

AN ABSTRACT OF THE THESIS OF

Philip R. Gafken for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on October 2, 2000. Title: Characterization of UV-Crosslinked Protein-Nucleic Acid Interfaces by MALDI MS and ESI MS/MS.

Redacted for Privacy

Abstract Approved: _____

 Douglas F. Barofsky

A procedure was developed for the analysis of UV-catalyzed protein-DNA complexes by mass spectrometry. It used matrix-assisted laser-desorption ionization mass spectrometry (MALDI MS) to tentatively identify protein regions taking part in crosslinking and electrospray ionization tandem mass spectrometry (ESI MS/MS) to definitively identify the regions and specific amino acids within those regions taking part in crosslinking.

The DNA-binding domain of *E. coli* uracil-DNA glycosylase (Ung) was characterized using the developed procedure. The combination of MALDI MS and LC/ESI MS/MS identified the tryptic peptides T6 (58-VVILGQDPYHGP-GQAHGLAFSVR-80), T11 (130-AGQAHSHAS-LGWETFTDK-147), and T18 (185-APHPSPLSAHR-195) as taking part in crosslinking with oligonucleotide dT₂₀. Within T6, LC/ESI MS/MS definitively located His₆₇ as a site of crosslinking. Other sites of crosslinking were assigned to Tyr₆₆, His₁₃₄, and His₁₃₆. Tandem mass data also indicated sites of crosslinking between Ala₁₃₇ and Trp₁₄₁ of T11 and His₁₈₇ and His₁₉₄ of T18. Crystallographic studies by

others of the human enzyme in complex with double-stranded DNA corroborated the mass spectrometric identifying peptides T6 and T18 as making up the DNA-binding region of Ung. The identification of T11, however, has not been identified by other biochemical studies. Preliminary crosslinking studies with human replication protein A demonstrated the general utility of this procedure.

The complexity of Ung nucleopeptide tandem mass spectra prompted an investigation to find means to simplify these spectra. Using protein and peptide models, the addition of a sulfoacetyl group to the N-terminus of peptides by chemical derivatization was found to greatly simplify tandem mass spectra. Singly charged derivatized tryptic peptides fragmented to produce primarily y-series ions whereas nonderivatized tryptic peptides produced mixtures of b- and y-series ions. Derivatization was also found to simplify the spectra of doubly charged tryptic peptides and non-tryptic peptides. These results set the groundwork for future application of chemical derivatization to nucleopeptides.

©Copyright by Philip R. Gafken
October 2, 2000
All Rights Reserved

**Characterization of UV-Crosslinked Protein-Nucleic Acid Interfaces by
MALDI MS and ESI MS/MS**

by

Philip R. Gafken

A THESIS

submitted to

Oregon State University

**in partial fulfillment of
the requirements for the
degree of**

Doctor of Philosophy

**Completed October 2, 2000
Commencement June 2001**

Doctor of Philosophy thesis of Philip R. Gafken presented October 2, 2000

APPROVED:

Redacted for Privacy

Major Professor, ~~representing Biochemistry~~ and Biophysics

Redacted for Privacy

Chair of Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of Graduate School

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

Redacted for Privacy

✓ Philip R. Gafken, Author

TABLE OF CONTENTS

	<u>Page</u>
1. INTRODUCTION.....	1
1.1 Analysis of Protein-DNA Interactions.....	1
1.2 UV-catalyzed Crosslinking of Nucleobases.....	3
1.2.1 Photophysics of Crosslinking.....	3
1.2.1.1 Low Intensity UV-Crosslinking.....	3
1.2.1.2 UV-Laser Crosslinking.....	5
1.2.2 Photochemistry of Crosslinking.....	6
1.2.2.1 Photoreactivity of Nucleobases.....	6
1.2.2.2 Selectivity of Crosslinking.....	6
1.2.2.3 UV-Catalyzed Crosslinking of Nucleobase Analogs...	7
1.3 Characterization of UV-Crosslinked Protein-Nucleic Acid Complexes.....	8
1.3.1 Methodology for Identifying Protein-Nucleic Acid Contacts..	8
1.3.2 Application of Mass Spectrometry to the Analysis of Protein Nucleic Acid Complexes.....	9
1.3.2.1 Fast Atom Bombardment Mass Spectrometry.....	9
1.3.2.2 Matrix-Assisted Laser Desorption-Ionization Mass Spectrometry.....	10
1.3.2.3 Electrospray Ionization Mass Spectrometry.....	14
1.4 Biological Systems Studied.....	17
1.4.1 Uracil-DNA Glycosylase.....	17
1.4.1.1 <i>E. coli</i> Uracil-DNA Glycosylase.....	18
1.4.1.2 Crystal Structures of Human Uracil-DNA Glycosylases.....	19
1.4.1.3 Crystal Structures of <i>E. coli</i> Uracil-DNA Glycosylase.	24
1.4.2 Human Replication Protein A.....	24
1.4.2.1 DNA Binding Domains of Human Replication Protein A.....	25
1.4.2.2 Binding of Human Replication Protein A to Single-Stranded DNA.....	27
1.5 Research Objectives.....	29

TABLE OF CONTENTS (Continued)

	<u>Page</u>
2. MATERIALS AND EXPERIMENTAL PROCEDURES.....	31
2.1 Materials.....	31
2.1.1 Chemicals.....	31
2.1.2 Enzymes.....	32
2.2 General Experimental Procedures.....	32
2.2.1 Chromatographic Resin Preparation.....	32
2.2.1.1 Preparation of Sephadex G-25, Bio-Gel P4 Gel, and Hydroxylapatite Bio-Gel HTP.....	32
2.2.1.2 Preparation of Dowex AG 1-X8 Resin.....	33
2.2.1.3 Preparation of Affi-gel Blue Gel.....	34
2.2.2 Dialysis Tubing.....	34
2.2.3 Preparation of dT ₂₀	34
2.2.4 Preparation of dT ₁₀ , dT ₃₀ , and dT ₇₀	35
2.2.5 5'-End Phosphorylation of Oligonucleotides.....	36
2.2.6 Polyacrylamide Gel Electrophoresis.....	36
2.2.6.1 Large Gel Denaturing Polyacrylamide Gel Electrophoresis.....	36
2.2.6.2 Mini Gel Denaturing Polyacrylamide Gel Electrophoresis.....	37
2.2.6.3 Native Polyacrylamide Gel Electrophoresis	38
2.2.7 Gel Staining.....	38
2.2.7.1 Coomassie Staining.....	38
2.2.7.2 Copper Staining.....	39
2.2.7.3 Silver Staining.....	39
2.2.8 In-gel Reduction and Alkylation.....	40
2.2.9 In-gel Proteolytic Digestion.....	40
2.2.10 Nuclease P1 Digestion.....	41
2.2.11 Off-line High Performance Liquid Chromatography.....	41
2.2.12 Matrix-Assisted Laser-Desorption Ionization Mass Spectrometry.....	42
2.2.12.1 MALDI Analysis of Proteins.....	42
2.2.12.2 MALDI Analysis of Peptides.....	42
2.2.12.3 MALDI Analysis of dT ₂₀ and Nucleopeptides.....	43
2.2.13 Electrospray Ionization Mass Spectrometry.....	43
2.2.13.1 Nanospray Triple Quadrupole Mass Analysis.....	43
2.2.13.2 LC Triple Quadrupole Mass Analysis.....	44

TABLE OF CONTENTS (Continued)

	<u>Page</u>
2.2.13.3 Quadrupole Ion-trap Mass Analysis.....	45
2.2.14 Database Search Algorithms.....	45
2.3. Experimental Procedures for Ung Project.....	46
2.3.1 Uracil-DNA Glycosylase Activity Assay.....	46
2.3.2 Purification of Uracil-DNA Glycosylase.....	46
2.3.3 Preparative Crosslinking of Ung to dT ₂₀	49
2.3.4 Purification of Ung Nucleopeptides.....	49
2.3.4.1 "In-pot" Method.....	49
2.3.4.2 Gel-based Method.....	50
2.4 Experimental Procedures for Sulfonic Acid Derivatization.....	52
2.5 Experimental Procedures for hRPA Project.....	53
2.5.1 Overexpression of hRPA.....	53
2.5.1.1 Mini Cultures.....	53
2.5.1.2 Preparative Cultures.....	54
2.5.2 Purification of hRPA.....	54
2.5.3 Protein Concentration Determinations.....	55
2.5.4 Electrophoretic Mobility Shift Assay.....	56
2.5.5 UV-Catalyzed Crosslinking of hRPA.....	56
 3. PURIFICATION OF NUCLEOPEPTIDES FROM THE UV-CROSSLINKED UNG × dT ₂₀ NUCLEOPROTEIN COMPLEX FOR CHARACTERIZATION BY MALDI MS	 57
3.1 Results.....	58
3.1.1 "In-pot" Purification of Ung Nucleopeptides.....	58
3.1.1.1 Crosslinking of Ung to dT ₂₀	58
3.1.1.2 Isolation and MALDI Analysis of Ung Nucleopeptides from the Ung × dT ₂₀ Nucleoprotein Complex.....	62
3.1.2 Gel-Based Purification of Ung Nucleopeptides.....	66
3.1.2.1 Crosslinking of Ung to dT ₂₀	66
3.1.2.2 Isolation of Ung Nucleopeptides from the Ung × dT ₂₀ Nucleoprotein Complex.....	72

TABLE OF CONTENTS (Continued)

	<u>Page</u>
3.1.2.3 Tentative Identification of Peptide × dT ₂₀ Complexes....	72
3.2 Discussion.....	79
 4. AMINO ACID SEQUENCING OF PURIFIED UNG NUCLEOPEPTIDES USING ELECTROSPRAY IONIZATION MASS SPECTROMETRY.....	 88
4.1 Results.....	88
4.1.1 Analytical Outline for Mass Spectrometric Analysis of Ung Nucleopeptides.....	88
4.1.2 MALDI MS Characterization of Nucleolytically Digested Ung Nucleopeptides.....	88
4.1.3 MALDI MS Characterization of HPLC-Purified Ung × dT _n Nucleopeptides.....	92
4.1.4 Nanospray/ESI MS/MS Amino Acid Sequencing of HPLC Purified Ung-peptide × dT _n Nucleopeptides.....	97
4.1.4.1 Nanospray/ESI MS/MS of T18 × dT ₄	97
4.1.4.2 Nanospray/ESI MS/MS of T6 ₍₋₇₎ × dT ₂ and T6 ₍₋₅₎ × dT ₂ ...	100
4.1.5 LC MS Analysis of Ung-peptide × dT _n Complexes.....	106
4.1.6 LC MS/MS Peptide Sequencing of Ung-peptide × dT _n Complexes.....	112
4.1.6.1 LC MS/MS of T18 × dT ₂	112
4.1.6.2 LC MS/MS of T6 ₍₋₇₎ × dT ₂	115
4.1.6.3 LC MS/MS of T11 × dT ₂	119
4.1.6.4 LC MS/MS of T6 ₍₋₃₎ × dT ₂	119
4.2 Discussion.....	126
 5. SULFONIC ACID DERIVATIZATION OF PEPTIDES FOR ENHANCED SEQUENCING BY ELECTROSPRAY IONIZATION MASS SPECTROMETRY.....	 144
5.1 Results.....	145

TABLE OF CONTENTS (Continued)

