL-248), F269 is located within $T_{28/29}$ (264-IANKQ**F**TAVK-273) and F386 is from T_{43} (383-VSD**F**GGR-389). According to the crystallography data [56], W361 has a weak base stacking interaction in comparison with the phenylalanine residues (F238,

F269 and F386); hence, it is not surprising that no mass spectrometric evidence was found for the involvement of W361 in crosslinking.

Figure 4.8 shows the three-dimensional structure of an hRPA70 fragment (residues 183-420) obtained from crystallography data [56,99].

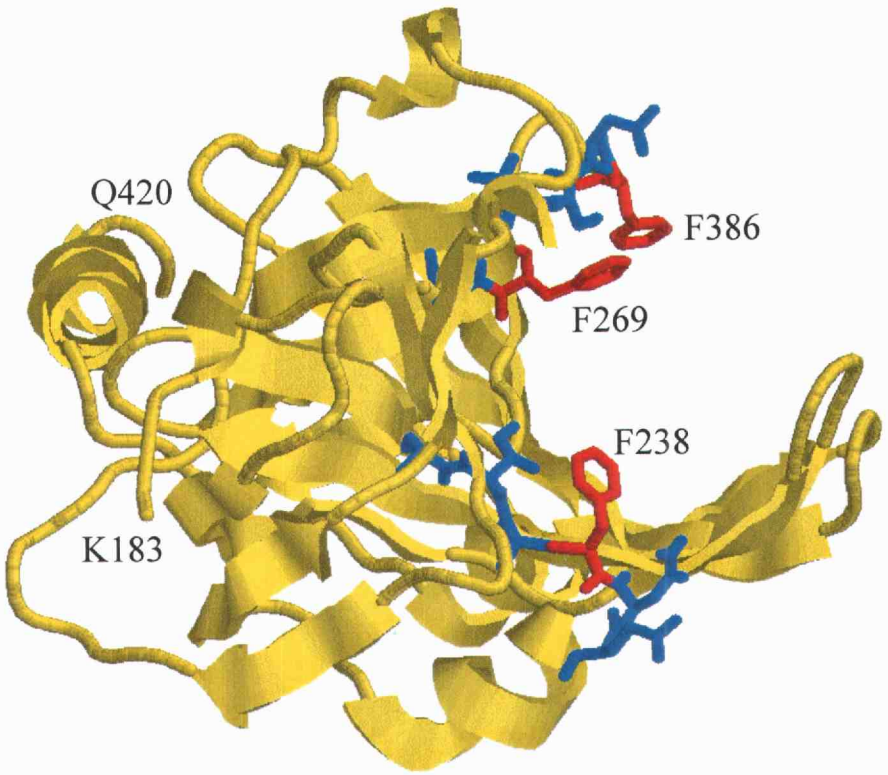


Figure 4.8. 3D structure of an hRPA70 fragment (residues 183-420) containing the major DNA-binding region of the protein obtained from crystallography data [56,99]. Several amino acids located in the DNA-binding channel of the protein were identified by UV-photocrosslinking and mass spectrometry as being crosslinked to oligonucleotide dT_{30}. Each of the three distinct peptide regions colored in blue have at least one crosslinked amino acid and these peptides are located in the DNA-binding subdomain A (F269/T270 and 235-ATAFNE-240) or in subdomain B (383-VSDF-386). All aromatic amino acids contained in these peptides are colored in red. The drawing was generated using the RasMol program [86].

In this picture, helices are represented as spirals and β sheets are represented as ribbons with arrows pointing in the C-terminal direction. All information obtained to date by UV-photocrosslinking and mass spectrometry is summarized in this figure.

Three distinct peptide regions colored in blue contain at least one crosslinked amino acid. These peptides are located in the DNA-binding subdomain A (F269/T270 and 235-ATAFNE-240) or in subdomain B (383-VSDF-386). There is no evidence that a third binding domain (C) of hRPA is essential for binding DNA. All aromatic amino acids contained in these peptides are colored in red. They can form base stacking interactions with the DNA and, in the authors' opinion, are the residues most likely to be involved in crosslinking. At this point, one can only speculate that these aromatic amino acids are required for the formation of a stable protein-DNA complex. More experiments will be required to confirm this speculation and to show whether other amino acids are involved in crosslinking.

5. SUMMARY AND CONCLUSIONS

The use of mass spectrometry for structurally characterizing nucleopeptide complexes produced from UV-catalyzed crosslinking reactions is increasing [10]. As with mass spectrometric analysis, the key to success in this application lies more in the preparation of the samples than in the operation of the instrumentation. In this respect, the dual chemical nature of the nucleopeptides makes it particularly difficult to prepare them for MALDI MS and LC/ESI MS.

One of the goals of the work presented in this thesis was test the general applicability of a previously developed nucleopeptide purification protocol [11,12]. This procedure, developed initially for the {Ung (*E. coli* uracil DNA glycosylase) dT_{20} system}, has been successfully applied in the present study to the isolation of nucleopeptides from the UV-catalyzed crosslinking reaction of hRPA and dT_{30}, thus illustrating the procedure's robustness. In the case of {Ung/dT_{20}} system, five peptides corresponding to Ung T_{18}, T_{11}, and three truncated forms of T_6 (minus the last three, five, and seven C-terminal amino acids) were identified as being crosslinked to dT_{20} [11,12]. In the case of the {hRPA70/dT_{30}} system, three peptides corresponding to hRPA70 T_{43}, $T_{28/29}$, and one truncated form of $T_{24/25}$ (minus the last five C-terminal aminoacids) have been shown to crosslink to dT_{30}.

Comparison of the hRPA-results presented in this thesis with those obtained in the earlier Ung-studies reveals at least three characteristics that may be general to the

purification and mass analysis of nucleopeptides. First, the tryptic peptide part of a nucleopeptide can undergo partial hydrolysis from the C-terminus during the isolation procedure. Since this hydrolysis has not been observed for all nucleopeptides in an isolated pool, it may be specific to a particular nucleopeptide's amino acid sequence in relation to the location of the crosslinked amino acid residue. Tentative identification of an individual peptide from its signal in the MALDI spectrum relies heavily on knowing the specific pattern of cleavage that produces the peptide during proteolysis. Since the peptide component of a partially hydrolyzed nucleopeptide is not tryptic as would be expected from the history of its preparation, it is nearly impossible to recognize it from its signal in a MALDI mass spectrum.

The accuracy of the MALDI measurement is usually not good enough by itself to identify a segment in a protein's sequence without knowing something about the proteolytic process that produced it. Therefore, tandem MS experiments become necessary to establish the identity of any nucleopeptides that hydrolyze during sample preparation. Second, nuclease P1 digestion of the oligonucleotide component of a nucleopeptides has never been extensive enough to result in a peptide crosslinked to a single nucleotide. In the cases of both Ung and hRPA70, P1 digestion was never observed to proceed beyond a dinucleotide. Third, the ESI MS/MS spectra of the nuclease digested nucleopeptides produce, as might be well expected, very complex spectra. In most cases, but not all, CID induced fragmentation of the oligonucleotide part occurs more readily than fragmentation of

the peptide part. Before tandem mass analysis of nucleopeptides becomes simple and routine, methods must be found that enhance fragmentation of the peptide part at the expense of the fragmentation of the oligonucleotide part.

Mass spectrometric analysis of nucleoprotein complexes seems to be primarily limited by the amount of crosslinked product that is required prior to starting the purification. In isolating nucleopeptides from the crosslinking of *E. coli* Ung to dT_{20} [11,12], over 1 nmol of crosslinked product was required to subsequently isolate nucleopeptides for mass spectrometric analysis. In the present work with the hRPA, approximately 800 pmol of crosslinked material was needed prior to starting the nucleopeptide purification. The requirement for these large amounts of starting material is a result of four factors: i) the number of steps in the purification procedure, ii) the distribution of nucleopeptide product amongst a number of peptides, iii) the hydrolytic loss of portions of the nucleopeptide's peptidic component during purification, iv) the cleavage of the nucleopeptide into free peptide and free oligonucleotide during purification. Aside from factor ii), which is dependent on the biological interactions of the protein-DNA system under study, efforts to minimize these factors must be made to reduce the amount of starting material needed for mass spectrometric characterization. An example of a move toward this goal is a recent introduction of a microscale method for the analysis of crosslinked nucleoprotein complexes [10].

The ultimate outcome of using a mass spectrometric analysis of a peptide-nucleic acid complex is the identities of specific amino acids that are crosslinked to

oligonucleotide. The rationale for seeking such data is that it can provide insight into the mechanism by which the protein interacts with its nucleic acid substrate. To date, tandem mass spectrometry has been successfully used to identify specific aminoacids in Ung x dT_{20} and hRPA x dT_{30} protein-DNA complexes that crosslinked to the DNA strand. The fact that the crosslinked residues discovered by mass spectrometry are several of the same amino acids that have been indicated by X-ray crystallography to play key roles in the biological function of Ung and hRPA suggests that it may not be too long before it will be possible to use mass spectrometry to learn something about the biology of protein-nucleic acid interactions.