	<u>Page</u>
5.1.1 The Derivatization Reaction.....	145
5.1.2 Tandem Mass Spectra of Singly Charged, Derivatized, Tryptic Peptides.....	147
5.1.3 Tandem Mass Spectra of Doubly Charged, Derivatized, Tryptic Peptides.....	152
5.1.4 Tandem Mass Spectra of Derivatized, Non-tryptic Peptides...	154
5.2 Discussion.....	154
 6. UV-CATALYZED CROSSLINKING OF HUMAN REPLICATION PROTEIN A TO SINGLE-STRANDED DNA SUBSTRATES.....	 160
6.1 Results.....	161
6.1.1 Overproduction of Recombinant hRPA.....	161
6.1.2 Purification of hRPA.....	161
6.1.3 Single-Stranded DNA Binding Capability of hRPA.....	173
6.1.4 UV-Catalyzed Crosslinking of hRPA to dT ₃₀	175
6.1.5 UV-Dose Response of Crosslinking hRPA to dT ₁₀ , dT ₃₀ , dT ₇₀ ..	178
6.1.6 Peptide Mapping of hRPA70.....	181
6.1.7 MALDI MS Analysis of hRPA70 × dT ₃₀ Nucleopeptides.....	181
6.1.8 Tandem Mass Spectrometry of Nuclease P1-digested hRPA70 × dT ₃₀ Nucleopeptides.....	189
6.2 Discussion.....	190
 BIBLIOGRAPHY.....	 194
 APPENDIX.....	 209

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
3.1 Effect of 254 nm irradiation time on the crosslinking of Ung to dT ₂₀ ..	59
3.2 MALDI mass spectra of Ung and dT ₂₀	60
3.3 "In-pot" purification scheme for isolating Ung-nucleopeptides from Ung × dT ₂₀ nucleoprotein complexes.....	63
3.4 Anion exchange chromatography elution profile of Ung tryptic nucleopeptides.....	64
3.5 MALDI mass spectrum of purified Ung × dT ₂₀ tryptic peptides isolated using the "in-pot" purification scheme.....	65
3.6 Effect of 254 nm irradiation time on the crosslinking of Ung to [³² P]dT ₂₀	67
3.7 Effect of 254 nm light on MALDI MS analysis of Ung and dT ₂₀	70
3.8 Effect of 312 nm light on MALDI MS analysis of Ung and dT ₂₀	71
3.9 In-gel purification scheme for isolating nucleopeptides from nucleoprotein complexes.....	73
3.10 MALDI tryptic map of <i>E. coli</i> Ung.....	74
3.11 MALDI mass spectrum of purified Ung × dT ₂₀ tryptic peptides.....	78
3.12 <i>E. coli</i> uracil-DNA glycosylase peptides tentatively identified by MALDI MS as being crosslinked to dT ₂₀	80
3.13 Correlation of Edman sequencing and MALDI MS data with the crystal structures of uracil-DNA glycosylase.....	85
3.14 Correlation of MALDI MS data with the crystal structures of uracil-DNA glycosylase.....	86
4.1 Analysis scheme for the identification of covalently attached amino acids by mass spectrometry.....	90

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
4.2 MALDI mass spectrum of purified tryptic Ung × dT ₂₀ peptides after digestion with nuclease P1.....	91
4.3 HPLC chromatogram of nuclease P1-digested nucleopeptides.....	95
4.4 MALDI MS spectrum of HPLC fraction A.....	96
4.5 MALDI MS spectrum of HPLC fraction B.....	98
4.6 Nanospray/ESI MS spectrum of HPLC elution fraction A.....	99
4.7 Nanospray/ESI MS/MS spectrum of triply charged T18 × dT ₄	101
4.8 Nanospray/ESI MS spectrum of HPLC elution fraction B.....	103
4.9 Nanospray/ESI MS/MS spectrum of triple charged T6 ₍₋₇₎ × dT ₂	104
4.10 Nanospray/ESI MS/MS spectrum of triple charged T6 ₍₋₅₎ × dT ₂	107
4.11 LC/ESI MS base peak ion chromatogram of Ung-peptide × dT _n complexes.....	109
4.12 LC/ESI MS spectra of Ung-peptide × dT _n complexes.....	110
4.13 LC/ESI MS/MS spectrum of T18 × dT ₂	113
4.14 LC/ESI MS/MS spectrum of triply charged T6 ₍₋₇₎ × dT ₂	116
4.15 LC/ESI MS/MS spectrum of doubly charged T6 ₍₋₇₎ × dT ₂	118
4.16 LC/ESI MS/MS spectrum of T11 × dT ₂	120
4.17 LC/ESI MS/MS spectrum of T6 ₍₋₃₎ × dT ₂	122
4.18 LC/ESI MS/MS spectrum of doubly charged T6 ₍₋₃₎ × dT ₂	125
4.19 Summary of fragmentation patterns observed in nanospray/ESI MS/MS spectra of Ung nucleopeptides.....	130

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
4.20 Summary of fragmentation patterns observed in LC/ESI MS/MS analysis of Ung nucleopeptides.....	134
4.21 Correlation of LC/ESI MS/MS nucleopeptide mapping with the crystal structures of uracil-DNA glycosylase.....	141
5.1 N-terminal peptide derivatization scheme.....	146
5.2 Effect of derivatization on the CID spectra of singly charged peptide LFSLELVDESGEIR.....	148
5.3 Effect of derivatization on the CID spectra of singly charged peptide QGVLLNNTVLTVR.....	150
5.4 Effect of derivatization on tryptic peptide containing C-terminal lysine.....	151
5.5 Effect of derivatization on the CID spectrum of the doubly charged peptide ELENTIPGFTRPNHGYLESWAR.....	153
5.6 Effect of derivatization on the CID spectrum of doubly charged angiotensin.....	155
5.7 Proposed ESI CID fragmentation scheme for tryptic peptides N-terminally derivatized with sulfonic acid.....	157
6.1 Overexpression of hRPA.....	162
6.2 Purification scheme for hRPA.....	164
6.3 Affigel Blue purification of hRPA.....	165
6.4 Single-stranded DNA agarose column purification of hRPA.....	168
6.5 Mass spectrometric analysis of purified hRPA.....	170
6.6 Database search results for identification of ~14 kDa purified protein.....	172

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
6.7 Electrophoretic mobility shift analysis to measure the capability of hRPA to bind DNA.....	174
6.8 Dependence of UV light for the formation of UV-catalyzed crosslinks between hRPA and dT ₃₀	176
6.9 Dose response of crosslinking hRPA to dT ₁₀ , dT ₃₀ , and dT ₇₀	179
6.10 Quantitation of UV dose response during crosslinking of hRPA to dT ₁₀ , dT ₃₀ , and dT ₇₀	182
6.11 Analysis of a preparative scale crosslinking reaction of hRPA to dT ₃₀	186
6.12 MALDI mass spectrum of purified hRPA70 × dT ₃₀ tryptic nucleopeptides.....	187

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1 Tentative identification of Ung tryptic nucleopeptides measured by MALDI MS.....	76
4.1 Summary of tentatively identified Ung × dT ₂₀ tryptic peptides.....	89
4.2 Analysis of nuclease P1 digestion products.....	93
6.1 Tryptic peptide mapping of carboxyamidomethylated human RAP70 by LC/ESI MS.....	183

LIST OF APPENDIX FIGURES

<u>Figure</u>	<u>Page</u>
A.1 Nomenclature for common peptide backbone fragments produced from low energy collisionally induced dissociation.....	210
A.2 Nomenclature for common oligonucleotide backbone fragments produced from low energy collisionally induced dissociation.....	211

Characterization of UV-Crosslinked Protein-Nucleic Acid Interfaces by MALDI MS and ESI MS/MS

1. INTRODUCTION

1.1 Analysis of Protein-DNA Interactions

The study of protein-nucleic interactions is a major focus of biochemical research because they are at the heart of such basic cellular processes as transcription, replication, recombination, and DNA repair. Knowledge of the structures of protein-nucleic acid complexes is important to understanding how these biological processes occur. Numerous methods are available to study protein-nucleic acid interactions and they can generally be categorized as either indirect or direct methods. Gel mobility shift assays, filter binding, and footprinting are considered indirect methods because they reveal the character and level of interaction between proteins and nucleic acids but do not expose any structural features of the interactions. X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are considered direct methods for determining the topology of protein-nucleic acid interactions because they provide three-dimensional structural data at the atomic level. Obtaining high-resolution structures of protein-nucleic acid complexes by these physical techniques, however, can be problematic. In particular, large amounts of homogeneous sample are required to produce high quality crystals for X-ray studies and the inherent size of many protein-nucleic acid complexes limits the amount of information that can be obtained in NMR studies (71, 115).

Another method important for studying protein-nucleic acid interactions involves the use of crosslinking agents. These methods are used primarily to gain information about amino acids and nucleobases that nearly touch at a protein-nucleic acid interface (88). A major drawback associated with chemical crosslinking reagents is that they have finite lengths; thus, regions within a protein-nucleic acid complex that are not naturally in contact with one another may become crosslinked via the length of the reagent. Another drawback associated with chemical crosslinking is the rate of crosslinking. For bifunctional crosslinking reagents (two reactive groups contained within a reagent separated by a spacer length), initial crosslinking to the nucleobase (or amino acid) must be immediately followed by crosslinking with the amino acid partner (or nucleobase). If the second reaction is slow, the protein-nucleic acid complex may take up a new, more stable conformation that brings nucleobases and amino acids not originally in contact (non-native contacts) into close proximity and in turn form a non-natural crosslink. Because of these problems, experiments using crosslinking agents must be carefully designed and the information obtained must be interpreted with caution. These problems can be avoided by simultaneously forming a crosslink between a protein and a nucleic acid without an intervening species, such as a chemical crosslinking agent. This sort of crosslink is considered a "zero-length" bond and its formation can be catalyzed by ultraviolet light.

1.2 UV-Catalyzed Crosslinking of Nucleobases

Two reports in the early 1960s established that the absorption of ultraviolet (UV) light by living cells could induce crosslinking of protein to DNA (3, 129). Along with studying the deleterious effects of UV light on living cells, studies over the past four decades have examined the feasibility of using UV-crosslinking as a direct means of determining contact points between nucleic acids and proteins. Today, photochemical crosslinking has become a powerful biochemical procedure for characterizing protein-nucleic acid interfaces.

1.2.1 Photophysics of Crosslinking

The optimum absorption of nucleobases is around 260 nm, whereas the optimum absorption for proteins is around 280 nm. The maximum absorption of proteins is determined by their content of aromatic amino acids, which are typically present in proteins at an abundance of 3-4% (136). This means that UV irradiation of a nucleoprotein complex in the range of 250 to 270 nm will be absorbed mainly by the nucleobases.

1.2.1.1 Low Intensity UV-Crosslinking

When a nucleobase absorbs a single photon, under typical low intensity light (less than 10^{18} photons/cm²•sec at 260 nm), it is promoted to the first excited singlet state (S_1) (19). From S_1 , which has a lifetime of about 10 ps, the excited nucleobase can either react, relax to the ground state, or undergo

intersystem crossing to the first excited triplet state (T_1). T_1 , which has a lifetime of about 1 μ s, can also react or relax to the ground state. The crosslinking of a nucleobase to an amino acid can occur from either the S_1 or the T_1 states.

Because the S_1 state has a very short lifetime, crosslinking from this state must take place between groups in contact at the time of excitation. Even though the T_1 state has a longer lifetime than the S_1 state, it is still relatively short compared to the time scale for macromolecular motion, which is on the order of milliseconds. Therefore, crosslinking via the T_1 state most likely involves groups in close proximity at the time of excitation (19).

One drawback to conducting crosslinking experiments using low intensity UV light sources is low sample yield. In order to obtain sufficient sample, crosslinking times ranging from minutes, to hours, or even days, are often used. Use of continuous UV irradiation permits side reactions to take place within the protein and the DNA. These side reactions can result in modifications to amino acids (e.g., destruction of tryptophan), reactions between bases (e.g., formation of cyclobutane dimers), and modifications to the DNA (e.g. strand breaks) (19, 51). As irradiation proceeds, modified, uncrosslinked proteins and nucleic acid residues accumulate. These modified macromolecules can take up new, non-native conformations that can potentially crosslink. Since native complexes crosslinked at the start of irradiation cannot be distinguished from those crosslinked after disruption to the native structures, the crosslinking conditions must be optimized in order to maximize crosslinking yield and minimize side reactions.