The research objective of this thesis was to identify specific amino acids from Ung and hRPA that are involved in binding DNA, to characterize the DNA-binding domains of these proteins. Following a protocol developed in our laboratory, three amino acids were identified from a short peptide region from Ung (H187, S189 and H194) that are involved in binding to the dT_{20} oligonucleotide. Using the same protocol, three aromatic amino acids (F238, F269 and F386) from hRPA70 that photocrosslink to dT_{30} were putatively identified. These results illustrate that photochemical crosslinking and mass spectrometry can be successfully employed in the study of non-covalent nucleoprotein complexes. In addition, these experimental results demonstrate that mass spectrometry is an extremely versatile and powerful tool in biology.

I believe that my work will contribute to the overall understanding of the protein-DNA interactions and I hope that more new applications of mass spectrometry in this field will be seen in the future.

BIBLIOGRAPHY

1. Jensen, O. N., Barofsky, D.F., Young, M.C., Von Hippel, P.H., Swenson, S., and Seifried, S (1993) "Direct observation of UV-crosslinked protein-nucleic acid complexes by matrix-assisted laser desorption ionization mass spectrometry" *Rapid Commun. Mass Spectrom.* **7**, 496-501.

2. Veenstra, T. (1999) "Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes" *Biochem. Biophys. Res. Commun.* **257**, 1-5.

3. Bennett, S. E., Jensen, O.N., Barofsky, D.F., and Mosbaugh, D.W. (1994) "UV-catalyzed cross-linking of *Escherichia coli* uracil-DNA glycosylase to DNA. Identification of amino acid residues in the single-stranded DNA binding site" *J. Biol. Chem.* **269**, 21870-21879.

4. Connor, D. A., Falick, A.M., Young, M.C., and Shetlar, M.D. (1998) "Probing the binding region of the single-stranded DNA-binding domain of rat DNA polymerase beta using nanosecond-pulse laser induced cross-linking and mass spectrometry" *Photochem. Photobiol.* **68**, 299-308.

5. Qin, J. and Chait, B.T. (1997) "Identification and characterization of posttranslational modifications of proteins by MALDI ion trap mass spectrometry" *Anal. Chem.* **69,** 4002-4009.

6. Urlaub, H., Thiede, B., Muller, E.C., Brimacombe, R., and Wittmann-Liebold, B. (1997) "Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30 S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing" *J. Biol. Chem.* **272**, 14547-14555.

7. Golden, M. C., Resing, K.A., Collins, B.D., Willis, M.C., and Koch, T.H. (1999) "Mass spectral characterization of a protein-nucleic acid photocrosslink" *Protein Science* **8**, 2806-2812.

8. Wang, Q., Shoeman,R., and Traub, P. (2000) "Identification of amino acid residues of the amino terminus of vimentin responsible for DNA binding by enzymatic and chemical sequencing and analysis by MALDI-TOF" *Biochemistry* **39,** 6645-6651.

9. Rieger, R.A., McTigue, M. M., Kycia, J.H., Gerchman, S.E., Grollman, A.P. and Iden, C.R. (2000) "Characterization of a cross-linked DNA-Endonuclease VIII repair complex by electrospray ionization mass spectrometry" *J. Am. Soc. Mass Spectrom.* **11**, 505-515.

10. Steen, H., Petersen, J., Mann, M., and Jensen, O.N (2001) "Mass spectrometric analysis of a UV-cross-linked protein-DNA complex: tryptophans 54 and 88 of *E. coli* SSB cross-link to DNA" *Protein Sci*ence **10**, 1989-2001.

11. Gafken, P. R., Mosbaugh, D.W., and Barofsky, D.F. (2000) "Characterization of the UV-crosslinked *E. coli* uracil-DNA glycosylase-dT_{20} nucleoprotein interface by MALDI-MS and ESI-MS/MS" *Proc. 48th ASMS Conf. Mass Spectrom. Allied Top., Long Beach, CA*, 54-55.

12. Gafken, P. R. (2000) "Characterization of UV-crosslinked protein-nucleic acid interfaces by MALDI-MS and ESI-MS/MS" *PhD thesis, Oregon State University.*

13. Doneanu, C. E., Gafken, P.R., and Barofsky, D.F. (2001) "Mass spectrometric analysis of UV-crosslinked protein-nucleic acid complexes" *Proc. 49th ASMS Conf. Mass Spectrom. Allied Top., Chicago, IL.*

14. Doneanu, C. E., Gafken, P.R., and Barofsky, D.F. (2002) "Mass spectrometric analysis of UV-crosslinked protein-nucleic acid complexes: identification of amino acid residues in the single-stranded DNA-binding domain of human replication protein A" *Proc. 50th ASMS Conf. Mass Spectrom. Allied Top., Orlando, FL.*

15. Smith, K. C. (1962) "Dose dependent decrease in extractibility of DNA from bacteria following irradiation with ultraviolet light or with visible light plus dyes" *Biochem. Biophys. Res. Commun.* **8**, 157-163.