1.2.1.2 UV-Laser Crosslinking

During irradiation with low intensity light, only the levels S_1 and T_1 of the nucleobases are populated. As the intensity of light increases the probability that a second photon will be absorbed also increases. The second photon can then excite a molecule already in the S_1 or T_1 level into higher excited levels (termed H_S and H_T). This results in a population of molecules at higher excited levels, the lifetimes of which do not exceed 25 ns (19). Photochemical reactions via higher excited states are more efficient than via lower excited states. When lasers are used as the high intensity light source, the efficiency of this process (the percentage of proteins crosslinked per absorbed photon) has been estimated to be 20-100 times higher than those originating from lower energy levels (20). Like the crosslinks formed from low intensity UV light, the crosslinks produced from laser crosslinking are also “zero-length” bonds.

Lasers can be used to produce a single, high intensity, short pulse (nanoseconds) of monochromatic UV light to catalyze protein-nucleic acid crosslinking. A single laser pulse used in crosslinking is on a timescale that is too short to allow any large-scale motion within the macromolecules (19, 20, 51, 52). Thus, the interactions between the protein and the nucleic acid are frozen in a native complex. This is the primary advantage of using lasers in place of low intensity UV irradiation. The increased intensity from lasers, however, can result in the absorption of two photons by the nucleic acid base, in turn, exceeding the ionization potential of the molecule. This produces longer-lived ions or radicals

that can react non-specifically with the protein or lead to strand breaks within DNA (19). Another drawback to the use of laser irradiation is the photodegradation of proteins (51, 73). Similar to low-intensity UV irradiation, laser crosslinking conditions must be optimized for a given system.

1.2.2 Photochemistry of Crosslinking

1.2.2.1 Photoreactivity of Nucleobases

Radiation sources with outputs in the range of nucleobase absorption maxima could potentially catalyze the crosslinking of any nucleobase within DNA or RNA. A study of bacteriophage T4 gene 32 single-stranded DNA binding protein (gp32) showed that thymine was the most reactive base (51). At a wavelength of 252.7 nm, it was found that the crosslinking efficiency followed the order: dT (1) >> dC (0.0084) > rU (0.0041) > rC, dA, dG (0.001).

1.2.2.2 Selectivity of Crosslinking

After excitation of the nucleobases with UV light, crosslinking reactions within a macromolecular complex can take place with numerous amino acids. Taking into account four common nucleobases and twenty common amino acids, the number of nucleobase-amino acid combinations that could potentially be formed during irradiation is large. Therefore, it is of great interest to know which, if any, of these combinations has the potential for crosslinking. Shetlar and colleagues performed experiments with model systems to test the reactivity of

the nucleobases to amino acids. They showed that all 20 common amino acids have the potential to crosslink to DNA, poly A, poly T, poly G, poly C and poly U, with cysteine, lysine, phenylalanine, tryptophan, and tyrosine being the most reactive. Histidine, glutamic acid, and aspartic acid were moderately reactive, and arginine, leucine, and isoleucine were the least reactive (122-124). Other studies have been conducted with individual amino acids and individual nucleobases. Specifically, crosslinking reactions have produced conjugates of uridine-cysteine (56), thymine-lysine (113), 5-methylcytosine-serine and 5-methylcytosine-threonine (121), and thymidine-tyrosine (120). Based on the results from these model studies, photochemical mechanisms for the formation of each of these conjugates were proposed. It is likely that numerous mechanisms exist between the potential nucleobase-amino acid crosslinking possibilities, the pathways of which may be somewhat different in macromolecular systems.

1.2.2.3 UV-Catalyzed Crosslinking of Nucleobase Analogs

An alternative to using common nucleobases as substrates for crosslinking experiments is the use of photosensitive nucleobase analogs in their place (for review, see (85)). Photosensitive analogs include 5-bromouracil, 5-iodouracil, 4-thiouracil, 5-iodocytosine, and azido derivatized nucleobases. The use of these analogs may increase the crosslinking efficiency. This permits the study of protein-nucleic acid complexes that would not normally be detected due to low reaction yields. An additional benefit is that the photon flux required to obtain crosslinking is greatly reduced when using these analogs, thus minimizing

unwanted side reactions. Typically, these analogs are irradiated at wavelengths greater than 300 nm, far beyond the absorption ranges of nucleic acids and proteins, which decreases the photodegradation. It should be noted that photosensitive nucleobases are unnatural nucleobases and as a result the possibility for structural perturbations to the nucleoprotein complex is present.

1.3 Characterization of UV-Crosslinked Protein-Nucleic Acid Complexes

1.3.1 Methodology for Identifying Protein-Nucleic Acid Contacts

For the identification of amino acids that take part in crosslinking, a general methodology has typically been used (146). It consists of first irradiating the protein-nucleic acid mixture to form a nucleoprotein complex. The complex is enzymatically broken apart into free peptides and peptides crosslinked to the nucleic acid substrate (termed nucleopeptides). Finally, the nucleopeptides are isolated and sequenced. Numerous biochemical techniques have been employed to isolate nucleopeptides. Characterization of these complexes, however, has been primarily conducted by Edman gas-phase sequencing. To identify the site of crosslinking by Edman sequencing, a "hole" is observed in a sequencing cycle. With prior knowledge of the primary structure of the DNA-binding protein, and knowledge of the peptide region of the protein being sequenced, the site of crosslinking can be inferred by the position of the hole in the peptide sequence (146). Edman sequencing has proven quite successful in this regard; however, there are drawbacks to this technique. Isolated

nucleopeptides containing multiple species must be further separated into individual species prior to analysis. This means that crosslinked species in low abundance could be lost during extensive sample purification. Also, the harsh conditions of gas-phase sequencing may break some labile amino acid-nucleic acid crosslinks. Thus, it would be advantageous to have direct, sensitive, non-destructive techniques available that could speed up the identification of crosslinked amino acids in UV-irradiated complexes. One technique that holds such promise is mass spectrometry.

1.3.2 Application of Mass Spectrometry to the Analysis of Protein-Nucleic Acid Complexes

Mass spectrometry has emerged as an important tool for the study of a wide variety of macromolecular structures. For a better part of the last two decades, the molecular weight determinations made by mass spectrometers have been used in numerous studies to characterize the structures of proteins, peptides, oligonucleotides, lipids, and carbohydrates. More recently, mass spectrometry has been applied to the analysis of protein-nucleic acid and peptide nucleic acid complexes.

1.3.2.1 Fast Atom Bombardment Mass Spectrometry

The first applications of mass spectrometry to characterize UV-crosslinked protein-nucleic acid interactions used fast atom bombardment (FAB) mass spectrometers. In one study, FAB MS was able to determine the mass of a

modified peptide produced from a UV-catalyzed crosslinking reaction of glyceraldehyde-3-phosphate dehydrogenase and the nucleic acid analog arylazido- β -alanyl-NAD⁺ (24). In a later study, FAB tandem mass spectrometry (MS/MS) was used to sequence the peptide-nucleic acid complex of angiotensin I to identify the site of crosslinking to thymidine (120). Even though these studies required large amounts of sample (greater than 50 pmol), they demonstrated the potential of mass spectrometry for analyzing peptides crosslinked to nucleic acids. Since these initial studies on peptide-nucleic acid complexes by FAB MS, major advances have taken place in the development of soft ionization techniques for mass spectrometry. Specifically, these advances have been in MALDI MS and ESI MS.

1.3.2.2 Matrix-Assisted Laser Desorption-Ionization Mass Spectrometry

Since the first descriptions of matrix-assisted laser desorption-ionization (MALDI) mass spectrometry for the analysis of macromolecules (49, 50, 60, 61), MALDI MS has become a widespread analytical technique for proteins and peptides (34, 63, 130), as well as oligonucleotides (35, 44).

Commercially available instruments that are combined with time-of-flight analyzers (29) can determine molecular weights to over 100 kDa from sample amounts on the order of 0.1 to 10 pmol.

Barofsky and co-workers first demonstrated the feasibility of using mass analysis for protein-nucleic acid complexes (57). MALDI MS was used to monitor crude UV-catalyzed crosslinking reactions between gp32 and oligonucleotide

dT₂₀, and reactions between *E. coli* termination factor rho (Rho) and the nucleic acid analog 4-thio-uridine-diphosphate (4-S-UDP). Mass analysis showed that both gp32 and rho crosslinked with their nucleic acid substrates at a ratio of 1:1. This work also presented a systematic study to identify appropriate matrix and solvent conditions for MALDI analysis of these complexes. In subsequent research performed by Barofsky and co-workers, MALDI MS was used in identifying the *E. coli* uracil-DNA glycosylase (Ung) DNA binding domain (6). Mass analysis of purified Ung:dT₂₀ complexes established the stoichiometry of crosslinking to be 1:1. To date, these are the only demonstrations of MALDI MS in the analysis of protein-nucleic acid complexes.

MALDI MS has also been used in characterizing peptide-nucleic acid complexes produced from enzymatic digestions of protein-nucleic acid complexes. In this role, MALDI serves as a technique to tentatively identify the peptide portion of a complex. Subtracting the mass of the nucleic acid from the complex and then comparing the resulting mass of the peptide to a peptide mass map of the protein identifies the peptide region taking part in crosslinking. By using this procedure, the DNA-binding domains of Ung (6) and rat DNA polymerase β (rat pol β) (28) have been defined. In the case of Ung, four peptides were identified as being crosslinked to dT₂₀ by using a combination of MALDI MS and gas-phase sequencing. The region that the peptides spanned within the protein (amino acids 57-79 and 184-212) were proposed to make up the DNA-binding domain (6). This was the first time MALDI had ever been used to provide structural information on a protein-nucleic acid complex. Subsequent

X-ray crystal structures of human uracil-DNA glycosylase in complex with double-stranded DNA (96, 128) revealed that the regions of the human enzyme predicted to be in contact with DNA were homologous to those regions of the *E. coli* enzyme identified by UV-crosslinking. The combination of MALDI MS and gas-phase sequencing was also used for rat DNA polymerase β (pol β) (28). In this study, nanosecond laser pulses were used as the radiation source and MALDI was used to tentatively identify six nucleopeptides involved in crosslinking. The regions of the protein identified as making up the DNA-binding domain (amino acids 28-40, 55-60, and 62-81) corroborated NMR data (82) and low intensity UV-crosslinking data obtained for the same protein-nucleic acid complex (103).

Other studies have used MALDI MS in both tentative and definitive identification of peptide-nucleic acid complexes. In the following two studies, photoaffinity-labeled nucleic acid substrates were used. Recently, Traub and colleagues crosslinked the protein vimentin to oligo (dG•BrdU)₁₂dG-fluorescein isothiocyanate (143). The protein-nucleic acid complex was proteolytically digested and subjected to several purification steps resulting in three individually isolated peptide-nucleic acid complexes. After tentatively identifying the peptide portion of the complexes, they were nucleolytically digested individually and subjected to carboxypeptidase Y digestion as well as manual Edman degradation. The resulting products were analyzed by MALDI MS, which produced sequencing ladders that identified the sites of crosslinking to Tyr₂₉, Tyr₃₇, and Tyr₅₂ within the protein (143). Even though this methodology was

unique in using MALDI MS to find sites of crosslinking, large amounts of purified peptide-nucleic acid analyte was needed to conduct multiple cycles of chemical and enzymatic sequencing. A peptide-nucleic acid complex from *E. coli* Rec A has also been identified using a MALDI ion-trap instrument. The purpose of this study was to demonstrate the capability of this specialized instrument (107). Nonetheless, analysis of the complex in the MS¹ mode was used for tentative identification, and analysis in the MSⁿ mode was used for definitive identification. After three stages of tandem mass spectrometry, Met₁₆₅ was identified as a site of crosslinking, which had been previously identified by Edman sequencing (144).