16. Alexander, P. and Morison, H (1962) "Crosslinking of deoxyribonucleic acid to protein followind ultraviolet irradiation of different cells" *Nature* **194**, 882-883.

17. Shetlar, M. D. (1980) "Cross-linking of proteins to nucleic acids by ultraviolet light" *Photochem. Photobiol. Rev.* **5**, 105-197.

18. Meisenheimer, K. and Koch, T. (1997). "Photocross-linking of nucleic acids to associated proteins" *Crit. Rev. Biochem. Mol. Biol.* **32,** 101-140.

19. Wiliams, K. R., and Konigsberg, W.H. (1991) "Identification of amino acid residues at interface of protein-nucleic acid complexes by photochemical cross-linking" *Methods Enzymol.* **208**, 516-539.

20. Karas, M., Bachmann, D., Bahr, U. and Hillenkamp, F. (1987) "Matrix-assisted ultraviolet laser desorption of non-volatile compounds" *Int. J. Mass Spectrom. Ion Proc.* **78**, 53-68.

21. Karas, M. and Hillenkamp, F. (1988) "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons" *Anal. Chem.* **60**, 2299-2301.

22. Vestal, M. L., Juhasz, P. and Martin, S.A. (1995) "Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry" *Rapid Commun. Mass Spectrom.* **9**, 1044-1050.

23. Brown, R.S. and Lennon, J.J. (1995) "Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer" *Anal. Chem.* **67**, 1998-2003.

24. Tang, X., Beavis, R., Ens, W., Lafortune, F., Schueler, B. and Standing, K.G. (1988) "A secondary ion time-of-flight mass spectrometer with an ion mirror" *Int. J. Mass Spectrom. Ion Proc.* **85**, 43-67.

25. Spengler, B., Kirsch, D., Kaufmann, R. and Jaeger, E. (1992) "Peptide sequencing by matrix-assisted laser-desorption mass spectrometry" *Rapid Commun. Mass Spectrom.* **2**, 105-108.

26. Shevchenko, A., Loboda, A., Shevchenko, A., Ens, W. and Standing, K.G. (2000) "MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research" *Anal. Chem.* **72**, 2132-2141.

27. Krutchinsky, A.N., Kalkum, M. and Chait, B. T. (2001) "Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer" *Anal. Chem.* **73**, 5066-5077.

28. Laiko, V.V., Moyer S. C. and Cotter, R. J. (2000) "Atmospheric pressure MALDI/ion trap mass spectrometry" *Anal. Chem.* **72**, 5239-5243.

29. Medzihradszky, K.F., Campbell, J. M., Baldwin, M.A., Falick, A.M., Juhasz, P, Vestal, M.L. and Burlingame, A.L. (2000) "The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer" *Anal. Chem.* **72**, 552-558.

30. Yamashita, M. and Fenn, J.B. (1984) "Electrospray ion source. Another variation on the free-jet theme" *J. Phys. Chem.* **88**, 4451-4459.

31. Meng, C. K., Mann, M. and Fenn, J.B. (1988) "Of protons or proteins - A beam's a beam for a' that" *Z. Phy. D: Atoms, Mol. Clusters* **10**, 361-368.

32. Taylor, G. (1964) "Disintegration of water drops in an electric field" *Proc. R. Society London Ser.* A **280**, 383.

33. Smith, R.D., and K.J. Light-Wahl (1993) "The observation of non-covalent interactions in solution by electrospray ionization mass spectrometry: promise, pitfalls and prognosis" *Biol. Mass Spectrom.* **22**, 493-500.

34. Dürcks, T. and Röllgen, F.W. (1988) "Ionization conditions and ion formation in electrohydrodynamic mass spectrometry" *Int. J. Mass Spectrom. Ion Processes* **148**, 123-144

35. Fenn, J.B. (1993) "Ion formation from charged droplets: roles of geometry, energy and time" *J. Am. Soc. Mass Spectrom.* **4**, 524-533.

36. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1990) "Electrospray ionization – principles and practice" *Mass Spectrom. Rev.* **9**, 37-90.

37. Kebarle, P and Tang, L., (1993) "From ions in solution to ions in the gas phase. The mechanism of electrospray mass spectrometry" *Anal. Chem.* **65,** 972A-986A.

38. Van Berkel, G. J. and Zhou, F. (1995) "Characterization of an electrospray source as a controlled-current electrolytic cell" *Anal. Chem.* **67**, 2916-2923.

39. Wilm, M.S. and Mann, M. (1994) "Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last" *Int. J. Mass Spectrom. Ion Proc.* **136**, 167-180.

40. Wilm, M.S. and Mann, M. (1996) "Analytical properties of the nanoelectrospray ion source" *Anal. Chem.* **68**, 1-8.

41. Cole, R.B. (1997) "Electrospray coupling to mass analyzers" book chapter in *Electrospray ionization mass spectrometry*, John Wiley & Sons, New York, 175-320.

42. Yost, R.A., Boyd, R. K. (1990) "Tandem mass spectrometry: quadrupole and hybrid instruments" *Methods Enzymol.* **193**, 154-200.

43. Morris, H.R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R.S., Hoyes, J. and Bateman R.H. (1996) "High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer" *Rapid Commun Mass Spectrom.* **10**, 889-896.

44. Zubarev, R.A., Horn, D. M., Fridriksson, E.K., Kelleher, N.L., Kruger, N.A., Lewis, M.A., Carpenter, B.K. and McLafferty, F.W. (2000) "Electron capture dissociation for structural characterization of multiply charged protein cations" *Anal. Chem.* **72**, 563-573.

45. Chait, B. T. and Kent, B.H. (1992) "Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins" *Science* **257**, 1885-1894.

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

47. McGuffin, V. L. and Novotny, M. (1983) "Optimization and evaluation of packed capillary colums for high-performance liquid chromatography" *J. Chromatogr.* **255**, 381-393.

48. Henderson, R.A., Cox, A. L., Sakaguchi, K., Appella, E., Shabanowitz, J., Hunt, D.F. and Engelhard, V.H. (1993) "Direct identification of an endogenous peptide recognized by multiple HLA-A2.1-specific cytotoxic T cells" *Proc. Natl. Acad. Sci. USA* **90**, 10275-10279.

49. Deterding, L.J., Moseley, M. A., Tomer, K.B. and Jorgenson J.W. (1991) "Nanoscale separations combined with tandem mass spectrometry" *J. Chromatogr.* **554**, 73-82.

50. Chervet, J. P., Ursem, M and Salzmann, J.P. (1996) "Instrumental requirements for nanoscale liquid chromatography" *Anal. Chem.* **68**, 1507-1512.

51. Davis, T.D. and Lee, M. T. (1998) "Rapid protein identification using a microscale electrospray LC/MS system on an ion trap mass spectrometer" *J. Am. Soc. Mass Spectrom.* **9**, 194-201.

52. Bombaugh, K. J. (1975) "A novel method of generating multislope gradients for high-pressure liquid chromatography" *J. Chromatogr.* **107**, 201-206.

53. Doneanu, C. E., Griffin, D.A., Barofsky, E.L., and Barofsky, D.F (2001) "An exponential dilution gradient system for nanoscale liquid chromatography in combination with MALDI or Nano-ESI mass spectrometry for proteolytic digests" *J. Am. Soc. Mass Spectrom* **12**, 1205-1213.

54. Mosbaugh, D. W. and Bennett, S.E. (1994). "Uracil-excision DNA repair", *Progress Nucleic Acid Res.* **48**, 315-370.

55. Wold, M. S., and Kelly, T. (1988) "Purification and characterization of replication protein A, a cellular protein required for *in vitro* replication of simian virus 40 DNA" *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2523-2527.

56. Pfuetzner, R.A., Bochkarev, A., Frappier L. and Edwards, A.M. (1997) "Replication protein A. Characterization and crystallization of the DNA binding domain" *J. Biol. Chem.* **272**, 430-434.

57. Jensen, O. N., Barofsky, D.F., Young, M.C., Von Hippel, P.H., Swenson, S., and Seifried, S (1994) "Mass spectrometric protocol for the analysis of UV-crosslinked protein-nucleic acid complexes" in *Techniques in Protein Chemistry V*, T.W. Crabb, Editor, Academic Press, San Diego, 27-37.

58. Karlsson, K.E., Novotny, M. (1988) "Separation efficiency of slurry-packed liquid chromatography microcolumns with very small inner diameters" *Anal. Chem.* **60**, 1662-1665.

59. Kennedy, R. T. and Jorgenson, J.W. (1989) "Preparation and evaluation of packed capillary liquid lhromatography lolumns with inner diameters from 20 to 50 microns" *Anal. Chem.* **61**, 1128-1135.

60. Kennedy, R.T. and Jorgenson, J.W. (1990) "Efficiency of packed microcolumns compared with large-bore packed columns in size-exclusion chromatography" *J. Microcolumn Sep.* **2**, 120-126.