The applications of MALDI MS to the analysis of protein-nucleic acid complexes described thus far have all focused on characterizing the interactions associated with the protein. It is worth mentioning one additional application in which specific sites of crosslinking were identified in both the protein and the nucleic acid portion of the crosslinked complexes. Urlaub and co-workers employed MALDI MS to determine sites of contact between the rRNA and individual ribosomal proteins of the 30 S ribosomal subunits from *E. coli* (135). Specifically, they employed N-terminal sequence analysis to determine the modified amino acids within chemically crosslinked peptide-oligoribonucleotide complexes. MALDI MS in conjunction with partial alkaline hydrolyses identified the corresponding sites on the oligoribonucleotides.

The aforementioned applications of MALDI MS to characterizing protein/peptide-nucleic acid complexes clearly demonstrates the value of this

method in determining sites of protein-DNA contact. In summary, MALDI MS was successfully used in measuring crosslinking stoichiometry, tentatively identifying peptide-nucleic acid species, and definitively identifying peptide-nucleic acid species when combined with enzymatic and chemical sequencing. This last application, however, requires large sample quantities and digestion procedures require optimization for each complex studied. It seems that MALDI MS is best suited for measuring crosslinking stoichiometry and tentatively identifying peptide-nucleic acid complexes. Tentative identification of these complexes is important since it can be used to direct subsequent gas-phase sequencing or tandem mass spectrometry sequencing efforts.

1.3.2.3 Electrospray Ionization Mass Spectrometry

Similar to MALDI MS analysis, electrospray ionization (ESI) mass spectrometry has been used to characterize both protein-nucleic acid and peptide-nucleic acid complexes. ESI-MS has shown utility in studying noncovalent interactions of protein-nucleic acid complexes due to the ability of ESI to generate molecular ions of intact complexes while producing little or no fragmentation (141). Stoichiometry of the binding partners can be deduced by accurate mass analysis of the ligands and the intact complex. Complexes between the DNA binding domain of human PU.1 and sequence specific double-stranded DNA (26), bacteriophage f1 gene V protein with single-stranded DNA (25), bacteriophage T4 RegA protein with RNAs of various sizes (81), and human XPA protein's minimal binding domain with cisplatin-adducted DNA (154) were

studied by Smith and colleagues. The complex between the HIV-1 Tat protein and RNA was studied by Loo and co-workers (114). ESI has also been used to measure the stoichiometry of crosslinking. Reich and colleagues measured the stoichiometry for the complex of *EcoRI* DNA methyltransferase crosslinked to 5-iodouracil substituted DNA (152).

ESI sources are commonly coupled to triple quadrupole (87) or ion-trap (11) mass analyzers that allow for multiple stages of mass spectrometry. Therefore, these instruments are heavily used for the sequencing of peptides. Barofsky and colleagues proposed a strategy that uses MALDI MS to monitor protein-nucleic acid crosslinking reactions and ESI MS/MS sequencing of peptide-nucleic acids species produced from proteolytic digestion of the protein-nucleic acid complex to identify amino acids that are crosslinked (58). The same laboratory used ESI MS/MS to characterize two synthetic peptide-nucleic acid complexes. They were prepared by joining an eleven-residue synthetic peptide containing one internal carboxyl group (Asp side chain) to amino-linker-5'pdT₆ and amino-linker-5'pdT₁₀ to demonstrate the utility of sequencing peptide-nucleic acid complexes by ESI MS/MS. The results of this study showed that a positive ESI tandem mass spectrum could sequence the peptide portion of the peptide-dT₆ complex to locate the point of the oligonucleotide's covalent attachment. This study also showed that as the size of the oligonucleotide portion of the complex decreased from dT₁₀ to dT₆, the ionization efficiency of the complex drastically increased in the positive ion mode. These results served as

guidelines for the future analysis of peptide nucleic acid complexes by positive mode ESI mass spectrometry.

Since this model study, ESI MS/MS has been successful in identifying sites of crosslinking within two DNA-binding proteins. One of the proteins studied was the 155-amino acid variant of basic fibroblast growth factor (38). This protein was UV-crosslinked to a 61-mer oligonucleotide containing seven bromodeoxyuridines. After enzymatic digestion and isolation of a single crosslinked peptide, digestion by snake venom phosphodiesterase and alkaline phosphatase produced a peptide crosslinked to a dinucleotide. ESI MS/MS revealed Tyr₁₃₃ as the site of crosslinking. The other DNA-binding protein studied was *E. coli* endonuclease VIII. This protein was chemically crosslinked to thymine glycol-containing double-stranded DNA in order to identify amino acids taking part in the catalytic function of the enzyme (110). After proteolytic digestion, a single crosslinked peptide was isolated (peptide covalently bound to the thymine glycol-containing oligonucleotide) and then sequenced by ESI MS/MS in the negative ion mode. This sequenced the oligonucleotide portion of the DNA locating the site of crosslinking to the thymine glycol-containing nucleotide. Analysis in the negative ion mode did not provide sufficient information to identify the amino acid taking part in crosslinking. To accomplish this, the complex was nucleolytically digested, and ESI MS/MS (negative ion) sequenced the resulting product to identify the N-terminal proline of the protein as the site of crosslinking. The results from these two proteins demonstrate the

utility of ESI tandem mass spectrometry in identifying amino acids and nucleotides that take part in crosslinking.

1.4 Biological Systems Studied

The research conducted for this thesis focused on elucidating the DNA-binding domains of two proteins—*E. coli* uracil-DNA glycosylase and human replication protein A. The following sections provide biological background on both of these proteins, with heavy emphasis being placed on the structural features that contribute to their abilities to bind DNA.

1.4.1 Uracil-DNA Glycosylase

Uracil is continually introduced into the genome of numerous biological systems, including bacteria, bacteriophages, yeast, mammalian cells, and viruses (93). The buildup of uracil in DNA primarily takes place by two processes. One is the incorporation of dUMP during DNA synthesis. DNA polymerases of both prokaryotes and eukaryotes do not readily distinguish between dUTP and dTTP during synthesis; thus, the intracellular pool size of dUTP at the time of synthesis dictates the frequency of uracil incorporation. The resulting accumulation of U•A base pairs in the genome has the potential to alter protein-DNA interactions (93). The other process that results in the buildup of uracil in DNA is cytosine deamination. Cytosine residues can undergo deamination by spontaneous or by hydrolytic, chemical, or UV-radiation induced

routes. These result in a G•U mispair from a previous G•C base pair that, if left unrepaired, can lead to a G•C to A•T transition mutation. Cytosine deamination is a major source of spontaneous mutations in both prokaryotes and eukaryotes (78). To counteract the effects of uracil buildup in DNA, cells contain uracil-DNA glycosylases that recognize and remove uracil from DNA (72).

Uracil-DNA glycosylase specifically recognizes uracil in DNA and hydrolyzes the N-C1' glycosylic bond between uracil and the deoxyribose sugar, beginning uracil-initiated base excision repair (72). This creates an abasic site that is removed by an apurinic/apyrimidinic (AP) endonuclease and a deoxyribophosphodiesterase, leaving a gap that is filled by DNA polymerase, and closed by DNA ligase. The base excision repair pathway is a major cellular defense mechanism against spontaneous DNA damage (79). Its importance is strongly indicated by the facts that the occurrence of uracil-DNA glycosylase is widespread, the protein is highly conserved between biological species, and the base excision repair pathway is very similar in almost every organism examined (72).

1.4.1.1 *E. coli* Uracil-DNA Glycosylase

Uracil-DNA glycosylase from *E. coli* (Ung) was the first enzyme in its class to be purified to apparent homogeneity (77). The native enzyme was determined to be monomeric with an apparent molecular weight of 24,500, as determined by denaturing polyacrylamide gel electrophoresis. The enzyme did not require a cofactor for activity, catalysis occurred in the presence of EDTA, and the enzyme

was predicted to be a globular protein based on a frictional ratio $f/f_0 \sim 1.2$. Ung also displayed maximum activity at pH 8.0 with 50% activity remaining at pH 7.0 and pH 9.5 (80). Subsequently, the *E. coli ung* gene was cloned and Ung was overexpressed (9, 32). The deduced nucleotide sequence indicates that the protein is made up of 228 amino acids with a molecular weight of 25,694. After post-translational removal of the N-terminal methionine, the protein has a molecular weight of 25,563 (138), which was verified by MALDI mass spectrometry (6). Comparison of the amino acid sequence of Ung with that of the human enzyme showed the two proteins share 56% identity (105). Ung was shown to use a processive search mechanism for locating sequential uracil residues on the same DNA strand (8, 48) and to prefer uracil residues located in single-stranded DNA over uracil residues in double stranded DNA by ~2 fold (8, 80). Similarly, the enzyme recognized U/G mispairs about 2.4-fold more efficiently than U/A basepairs in oligonucleotide substrates (8). The minimum substrate for Ung was determined to be pd(UN)p (139), with the 3'-phosphate to deoxyuridine being essential for binding and catalysis (104).

1.4.1.2 Crystal Structures of Human Uracil-DNA Glycosylases

The first uracil-DNA glycosylase structure solved was for the human enzyme, termed UDG. This structure was for a recombinant UDG in which the N-terminal 84 amino acids of the protein encoded by the *UNG* gene had been replaced by three N-terminal residues encoded by the vector (91). Hereafter, the recombinant protein is termed UDG Δ 84. Prior to crystallization, this truncated

protein had been determined to be active (127). Crystal structures of UDG Δ 84 and UDG Δ 84 in complex with the inhibitor 6-aminouracil were solved to 2.0 Å and 2.3 Å, respectively (91). The results showed the protein is made up of a central, four stranded, all parallel β -sheet surrounded on either side by a total of eight α -helices. Within this α/β fold, a groove is present with a width of 21 Å, approximately equal to the diameter of a DNA double helix, and lined with basic amino acids. The groove narrows to almost 10 Å between residue Pro₁₅₀ and Pro₁₆₅, suggesting that the groove floor cannot directly bind double-stranded DNA without significant conformational change. Site-directed mutagenesis studies and the structure of the UDG Δ 84:6-aminouracil complex located the catalytic site within the groove. Structural features of the groove, primarily the presence of rigid proline residues, suggested the groove and the active site would not undergo major conformational changes for binding DNA. The structure of the UDG Δ 84:6-aminouracil complex showed that the side chain of Tyr₁₄₇ of the active site sterically blocks bases substituted at the base 5 position, thus revealing how the enzyme discriminates between uracil and thymine. Hydrogen bonds between the protein and the 2, 3, and 4 positions of uracil allowed for discrimination against cytosine. The overall architecture of the enzyme suggested that for uracil to reach its binding pocket, the base must be “flipped out” from the DNA helix. A conserved Leu₂₇₂ residue directly above the uracil-binding pocket was positioned to potentially insert its side chain into the DNA helix to aid in melting the DNA helix. From this structure, base flipping as a structural mechanism for UDG damage recognition was proposed (91). Base

flipping was first observed in the co-crystal structure of *HhaI* DNA methyltransferase (70) and substrate DNA and then subsequently observed in the co-crystal structures of *HaeIII* methyltransferase (109) and T4 endonuclease V with substrate DNA (140). Base flipping has also been suggested from crystal structures of proteins without the presence of substrate DNA. These proteins include human apurinic/apyrimidinic endonuclease (42), human 3-methyladenine DNA glycosylase (74), *E. coli* photolyase (97), *E. coli* endonuclease III (134), *E. coli* exonuclease III (92), and *E. coli* mismatch-specific uracil DNA glycosylase (5).