61. Moseley, M.A., Deterding, L.J., Tomer, K.B. and Jorgenson J.W. (1991) "Nanoscale packed-capillary liquid chromatography coupled with mass spectrometry using a coaxial continuous–flow fast atom bombardment interface" *Anal. Chem.* **63**, 1467-1473.

62. Huczo, E.L., Bodnar, W.M, Benjamin, D., Sakaguchi, K., Zhu, N.Z., Shabanowitz, J., Henderson, R.A., Apella, E., Hunt, D.F. and Engelhard, V.H. (1993) "Characteristics of the endogenous peptides eluted from the class I MHC molecule HLA-B7 detrmined by mass spectrometry and computer modeling" *J. Immun.* **151**, 2572-2587.

63. Bennett, S.E. and Mosbaugh, D.W. (1992) "Characterization of the *Escherichia coli* uracil-DNA glycosylase inhibitor protein complex" *J. Biol. Chem.* **267**, 22512-22521.

64. Shevchenko, A., Wilm, M., Vorm, M. and Mann, M. (1996) "Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels" *Anal. Chem.* **68**, 850-858.

65. Dolan, J. (1989) "Gradient separation of peptides in HPLC" *LC-GC Magazine* **7**, 18-24.

66. Wilm, M. and Mann, M. (1994) "Micro electrospray source for generating highly resolved MS/MS spectra of 1μL sample volume" *Proc. 42nd ASMS Conf. Mass Spectrom. Allied Top., Chicago, IL*, 770.

67. Wilm, M., Neubauer, G. and Mann, M. (1996) "Parent ion scans of unseparated peptide mixtures" *Anal. Chem.* **68**, 527-533.

68. Preisler, J., Foret, F. and Karger B.L. (1998) "On-line MALDI-TOF MS using a continuous vacuum deposition interface" *Anal. Chem.* **70**, 5278-5287.

69. Preisler, J., Ping, H., Rejtar, T. and Karger B.L. (2000) "Capillary electrophoresis – matrix-assisted laser desorption/ionization time of flight mass spectrometry using a vacuum deposition interface" *Anal. Chem.* **72**, 4785-4795.

70. Jonscher, K.R. and Yates, J.R. III (1997) "The quadrupole ion trap mass spectrometer – a small solution to a big challenge" *Anal. Biochem.* **244**, 1-15.

71. Lindahl, T. (1974) "An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues" *Proc. Natl. Acad. Sci.* **71**, 3649-3653.

72. Duncan, B. K., and Chambers, J.A. (1984) "The cloning and overproduction of E. coli uracil-DNA glycosylase" *Gene* **28**, 211-219.

73. Friedberg, E. C., Walker, G.C. and Siede, W (1995) "DNA repair and mutagenesis" ASM Press, Washington, DC.

74. Krokan H.E., Standal, R. and Slupphaug G. (1997) "DNA glycosylases in the base excision repair of DNA" *Biochem. J.* **325**, 1-16.

75. Ravishankar R., Bidya Sagar, M., Roy S., Purnapatre K., Handa P., Varshney U. and Vijayan M. (1998) "X-ray analysis of a complex of *Escherichia coli* uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG" *Nucleic Acids Res.* **26**, 4880-4887.

76. Putnam C.D., Shroyer, M. J., Lundquist A.J., Mol C.D., Arvai A.S., Mosbaugh D.W. and Tainer J.A. (1999) "Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase" *J. Mol. Biol.* **287**, 331-346.

77. Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. and Gilliland, G. L. (1999) "Crystal structure of *Escherichia coli* uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited" *Proteins: Struct.,Funct., Genet.* **35**, 13-20.

78. Dong J., Drohat, A. C., Stivers J.T., Pankiewicz K.W. and Carey P.R. (2000) "Raman spectroscopy of uracil DNA glycosylase-DNA complexes: insights into DNA damage recognition and catalysis" *Biochemistry* **39**, 13241-13250.

79. Drohat A.C., Xiao, G., Tordova M., Jagadeesh J., Pankiewicz K.W., Watanabe K.A., Gilliland G.L. and Stivers J.T. (1999) "Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187" *Biochemistry* **38**, 11876-11886.

80. Slupphaug G., Mol, C. D., Kavli B., Arvai A.S., Krokan H.E. and Tainer J.A. (1996) "A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA" *Nature* **384**, 25-26.

81. Parikh S.S., Mol, C. D., Slupphaug G., Bharati S., Krokan H.E. and Tainer J.A. (1998) "Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA" *EMBO J.* **17**, 5214-5226.

82. Werner R.M., Jiang, Y. L., Gordley R.G., Jagadeesh G.J., Ladner J.E., Xiao G., Tordova M., Gilliland G.L. and Stivers J.T. (2000) "Stressing-out DNA? The contribution of serine-phosphodiester interactions in catalysis by uracil DNA glycosylase" *Biochemistry* **39**, 12585-12594.

83. Lee C., Levin, A. and Branton D. (1987) "Copper staining: a five-minute protein stain for sodium dodecyl sulfate-polyacrylamide gels" *Anal. Biochem.* **166**, 308-312.

84. Shroyer, M.J., Bennett, S.E., Putnam, C.D., Tainer, J.A. and Mosbaugh, D.W. (1999) "Mutation of an active site residue in Escherichia coli uracil-DNA glycosylase: effect on DNA binding, uracil inhibition and catalysis" *Biochemistry* **39**, 4834-4845.

85. Drohat A.C. and Stivers, J. T. (2000) "Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2" *Biochemistry* **39**, 11865-11875.

86. Sayle, R.A. and Milner-White, E.J. (1995) "RASMOL: biomolecular graphics for all" *Trends Biochem. Sci.* **20**, 374-377.

87. Wold, M. S. (1997) "Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism" *Annu. Rev. Biochem.* **66**, 61-92.

88. Wobbe, C.R., Weissbach, L., Borowiec, J.A., Dean, F.B., Murakami, Y., Bullock, P., and Hurwitz, J (1987) "Replication of simian virus 40 origin-containing DNA in vitro with purified proteins" *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1834-1838.

89. Wold, M. S., Li, J.J., and Kelly, T.J. (1987) "Initiation of simian virus 40 DNA replication in vitro: large-tumor-antigen- and origin-dependent unwinding of the template" *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3643-3647.

90. Fairman, M. P. and Stillman, B. (1988) "Cellular factors required for multiple stages of SV40 DNA replication in vitro" *EMBO J.* **7**, 1211-1218.

91. Kolpashchikov D.M., Khodyreva, S. N., Khlimankov D.Y., Wold M.S., Favre A., and Lavrik O.I. (2001) "Polarity of human replication protein A binding to DNA" *Nucleic Acids Res.* **29**, 373-379.

92. Kim, C., Snyder, R.O. and Wold, M.S. (1992) "Binding properties of replication protein A from human and yeast Cells" *Mol. Cell. Biol.* **12**, 3050-3059.

93. Kim, C., and Wold, M.S. (1995) "Recombinant human replication protein A binds to polynucleotides with low cooperativity" *Biochemistry* **34**, 2058-2064.

94. Blackwell, L. J., and Borowiec, J.A. (1996) "Human replication protein A binds single-stranded DNA in two distinct complexes" *Mol. Cell. Biol.* **14**, 3993-4001.

95. Lavrik, O. I., Kolpashchikov, D.M., Weisshart, K., Nasheuer, H.P., Khodyreva, S.N., and Favre, A. (1999) "RPA subunit arrangement near the 3'-end of the primer is modulated by the length of the template strand and cooperative protein interactions" *Nucleic Acids Res.* **27**, 4235-4240.

96. Braun, K. A., Lao, Y., He, Z., Ingles, C.J., and Wold, M.S. (1997) "Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase alpha by multiple mechanisms" *Biochemistry* **36**, 8443-8454.

97. Brill, S. J., and Bastin-Shanower, S.A. (1998) "Identification and characterization of the fourth single-stranded-DNA binding domain of replication protein A" *Mol. Cell. Biol.* **18**, 7225-7234.

98. Lao Y., Lee, C. G., and Wold, M.S. (1999) "Replication protein A interactions with DNA. 2. Characterization of double-stranded DNA-binding/helix-destabilization activities and the role of the zinc-finger domain in DNA interactions" *Biochemistry* **38**, 3974-3984.

99. Bochkarev, A., Pfuetzner, R.A., Edwards, A.M., and Frappier, L. (1997) "Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA" *Nature* **385**, 176-181.

100. Walther, A. P., Gomes X.V., Lao Y., Lee C.G., and Wold, M.S. (1999) "Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit" *Biochemistry* **38,** 3963-3973.

101. Murzin, A. G. (1993) "OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences" *EMBO J.* **12**, 861-867.

102. Bochkareva, E., Belegu, V., Korolev, S. and Bochkarev, A. (2001) "Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding" *EMBO J.* **20**, 612-618.

103. Bochkareva, E., Korolev, S., Lees-Miller, S.P. and Bochkarev, A. (2002) "Structure of the RPA trimerization core and its role in the multistep DNA binding mechanism of RPA" *EMBO J.* **21**, 1855-1863.

104. Bastin-Shanower, S. A. and Brill, S.J. (2001) "Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding" *J. Biol. Chem.* **276**, 36446-36453.

105. Studier, F. W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) "Use of T7 RNA polymerase to direct expression of cloned genes" *Methods Enzymol.* **185**, 60-89.