The UDG Δ 84 protein was used again to form a complex with the uracil-DNA glycosylase inhibitor protein (Ugi) and the crystal structure was solved to 1.9 Å (90). Ugi is a 9,474 molecular weight protein product from bacteriophage PBS2 that forms an essentially irreversible complex with UDG (90), as well as Ung (7, 9) and uracil-DNA glycosylases isolated from other species (93). The Ugi structure consists of a twisted five-stranded antiparallel β -sheet and two α -helices. The first strand of the β -sheet, β_1 , was inserted into the active site groove of UDG Δ 84 without contacting the catalytic pocket. Ugi was bound to the enzyme via shape and electrostatic complementarity and by insertion of the UDG Leu₂₇₂ residue into a hydrophobic pocket of Ugi. These interactions were thought to mimic those suspected to take place when UDG binds to DNA. The UDG Δ 84:Ugi structure was used to propose a model for UDG binding to DNA. The model suggested that Leu₂₇₂ penetrated the DNA base stack from the major groove resulting in uracil flipping out of the major groove side of the DNA and into

the catalytic pocket (90). To further access this model, UDG bound to substrate DNA was needed.

Enzyme kinetic and DNA-binding studies showed that mutating Leu₂₇₂ to Arg enhanced the binding of UDG to double-stranded DNA and the mutation Asp₁₄₅ to Asn decreased its catalytic activity (128). This resulted in the 2.9 Å resolved crystal structure of a double mutant UDG Δ 84 (Leu₂₇₂ to Arg and Asp₁₄₅ to Asn) bound to double-stranded uracil-containing DNA with free uracil bound in the specificity pocket (128). The structure confirmed flipping of the damaged nucleotide since the uracil, deoxyribose, and 5' phosphate were rotated 180° from their position in the DNA. Interestingly, the majority of the contacts with the DNA substrate took place with the uracil-containing strand. From this structure, five steps involved in damaged base recognition and catalysis were proposed. First, electrostatic interactions position the UDG active-site groove along the DNA. Damage is detected and the structural loop containing Leu₂₇₂ pushes into the base stack via the DNA minor groove to promote nucleotide flipping via the major groove. The binding of UDG to sugar phosphates promotes DNA backbone compression to favor the extrahelical nucleotide. Then the substrate specificity pocket recognizes, or “pulls”, the flipped-out nucleobase, deoxyribose and phosphate. Finally, active-site hydrogen bonds polarize the N-C1' bond making it susceptible to attack by a water nucleophile activated by an Asp residue (128). These steps are termed the “push-pull” mechanism.

Additionally, structures of UDG Δ 84 and a mutant of UDG Δ 84, Leu₂₇₂ to Ala, were co-crystallized with U•A and U•G 10-base-pair oligonucleotides and

refined to 1.9 Å and 2.25 Å, respectively (96). The UDG84-DNA complexes were shown to be trapped product complexes containing bound uracil in the catalytic pocket and an extrahelical abasic deoxyribose. The DNA both 5' and 3' to the flipped-out nucleotide was compressed 4 Å by rigid enzyme loops causing the DNA to kink by ~45°. These loops contained Ser, Pro, and Gly residues and it was determined that free UDG could not bind B-DNA without pinching and bending the DNA backbone (96). This pinching mechanism defined a means for lesion detection. In these UDG84-DNA complexes, Leu₂₇₂ penetrated into the DNA base stack and replaced the flipped-out uracil. In the Leu₂₇₂ to Ala mutant, the small side chain of alanine could not fill the hole in the base stack, but the protein loop containing Ala₂₇₂ (normally Leu₂₇₂ in UDG84) could still penetrate the base stack as deeply as in the UDG84 structure to promote flipping. This suggested that the Leu₂₇₂ side chain push was not essential for nucleotide flipping. The mutant protein-DNA structure showed only a flipped-out AP site and the catalytic pocket contained two ordered water molecules rather than a uracil base. It appeared as if the mutant enzyme had cleaved the uracil, disengaged from the substrate, released the cleaved uracil from the catalytic pocket, and rebound the AP site containing DNA. It was suggested that human UDG might serve a role in protecting cells from the cytotoxic effects of AP sites (96).

1.4.1.3 Crystal Structures of *E. coli* Uracil-DNA Glycosylase

The first published crystal structure for *E. coli* uracil-DNA glycosylase (Ung) was for the enzyme co-crystallized with the uracil-DNA glycosylase inhibitor protein (Ugi). The structure was refined to 3.2 Å and showed that Ung is made up of a four-stranded parallel β -sheet and 11 separate helical stretches, four of which are greater than 10 residues and the remaining helices are five residues or less. The structural comparison of the Ung:Ugi complex with the UDG:Ugi complex revealed that the nature of the binding site and mechanism of action for the two enzymes are essentially the same (108).

A second crystal structure of the Ung:Ugi complex has also been solved using the native enzyme and a His187 to Asp mutant (105). This study showed that Ugi resembled an intermediate of the DNA in the base-flipping process instead of directly mimicking DNA. This, along with the shape, charge, and hydrophobic complementarity between the two proteins indicates how Ugi reacts with Ung instead of other DNA-binding proteins.

1.4.2 Human Replication Protein A

Single-stranded DNA-binding proteins participate in almost every aspect of DNA metabolism, including replication, repair, and recombination. The primary role of these proteins is to bind single-stranded DNA and reduce secondary structure. In eukaryotic cells, this role is played by replication protein A (RPA). As the characterization of the RPA structure and function is revealed, it

becomes apparent that RPA plays a role in modulating DNA metabolism events through specific interactions with other proteins and with DNA.

RPA was originally purified from human (HeLa) cell extracts and identified as a protein required for simian virus 40 (SV40) DNA replication (149, 151). Analysis of human RPA (hRPA) indicated that it is a heterotrimer composed of 70-kDa (hRPA70), 32-kDa (hRPA32), and 14-kDa (hRPA14) subunits (150). Since its identification in human cells, heterotrimeric homologues of hRPA have been found in virtually every eukaryotic cell examined, including *S. cerevisiae* (17), *S. pombe* (98), *X. laevis* (1), *D. melanogaster* (89), and *C. fasciculata* (18). Comparison of the amino acid sequence of various RPA subunits has revealed significant homology between species, with the large subunit showing the highest homology. Amino acid sequence comparisons have shown 44.5%, 50.0%, and 80.7% similarity for the large RPA subunits of *S. cerevisiae*, *S. pombe*, and *X. laevis*, respectively, versus hRPA70 (55).

1.4.2.1 DNA Binding Domains of Human Replication Protein A

UV-crosslinking of hRPA to single-stranded DNA or oligonucleotide probes indicated that hRPA70 is the only subunit to have significant single-stranded DNA-binding activity (119). This result was supported by mapping a single DNA-binding domain to the center of hRPA70 (residues ~180 to 420) (40, 41, 69, 76, 101). The crystal structure of the central DNA-binding domain of hRPA70 complexed to a single molecule of octa-dCMP has been obtained at 2.2 Å resolution (14). The crystal structure revealed a DNA binding domain

composed of two copies of a single-stranded DNA binding subdomain that have the same three dimensional structure. These subdomains resemble an OB fold motif that has been previously found in various oligonucleotide- and oligosaccharide-binding proteins (94). The subdomains are oriented in tandem, and a DNA binding channel extends from one subdomain to the other. The two primary types of interactions observed between the protein and the DNA were 1) aromatic amino acids (Phe₂₃₈, Phe₂₆₉, Trp₃₆₁, and Phe₃₈₆) that stack with individual nucleic acid bases and 2) hydrogen bonds that occur between amino acid side chains and both the phosphate backbone and individual nucleic acid bases (14). The importance of hRPA70 aromatic amino acids interacting with single-stranded DNA has been questioned. Site directed mutants of hRPA70 in which Phe₂₃₈, Phe₂₆₉, Trp₃₆₁, or Phe₃₈₆ were mutated to alanine resulted in only slight changes in the hRPA binding constant. A double mutant, in which Phe₂₃₈ and Trp₃₆₁ were mutated to alanine, resulted in a decrease of the hRPA binding constant by 3 orders of magnitude (142). The double mutant, however, was determined by proteolytic sensitivity experiments to be structurally disrupted. Structural perturbation makes it difficult to draw conclusions about the double mutant phenotype, but the single aromatic residue mutants suggest that no one aromatic residue is critical for the formation of the hRPA-DNA complex (142).

Studies have indicated that hRPA70 has a 30 nt (67, 119) and a 8 nt (12) DNA-binding mode. They suggest the possibility that the protein may contain multiple binding sites, possibly within the hRPA32 and/or hRPA14 subunits. Some direct evidence exists that this is possible since hRPA32 can be

crosslinked to synthetic primer-template DNA molecules *in vitro* (75). Additionally, a complex of hRPA14 and N- and C-terminal deleted hRPA32 (residues 43 to 171) was found to bind single-stranded DNA weakly, as opposed to a full length hRPA14:hRPA32 complex that showed no significant single-stranded DNA binding activity (15). From these results it has been suggested that the hRPA32 single-stranded DNA-binding domain is normally cryptic, and only becomes accessible after hRPA70 binds to single-stranded DNA. Recently, the crystal structure of the hRPA14:hRPA32 complex (residues 43-171 of hRPA32) has been solved to 2.5 Å resolution (13). The crystal structure revealed that both hRPA14 and the central part of hRPA32 are structurally similar. Each subunit contains an OB-fold domain that resembles the DNA-binding domains in hRPA70. Even though limited biochemical evidence is available to show single-stranded DNA-binding activity in hRPA32 or hRPA14, these structural data strongly suggest that these two subunits may harbor DNA-binding activity.

1.4.2.2 *Binding of Human RPA to Single-Stranded DNA*

Heterotrimeric hRPA binds to single-stranded DNA with an intrinsic binding constant (K) measured between $10^9 - 10^{11} \text{ M}^{-1}$; this is three orders of magnitude higher than for double-stranded DNA and RNA (41, 67, 68). The binding preference of hRPA has been found to be ~50-fold higher for polypyrimidines versus polypurines. The affinity of hRPA binding to nucleic acids is single-stranded polypyrimidines > mixed sequence single-stranded DNA >

single-stranded polypurines $\gg$ RNA $\geq$ double-stranded DNA (67).

Measurements of hRPA binding to oligonucleotides are strongly dependent on oligonucleotide length. The preference of hRPA for oligodeoxythymidine 20 or 30 residues in length (dT₂₀ and dT₃₀) is high, with $K \approx 4 \times 10^9 \text{ M}^{-1}$, but as the length decreases the affinity constant decreases to $K = 1 \times 10^9 \text{ M}^{-1}$ for dT₁₅, $K = 2 \times 10^8 \text{ M}^{-1}$ for dT₁₂, and $K = 7 \times 10^7 \text{ M}^{-1}$ for dT₁₀ (66). This corresponds to a decrease in the binding constant of almost two orders of magnitude. The strong binding dependence of hRPA on oligonucleotide length suggests that it may adopt a different conformation when binding to longer oligonucleotides or that additional contacts occur between the protein and longer oligonucleotides. A conformational change in the protein has been measured upon binding to single-stranded DNA. This was demonstrated by changes in proteolytic sensitivity of the hRPA70 and hRPA32 subunits (39).