106. Henricksen, L. A., Umbricht, C.B., and Wold, M.S. (1994) "Recombinant replication protein A: expression, complex formation, and functional characterization" *J. Biol. Chem.* **269**, 11121-11132.

107. Wong, D.L. and Reich, N.O. (2000) "Identification of tyrosine 204 as the photo-cross-linking site in the DNA-EcoRI DNA methyltransferase complex by electrospray ionization mass spectrometry" *Biochemistry* **39**, 15410-15417.

108. Thiede B., Wittmann-Liebold, B., Bienert M. and Krause, E. (1995) "MALDI-MS for C-terminal sequence determination of peptides and proteins degraded by carboxypeptidase Y and P" *FEBS Letters* **357**, 65-69.

109. Patterson D.H., Tarr, G. E., Regnier F.E. and Martin S.A. (1995) "C-terminal ladder sequencing via matrix-assisted laser desorption mass spectrometry coupled with carboxypeptidase Y time-dependent and concentration-dependent digestions" *Anal. Chem.* **67**, 3971-3978.

110. Bonetto V., Bergman, A. C., Jornvall H. and Sillard, R. (1997) "C-terminal sequence analysis of peptides and proteins using carboxypeptidases and mass spectrometry after derivatization of Lys and Cys residues" *Anal. Chem.* **69**, 1315-1319.

111. Stensballe, A., Andersen, S. and Jensen, O.N. (2001) "Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis" *Proteomics* **1**, 207-222.

112. Busch, K.L. (1999) "Mechanisms of MALDI" *Spectroscopy*, **14**, 14-19.

113. Wiley, W.C. and McLaren I.H. (1955) "Time-of-flight mass spectrometer with improved resolution" *Rev. Sci. Instrum.* **26**, 1150-1157.

114. Piyadasa, C.K.G., Hakansson P., Ariyaratne, T.R. and Barofsky, D.F. (1998) "A high resolving power ion selector for post-source decay measurements in a reflecting time-of-flight mass spectrometer" *Rapid Commun. Mass Spectrom.* **12**, 1655-1664.

115. Barofsky, D.F., Hakanson, P., Piyadas, C.K.G. and Katz, D.L. (1998) "Tandem time-of-flight mass spectrometer", U.S. patent application OSU No.98-19; DOJ-OSU-12-98 (63 pages, 15 figures).

Publications related to this Thesis

Doneanu, C.E, Gafken, P.R., Bennett, S.E., and Barofsky, D.F. "Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes: Identification of Amino Acid Residues in the Single-Stranded DNA Binding Domain of Human Replication Protein A", *J. Biol. Chem.,* submitted.

Doneanu, C.E, Gafken, P.R., Bennett, S.E. and Barofsky, D.F. "Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes: Identification of Amino Acid Residues in the DNA Binding Domain of Uracil-DNA Glycosylase from *E. Coli*", *Anal. Chem.,* submitted.

Doneanu, C.E., Griffin, D.A., Barofsky, E.L., and Barofsky, D.F. "An Exponential Dilution Gradient System for Nanoscale Liquid Chromatography in Combination with MALDI or Nano-ESI Mass Spectrometry for Proteolytic Digests", *J. Am. Soc. Mass. Spectrom.*, **2001**, 12, 1205-1213.

Presentations and posters related to this Thesis

Doneanu, C.E., Gafken, P.R. and Barofsky, D.F. "Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes", oral presentation at the 57th Northwest Regional ACS Meeting, **2002**, June 19-22, Spokane, WA.

Doneanu, C.E., Gafken, P.R. and Barofsky, D.F. "Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes", oral presentation at the 56th Northwest Regional ACS Meeting, **2001**, June 14-17, Seattle, WA.

Doneanu, C.E., Gafken, P.R. and Barofsky, D.F. "Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes", poster presented at the 50th ASMS Conference on Mass Spectrometry and Allied Topics, **2002**, June 2-6, Orlando, FL.

Doneanu, C.E., Gafken, P.R. and Barofsky, D.F. "Mass Spectrometric Analysis of UV-Crosslinked Protein-Nucleic Acid Complexes", poster presented at the 49th ASMS Conference on Mass Spectrometry and Allied Topics, **2001**, May 27-31, Chicago, IL.

Doneanu, C.E., Griffin, D.A., Barofsky, E.L., and Barofsky, D.F. "NanoHPLC as a Preparative Technique for Isolating Peptides from Proteolytic Digests for Subsequent Characterization by MALDI-TOF MS and Nanospray MS/MS", poster presented at the 48th ASMS Conference on Mass Spectrometry and Allied Topics, **2000**, June 11-15, Long Beach, CA.