As mentioned above, the binding affinity of hRPA for non-specific double-stranded DNA is at least three orders of magnitude weaker than for single-stranded DNA. High-affinity binding of hRPA to a double-stranded DNA sequence, possibly involved in the regulation of transcription, has been reported (131). Since no consensus sequence has been identified for hRPA binding to double-stranded DNA, it is likely that hRPA recognizes a specific DNA structure associated with the sequence rather than the sequence itself. hRPA has also been shown to bind damaged double-stranded DNA. Double-stranded DNA that has been modified with *cis*-platin (99, 100), UV mimetic N-acetoxy-2-acetylaminofluorene (45), or UV light (22) can act as substrates for hRPA. In the

case of UV light, hRPA appears to bind preferentially to 6-4 photoproducts (22). Since hRPA is essential for the recognition of damaged DNA in nucleotide excision repair, this mode of binding is likely physiologically relevant.

1.5 Research Objectives

The studies of Barofsky and co-workers demonstrated the feasibility of using MALDI MS (57, 58) and ESI MS/MS (59) to identify sites of contacts within protein-nucleic acid interfaces. The results from the model systems, combined with those obtained from crosslinking studies of *E. coli* Ung (6) suggested that the analytical methodology associated with analyzing protein/peptide-nucleic acid complexes by mass spectrometry were worked out. Therefore, the primary objective of this thesis was to continue the application of mass spectrometry to characterizing the DNA-binding domain of Ung. Specifically, use ESI tandem mass spectrometry to identify amino acids within the DNA-binding domain of the protein that crosslink to dT₂₀ and are thus in contact with DNA. This research would give insight into how Ung binds to single-stranded DNA and the specific amino acids that take part in binding.

After initial crosslinking experiments were conducted with Ung and dT₂₀, it became apparent that additional analytical methodologies were required in order to utilize mass spectrometry for studying peptide-nucleic acid complexes of Ung. The development of a generalized procedure for the isolation and purification of peptide-nucleic acid complexes from protein-nucleic acid complexes was

undertaken in order to provide sample that was appropriate for mass spectral analysis.

At the beginning of this thesis research, only one ESI MS/MS spectrum of a peptide-nucleic acid complex had ever been reported (59) and this spectrum was for a synthetic peptide-nucleic acid complex. Little, if anything, was known about the fragmentation patterns associated with peptide-nucleic acid complexes produced from UV-catalyzed crosslinking reactions. Investigations were needed to know what effect the oligonucleotide portions of these complexes had on the fragmentation of the peptide portions and vice versa, how fragmentation patterns were influenced by the charge state of these complexes, and whether the means could be found to direct the fragmentation patterns of these types of molecules. Thus, the ESI tandem mass spectra from peptide-nucleic acid complexes of Ung were not only analyzed to identify sites of crosslinking, but thoroughly characterized to gain an understanding of the fragmentation characteristics for these types of molecules.

The final research objective was to characterize the DNA-binding domain of human replication protein A (hRPA) using UV-catalyzed crosslinking. It was surmised that crosslinking studies would give a better understanding of how this protein binds to single-stranded DNA in a length dependent manner and the role specific aromatic amino acids play in binding to DNA. The study of this protein would also test the general applicability of the developed purification protocol and produce additional tandem mass spectra that would give insight into the fragmentation mechanisms of peptide-nucleic acid complexes.

2. MATERIALS AND EXPERIMENTAL PROCEDURES

2.1 Materials

2.1.1 Chemicals

ATP, ampicillin, angiotensin (DRVYIHPFHL), bovine serum albumin (BSA), citric acid (diammonium salt), copper (II) chloride, chloramphenicol, ethylenediaminetetraacetic acid (EDTA), HEPES, sodium chloride, Trizma (Tris-base) and streptomycin sulfate were purchased from Sigma Chemical Co. Acrylamide (>99% pure), ammonium persulfate, 2-mercaptoethanol, bis N,N'-methyl-bis-acrylamide, and N,N,N',N'-tetramethylethylenediamine (TEMED) were from Bio-Rad. Ammonium sulfate, dithiothreitol (DTT), glycine, isopropyl- β -D-thiogalactopyranoside (IPTG), sodium dodecyl sulfate (SDS), and urea were from Life Technologies. Acetic acid, acetonitrile, ammonium bicarbonate, ammonium hydroxide (50%), hydrochloric acid, isopropanol, magnesium chloride, methanol, potassium chloride, sodium hydroxide, sodium thiosulfate, and trichloroacetic acid were supplied by Fisher Chemical. Glycerol, potassium phosphate (dibasic and monobasic salts) and triethylamine were from J.T. Baker. Chlorosulfonylacetyl chloride, α -cyano-4-hydroxycinnamic acid, N, N-diisopropylethylenamine, iodoacetamide, silver nitrate, sinapinic acid, 2-sulfobenzoic acid cyclic anhydride, tetrahydrofuran, 2',4',6'-trihydroxyacetophenone, and trimethylamine were supplied by Aldrich Chemical Co. Calcium chloride and formic acid were from EM Science, while formalin,

potassium chloride, potassium hydroxide, and sodium carbonate were from Mallinckrodt. Trifluoroacetic acid was from Fluka. Agar, tryptone, and yeast extract were supplied by Difco.

2.1.2 Enzymes

Trypsin (L-1-tosylamide-2-phenyl chloromethyl ketone-treated) was obtained from Worthington Biochemical as a lyophilized powder. The protease was resuspended in 0.1 N HCl, divided into 500- μ g portions in Eppendorf tubes, dried by vacuum centrifugation, and stored at -20 °C until used. Nuclease P1 was supplied by Pharmacia as a lyophilized powder. The nuclease was resuspended in 50 mM sodium acetate (pH 5.5) at 1 unit/ μ L, divided into 10- μ L portions, flash frozen in liquid nitrogen, and stored at -20 °C. All dilutions of the nuclease were made with 50 mM ammonium acetate. T4 polynucleotide kinase was purchased from New England Biolabs and used without further preparation.

2.2 General Experimental Procedures

2.2.1 Chromatographic Resin Preparation

2.2.1.1 Preparation of Sephadex G-25, Bio-Gel P4 Gel, and Hydroxyapatite Bio-Gel HTP

Resins were hydrated in 6- to 10-fold volumes of the appropriate buffer, "defined" a minimum of 3 times, and stored in the same buffer at 4 °C as 50% (v/v) slurries, unless stated otherwise. Sephadex DEAE A-50 was hydrated in 50

mM NH_4HCO_3 . Sephadex G-75 (Pharmacia) was hydrated in 10 mM triethylammonium bicarbonate (pH 7.0) and immediately poured. Bio-Gel P4 Gel (Bio-Rad) was hydrated in DAB buffer (30 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1 mM dithiothreitol, and 5% (w/v) glycerol), and hydroxyapatite Bio-Gel HTP (Bio-Rad) was hydrated in 20 mM potassium phosphate buffer (pH 7.5).

2.2.1.2 Preparation of Dowex AG 1-X8 Resin

A 1-L graduated cylinder was filled with 800 mL of 1 M NaOH and dry Dowex 1-X8 was added until the volume in the cylinder was displaced to a total volume of 1 L. The slurry was mixed by inversion, allowed to settle (approximately 30 min), and the supernatant was removed by using an aspirator. The cylinder was filled to 1 L with 1 M NaOH and mixed by inversion. The slurry was allowed to settle (approximately 30 min) and the supernatant was removed. The addition of 1 M NaOH and subsequent removal was repeated until a total of 4 L of NaOH had been used to wash the resin. The resin was transferred to a glass filter funnel (60-C), allowed to drain by gravity, washed with 600 mL of 1 M ammonium formate (pH 4.2), allowed to drain, and washed with 2 L of 10 mM ammonium formate (pH 4.2). After the resin fully drained, it was transferred to a storage container and stored at 4 °C as a 50% (v/v) slurry in 10 mM ammonium formate (pH 4.2).

2.2.1.3 Preparation of Affi-gel Blue Gel

One part Affi-gel Blue Gel resin slurry (Bio-Rad) was thoroughly mixed at 4 °C with six parts HI buffer (30 mM HEPES-KOH (pH 7.8), 1 mM dithiothreitol, 0.25 mM EDTA, 0.25% (w/v) inositol, and 0.01% (v/v) Nonidet P-40) containing 50 mM KCl and allowed to settle (approximately 30 min). The buffer was decanted and mixed with six parts HI-30 buffer. This defining process was repeated eight times and then the column was immediately poured.

2.2.2 Dialysis Tubing

SpectraPor dialysis tubing was cut into 30-cm lengths, placed into a 1-L beaker, and soaked in 1% acetic acid for 1 h. The tubing was rinsed with water. A solution of 1% (w/v) NaHCO_3 with 0.1% (w/v) EDTA was added to the beaker and heated to a boil. The boiled solution was replaced with a fresh solution of 1% (w/v) NaHCO_3 with 0.1% (w/v) EDTA and heated to a boil again. The tubing was washed once with water and then heated to a boil in water. The water was removed and the treated dialysis tubing was stored in 10 mM EDTA at 4 °C.

2.2.3 Preparation of dT₂₀

Oligonucleotide dT₂₀ was synthesized by the Biopolymer Core Facility at the University of Maryland at Baltimore. To desalt the dT₂₀, the lyophilized sample (approximately 5 μmole) was resuspended in 1 mL of 10 mM

triethylammonium bicarbonate buffer (pH 7.2) and applied to a $1.79 \text{ cm}^2 \times 10 \text{ cm}$ Sephadex G-25 column equilibrated in the same buffer. The column was operated at a 20 ml/h flowrate 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 dT₂₀ of $\epsilon_{260} = 1.7 \times 10^5 \text{ l/mol}\cdot\text{cm}$. The pooled peak was analyzed by MALDI MS to check for incomplete synthesis products and then divided into 1- or 1.5-mL fractions that were dried by vacuum centrifugation. The dried fractions were stored at -80°C until further use.

2.2.4 Preparation of dT₁₀, dT₃₀, and dT₇₀

Oligonucleotides dT₁₀, dT₃₀, and dT₇₀ were synthesized and HPLC purified by the Biopolymer Core Facility, at the University of Maryland at Baltimore. Oligonucleotide dT₁₀ was used without further purification. Oligonucleotides dT₃₀ and dT₇₀ (~30 nmol of each) were run out on a 15% native polyacrylamide gel. Oligonucleotide bands were located within the gel by UV shadowing, bands were cut from the gel, and oligonucleotides were electroeluted from the gel for 2 h by using an Elutrap device and 1x TAE (40 mM Tris-acetate, 1 mM EDTA) as the elution solution. Both oligonucleotides were buffer exchanged with 4 mL of TE buffer (10 mM Tris-HCl (pH 7.2) and 1 mM EDTA) and concentrated to about 130 μL in a Centricon-3. The absorption at 260 nm was used to estimate the quantity of oligonucleotide by using molar extinction coefficients of $8.4 \times 10^4 \text{ l/mol}\cdot\text{cm}$ for dT₁₀, $2.5 \times 10^5 \text{ l/mol}\cdot\text{cm}$ for dT₃₀, and $5.9 \times 10^5 \text{ l/mol}\cdot\text{cm}$ for dT₇₀.

2.2.5 5'-End Phosphorylation of Oligonucleotides

At 4 °C, desalted oligonucleotides were resuspended in 470 µL PNKB buffer (50 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, 5 mM dithiothreitol, and 0.1 mM EDTA) and 10 µL of [γ -³²P]ATP (6000 Ci/mmol) and 10 µL (100 units) of T4 polynucleotide kinase were added. Reaction mixtures were incubated at 37 °C for 15 min at which time 10 µL of 100 µM cold ATP was added. Incubation at 37 °C was continued for an additional 45 min. Labeling reactions were stopped by incubating the mixture at 65 °C for 5 min. Unreacted ATP was removed from 250-µL fractions by immediate passage through two consecutive 1-mL mini-columns containing P-4 resin equilibrated in DAB buffer (30 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1 mM dithiothreitol, and 5% (w/v) glycerol). The mini columns were operated in a clinical centrifuge (IEC) at setting 4 for 2.5 min. The concentrations of final [³²P]-labeled oligonucleotide solutions were measured spectrophotometrically and the radioactive specific activity was determined by liquid scintillation counting using Formula 989 cocktail.

2.2.6 Polyacrylamide Gel Electrophoresis

2.2.6.1 Large Gel Denaturing Polyacrylamide Gel Electrophoresis

Denaturing polyacrylamide gel electrophoresis was conducted using 17 cm x 15 cm x 1.5 mm slab gels (Gibco-BRL). Stacking gels (3 cm) were composed of 3% acrylamide, 0.1% N,N'-methylenebis acrylamide, 125 mM Tris-HCl (pH 6.8), and 0.1% SDS. Resolving gels were composed of acrylamide in

0.4% N,N'-methylenebis acrylamide, 375 mM Tris-HCl (pH 8.8), 0.1% SDS.

Samples were usually mixed with 5x cracking buffer (0.125 M Tris-HCl (pH 6.8), 360 mM 2-mercaptoethanol, 25% (w/v) glycerol, 2.5% SDS, 0.1% bromophenol blue), heated at 56 °C for 15 min and loaded. Electrophoresis was conducted with 25 mM Tris base, 200 mM glycine, and 0.1% SDS as the running buffer and at 200 V and room temperature until the tracking dye had migrated to 2 cm from the bottom of the gel. Gels were stained by Coomassie stain, copper, or silver as described below. When needed, gels were dried and autoradiography was performed using X-OMAT (Kodak) film. Wet gels were photographed with Polaroid-55 or -57 film using an orange filter.

2.2.6.2 Mini Gel Denaturing Polyacrylamide Gel Electrophoresis

Denaturing polyacrylamide gel electrophoresis was also conducted using the Mini-Protean II system (Bio-Rad) with 1 mm spacers. Stacking gels (2 cm) were composed of 3% acrylamide, 0.1% N,N'-methylenebis acrylamide, 125 mM Tris-HCl (pH 6.8), and 0.1% SDS. Resolving gels were composed of acrylamide in 0.4% N,N'-methylenebis acrylamide, 375 mM Tris-HCl (pH 8.8), 0.1% SDS. Samples were usually mixed with 5x cracking buffer (0.125 M Tris-HCl (pH 6.8), 360 mM 2-mercaptoethanol, 25% (w/v) glycerol, 2.5% SDS, 0.1% bromophenol blue), heated at 56 °C for 15 min and loaded. Electrophoresis was conducted with 25 mM Tris-base, 200 mM glycine, and 0.1% SDS as the running buffer and at 200 V and room temperature until the tracking dye had migrated 0.5 cm from the bottom of the gel. Gels were stained by Coomassie stain, copper, or silver as

described below. When needed gels were dried and autoradiography was performed.

2.2.6.3 Native Polyacrylamide Gel Electrophoresis

Native, continuous, polyacrylamide gel electrophoresis was conducted using 17 cm x 15 cm x 1.5 mm slab gels (Gibco-BRL). Gels were composed of acrylamide in 0.4% N,N'-methylenebis acrylamide and 375 mM Tris-HCL (pH 8.8). Samples were mixed with 5x loading buffer (0.125 M Tris-HCl (pH 8.8), 25% (w/v) glycerol, and 0.1% bromophenol blue) and immediately loaded. Electrophoresis was conducted with 25 mM Tris-base and 200 mM glycine as the running buffer and at 250 V until the tracking dye had migrated to about 2 cm from the bottom of the gel.

2.2.7 Gel Staining

2.2.7.1 Coomassie Staining

Staining was conducted by covering gels with a solution of 0.05% (w/v) Coomassie Brilliant Blue G-250 in 50% methanol and 10% acetic acid that had been filtered through Whatman 1 filter paper. For large gels, staining was performed overnight. For mini-gels, staining was performed for 1 h. Destaining was carried out in 5% methanol and 7% acetic acid until the desired band intensity was reached.

2.2.7.2 Copper Staining

Immediately following electrophoresis, large gels were rinsed in water for 5 min, the water was removed, and gels were stained with a solution of 300 mM copper (II) chloride for 5 min. The staining solution was removed and gels were rinsed with water for 5 min. For mini-gels, gels were immediately stained in 300 mM copper (II) chloride for 5 min. The staining solution was removed and gels were rinsed with water for 5 min. Negatively stained protein gel bands were visualized by holding a black piece of paper behind the gel and observing black protein bands against a blue-white background.

2.2.7.3 Silver Staining

Silver staining was performed as described by Schevchenko *et al.* (126) with modifications. Immediately following electrophoresis gels were fixed in 50% methanol and 5% acetic acid for 20 min. The fixing solution was replaced with 50% methanol, washed for 10 min, replaced with water, and the gel was washed for 1 h. Water was removed and gels were then washed with a sensitizing solution of 0.02% sodium thiosulfate for 1 min. The sensitizing solution was removed and gels were rinsed for 1 min in water, with one subsequent repetition. Gels were stained by submerging them in a 0.1% silver nitrate solution for 20 min. After staining, gels were rinsed twice in water, 1 min for each rinse, and then developed in a solution containing 0.04% formalin and 2% sodium carbonate. Once bands were visible at the desired intensity, an equal volume of

10% acetic acid was mixed with the developing solution to stop the reaction. The stopped developing solution was removed and replaced with 5% acetic acid.

2.2.8 In-gel Reduction and Alkylation

Gel slices were dehydrated with 100% acetonitrile, the solvent was removed, and gel slices were dried under vacuum. Once dried, the gel slices were covered with 10 mM DTT in 100 mM NH_4HCO_3 and reduced for 1 h at 56 °C. This solution was then replaced with freshly prepared 55 mM iodoacetamide in 100 mM NH_4HCO_3 and gel slices were incubated at room temperature for 45 min.

2.2.9 In-gel Proteolytic Digestion

Gel slices were prepared for in-gel proteolytic digestion by washing them with 100 mM ammonium bicarbonate, dehydrating with 100% acetonitrile, rehydrating with 100 mM ammonium bicarbonate, and dehydrating again with 100% acetonitrile. Each gel preparation step was carried out for 15 min with vortexing. The final dehydration liquid was removed and the gel pieces were dried under vacuum for 3 h. Gel pieces were rehydrated on ice in digestion buffer (50 mM ammonium bicarbonate, 5 mM calcium chloride) containing 36 ng/ μL trypsin. After 1 h, the trypsin solution was removed from each gel piece, replaced with 40 μL digestion buffer, and the digestion was continued at 37 °C

overnight. MALDI-MS analysis was conducted on the resulting digestion solution.

2.2.10 Nuclease P1 Digestion

Nuclease P1 was diluted with 50 mM ammonium acetate at a nuclease to buffer ratio of 1:25. Digestion was conducted by mixing one part of the resulting nuclease solution (0.04 units/ μ L) with one part sample and incubated at 37 °C for 2 h.

2.2.11 Off-line High Performance Liquid Chromatography

The off-line HPLC system consisted of an ABI 140B syringe pump (Perkin-Elmer) with a flow rate split to 5 μ L/min and a 0.32 mm x 250 mm column packed with 5- μ m particle, 150-Å pore, Beta Basic C₁₈ (Keystone Scientific) packing material. An ABI 759A UV detector equipped with a 35-nl dead volume and 8-mm pathlength flowcell (LC Packings) was used for detection. All system connections were made with 50- μ m (i.d) silica tubing. The solvent system of (A) 5 mM ammonium acetate and (B) 100% acetonitrile was delivered with a gradient of 5% to 50% B over 45 min. Eluted species were monitored at 217 nm and collected for identification by MALDI-MS and subsequent nano-ESI tandem mass analysis.

2.2.12 Matrix-Assisted Laser-Desorption Ionization Mass Spectrometry

MALDI mass spectrometry was performed at Oregon State University (Mass Spectrometry Core Facility) using a custom-built delayed extraction time-of-flight instrument. Samples were mixed with matrix, applied to the probe by using the wipe-wash method, allowed to dry on the probe by slow evaporation, and rinsed with Millipore-purified water. All mass spectra were collected in the positive-ion mode by summing the signals generated from 30 individual laser pulses.

2.2.12.1 MALDI Analysis of Proteins

The typical matrix solution used for the analysis of proteins (mainly Ung) was 10 mg/mL of sinapinic acid in 50% acetonitrile and 0.1% trifluoroacetic acid. Analyte was usually mixed with matrix at a ratio of 1:3 and α -chymotrypsinogen ($M+H^+=25,657$) was used as an external mass calibrant.

2.2.12.2 MALDI Analysis of Peptides

MALDI analysis of peptides was conducted using a 5-mg/mL solution of the matrix α -cyano-4-hydroxycinnamic acid (HCCA) in 0.1% trifluoroacetic acid, 33% acetonitrile. Analyte was typically mixed with matrix at a ratio of 1:5 (peptides to matrix). Angiotensin I ($MH^+=1297.5$) and bovine insulin ($MH^+=5734.6$) were used as external mass calibrants.

2.2.12.3 MALDI Analysis of dT₂₀ and Nucleopeptides

The matrix solution used for dT₂₀ and peptide × dT₂₀ samples was a saturated solution of 2',4',6' trihydroxyacetophenone (THAP) in 50 mM diammonium hydrogen citrate and 50% acetonitrile. Analyte was typically mixed with matrix at a ratio of 1:1 or 1:2 (analyte to matrix) and a synthetic polymer (M+H⁺=5043.4) was used as an external mass calibrant.

2.2.13 Electrospray Ionization Mass Spectrometry

2.2.13.1 Nanospray Triple Quadrupole Mass Analysis

All nanospray analyses were conducted using a SCIEX (Perkin Elmer) API III triple quadrupole mass spectrometer in which the commercial electrospray source was replaced with a nanospray source constructed by the Protein and Peptide Group, European Molecular Biology Laboratory, Heidelberg, Germany (147). A curtain gas of nitrogen was used at a flow rate of 0.6 L/min, the needle voltage was 700 V, the interface plate voltage was 100 V, and the orifice potential was 80 V. Samples collected from off-line HPLC fractions were adjusted to 0.1% formic acid and 40% acetonitrile and 0.6 µL of the adjusted sample was loaded into a nanospray needle (Protana). Scans in the Q₁ mode were performed with 0.3 mass unit steps for 3 seconds per scan. For operation in the MS/MS mode, Q₁ was set to transmit a mass window of 2 mass units and mass spectra were accumulated with 0.3 mass unit steps for 2 min per scan. Resolution was adjusted to provide maximum sensitivity and the collision gas

thickness (CGT) was set to 230×10^{12} molecules/cm², with argon being the collision gas. Collision energy was increased in a stepwise manner between 15 V and 35 V. Typically, the collision energy was increased in 5 V increments every two to four scans and the resulting scans were averaged. All analyses were performed in the positive ion mode.

2.2.13.2 LC Triple Quadrupole Mass Analysis

A Shimadzu VP HPLC system was operated at a flow rate of 80 μ L/min that was split to 5 μ L/min with a splitter box (LC Packings). A column of 0.32 mm x 100 mm packed with 5- μ L particle size, 300-Å pore size, Luna C18 (2) packing material was directly in-line with a SCIEX API-III triple quadrupole mass spectrometer (Perkin Elmer). The solvent system used was (A) 0.1% acetic acid, 0.01% trifluoroacetic acid, 5% methanol and (B) 0.1% acetic acid, 0.01% trifluoroacetic acid, and methanol. A curtain gas of nitrogen was used at a flow rate of 0.8 L/min, the needle voltage was 5000 V, and the orifice potential was 100 V. Scans in the Q₁ mode were performed with 0.3 mass unit steps for 2 s/scan. For operation in the MS/MS mode, Q₁ was set to transmit a mass window of 2 mass units and mass spectra were accumulated with 0.3 mass unit steps for 2 s/scan. Resolution was adjusted to provide maximum sensitivity and the collision gas thickness (CGT) was set to 230×10^{12} molecules/cm², with argon being the collision gas. The collision energy was optimized for each peptide analyzed, but was typically between 15 V to 30 V. All analyses were performed in the positive ion mode.

2.2.13.3 *Quadrupole Ion-trap Mass Analysis*

An ABI 140B syringe pump (Perkin-Elmer) and a 0.17 mm x 100 mm column packed with 5- μ m particle, 100-Å pore, Luna C18 (2) packing material (Phenomenex) were directly coupled to a Finnegan ESI-LCQ quadrupole ion trap mass spectrometer (ThermoQuest Corp). A solvent system of (A) 0.1% acetic acid, 0.01% trifluoroacetic acid and (B) 0.1% acetic acid, 0.01% trifluoroacetic acid in acetonitrile was delivered at a flow rate of 2 μ L/min and a gradient of 5% to 50% B over 30 min. The HPLC column was connected by 30 μ m (i.d.) silica tubing to a custom built ESI source in which the spray formed from the tip of a 2.5-cm length, 35 gauge, stainless steel hypodermic tubing that was positioned approximately 5 mm from the inlet capillary. The ESI source was operated with a needle voltage of 2300 V and a heated capillary was maintained at 180 °C. For MS/MS operations an isolation width of 1.5 m/z for the precursor ion and a collisional energy parameter at 27% of the highest available energy, as defined by the LCQ software, were used. All analyses were conducted in the positive ion mode.

2.2.14 Database Search Algorithms

Peptide sequencing and protein identifications were conducted with the aid of Protein Prospector (<http://prospector.ucsf.edu>) and SEQUEST (33, 155-157) database search algorithms.

2.3 Experimental Procedures for Ung Project

2.3.1 Uracil-DNA Glycosylase Activity Assay

Enzymatic activity of Ung was measured using a procedure similar to that of Bennett and Mosbaugh (7). Reaction mixtures of 100 μ L containing 70 mM HEPES-KOH (pH 8.0), 1 mM EDTA, 1mM DTT, 8.2 nmol of calf thymus [uracil- 3 H]DNA (180 cpm/pmol of uracil), and 25 μ L of Ung diluted in 50 mM HEPES-KOH (pH 8.0), 1 mM EDTA, 1 mM DTT and 100 μ g/ml bovine serum albumin. Reaction assays were mixed on ice and then incubated at 37 °C for 30 min. Assays were stopped by the addition of 250 μ L of ice cold 10 mM ammonium formate (pH 4.2). Aliquots from each stopped reaction mixture (300 μ L) were applied to Dowex AG 1-X8 columns (Pasteur pipette, 0.2 cm² $\times$ 2 cm) equilibrated in 10 mM ammonium formate. The columns were washed with 1.7 mL of 10 mM ammonium formate to elute free [3 H]uracil in two 1-mL fractions that were collected directly in separate scintillation vials. Formula 989 (5 mL) scintillation cocktail was added to each vial and 3 H radioactivity was measured in a Beckman (LS6800) liquid scintillation counter.

2.3.2 Purification of Uracil-DNA Glycosylase

The purification of Ung was based on a procedure previously described (6). *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 OD₆₀₀ 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 at 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 cm² × 88 cm) equilibrated in UE buffer, eluted with the same buffer,

and 7-mL fractions were collected. 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 under 55 psi of N₂ gas using a YM10 membrane. The sample was applied to a hydroxylapatite column (4.9 cm² × 15 cm) equilibrated in HA buffer, eluted with the same buffer, and 5-mL fractions were collected. Fractions containing uracil-DNA glycosylase activity were pooled and dialyzed overnight against DAB buffer (30 mM Tris-HCl (pH 7.4), 1 mM dithiothreitol, 1 mM EDTA, 5% (w/v) glycerol). The sample was applied to a single-stranded DNA agarose column (19 cm² × 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 N₂ gas and a YM10 membrane. The protein concentration was determined spectrophotometrically by using a molar extinction coefficient for Ung as $\epsilon_{280} = 4.2 \times 10^4$ l/mol·cm. The purified protein was stored at -80 °C until used.

2.3.3 Preparative Crosslinking of Ung to dT₂₀

Purified Ung (36 nmol) was combined with a 3-fold molar excess of dT₂₀ 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) 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 ($\lambda_{\text{max}}=254$ nm) in a Stratalinker 1800 (Stratagene Cloning Systems) for an appropriate period of time (see Chapter 3 sections 3.1.1.1 and 3.1.2.1). This solution was either used immediately or stored at -80 °C until future use.

2.3.4 Purification of Ung Nucleopeptides

2.3.4.1 *"In-pot" Method*

Five 1-mL crosslinked reaction mixtures were pooled in a 15-mL Corex tube on ice. Ice-cold trichloroacetic acid (TCA) was added to the pooled solution to give a final TCA concentration of 10% in the final volume. Free protein and crosslinked protein were allowed to precipitate on ice for 20 min and then the solution was cleared by centrifugation at 10,000 rpm in an SS-34 rotor at 4°C. The supernatant was removed and the pellet was washed twice with ice-cold 10% TCA, once with -20 °C acetone, and the pellet was dried under vacuum for 45 min. The pellet was resolubilized in 500 μ L of a freshly prepared solution of 8 M urea, 0.4 M NH₄HCO₃, and 5 mM DTT. This solution was incubated at 50 °C for 20 min at which time 10 μ L of fresh 500 mM iodoacetamide was added and

the solution was incubated again at 50 °C for 20 min. Water was added to dilute the solution to a final urea concentration of 2 M and trypsin was added in a 1:50 (w/w) ratio of trypsin to Ung and incubated at 37 °C for 2 h. At this time, an additional 1:50 (w/w) ratio of trypsin was added to the solution and incubation was continued at 37 °C for 19 hr. The digestion mixture was diluted 1:4 with 50 mM NH_4HCO_3 and applied to a Sephadex DEAE A-50 column ($1.79 \text{ cm}^2 \times 9 \text{ cm}$) equilibrated in the same buffer. The column was washed at 20 mL/hr with 21 mL of 50 mM NH_4HCO_3 , 40 mL of NH_4HCO_3 containing 100 mM NaCl, and eluted with a 90-mL linear gradient of 100 mM to 800 mM NaCl in NH_4HCO_3 . The column was washed with an additional 30 mL of 800 mM NaCl in NH_4HCO_3 to complete the gradient. Fractions of 1 mL were collected and surveyed spectrophotometrically at 260 nm and 280 nm and by conductivity to generate an elution profile. The elution peak fractions were pooled and loaded onto a Nensorb-20 (DuPont) column that had been prewetted with 100% methanol and equilibrated with 20 mM triethylammonium acetate (pH 7.2). The column was washed with 10 mL of equilibration buffer, then with 10 mL of water, and eluted with 1 mL of 50% methanol. The eluate was concentrated by vacuum centrifugation to about 10 μL and analyzed by MALDI-MS.

2.3.4.2 Gel-Based Method

Aliquots of the irradiated crosslinking mixture, typically 200 μL , were mixed with 50 μL of 5x cracking buffer, and SDS-polyacrylamide electrophoresis was carried out on Gibco-BRL gels. After removing the stacking gel, the lower 3-

cm portion of the resolving gel corresponding to the region of free dT_{20} was removed with a razor blade and the remaining portion of gel was stained with copper. Bands corresponding to $Ung \times dT_{20}$ were cut from the gel with a clean, sharp razorblade and placed into 1.5-ml Eppendorff tubes. Gel slices were washed by adding a volume of destain solution (Bio-Rad) to cover the gel pieces. The tubes were vortexed for 15 min, the solution in each tube was removed, and the gel pieces were washed with fresh destain solution three more times. At this time, the gel slices were considered "destained". In-gel digestion (Section 2.2.9) was then performed on the gel-slices with one modification. After 2 h of digestion at 37 °C, 400 μ L 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 μ L 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 μ L TE-100. Both the adjusted aqueous digestion solution and the resuspended organic extraction

solution were applied to a NAC-52 (Gibco-BRL) anion exchange cartridge 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 $\times$ dT₂₀ fragments were eluted with 1 mL TE-1000 directly into a Centricon-3 (Amicon). The Centricon-3 was centrifuged in an SS-34 rotor at 5500 rpm. 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 μ L and then transferred to a 750- μ L Eppendorff tube and further concentrated to about 10 μ L by vacuum centrifugation.

2.4 Experimental Procedures for Sulfonic Acid Derivatization

All derivatizing steps were conducted in 500 μ L Eppendorff tubes and at ambient temperature. Peptide solutions (5 μ L) were dried under a stream of nitrogen (approximately 1 h) at which time 20 μ L of THF:diisopropylethylamine (4:1) and 2 μ L of THF:chlorosulfonylacetyl chloride (100:1) were added to the tube. The reaction mixture was vortexed for 1 min at which time 1 μ L of water was added and the mixture was again vortexed for an additional 1 min. The reaction mixture was then thoroughly dried under a stream of nitrogen to remove the volatile reagents. The derivatized sample was resuspended in water and stored at -20 °C until further use.

2.5 Experimental Procedures for hRPA Project

2.5.1 Overexpression of hRPA

2.5.1.1 Mini Cultures

The plasmid p11d-tRPA, which contains a single T7 RNA polymerase promoter followed by the coding sequences for hRPA70, hRPA14, and hRPA32, was obtained as a gift from Dr. Marc Wold (University of Iowa). This plasmid was electroporated into *E. coli* strain BLR21(DE3) and plated on LB plates (1% tryptone, 0.5% yeast extract, 1% NaCl, and 1.5% agar) containing 100 µg/ml ampicillin. A single colony was divided in half with a toothpick and each half was used to inoculate two separate flasks containing 20 mL LB medium and 100 µg/ml ampicillin. The contents of both flasks were grown at 37 °C with shaking to an O.D.₆₀₀ of 0.6 at which time one was induced with IPTG to 0.4 mM for 2 h and the other was uninduced. A fraction (5 mL) from each flask was removed and centrifuged for 10 min using a SS-34 rotor at 5000 × g and 4 °C. The supernatant was removed from each tube, the cells were resuspended in 10% (w/v) glycerol, then centrifuged again for 10 min using a SS-34 rotor at 5000 × g and 4 °C. Pellets in each tube were resuspended in 250 µL of HI buffer (30 mM HEPES-KOH (pH 7.8), 1 mM dithiothreitol, 0.25 mM EDTA, 0.25% (w/v) inositol, and 0.01% (v/v) Nonidet P-40), SDS was added to a final concentration of 0.1% to each tube, and the tubes were heated in boiling water for 10 min. The soluble protein (~ 2 mg/mL) from the induced and the uninduced fractions was analyzed on a denaturing 12.5% polyacrylamide gel.

2.5.1.2 Preparative Cultures

The plasmid p11d-tRPA, along with the plasmid pRIL, were electroporated into *E. coli* strain BLR21(DE3) and plated on LB (1% tryptone, 0.5% yeast extract, 1% NaCl, 1.5% agar) plates 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 KH₂PO₄, and 0.07 M K₂HPO₄) 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 and 275 rpm. 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 in a GSA rotor and frozen at -80 °C until used.

2.5.2 Purification of hRPA

All purification steps were carried out at 4 °C. Frozen cells (from 1 L of culture) were resuspended in 50 mL of HI buffer containing 50 mM KCl, 1 mM phenylmethylsulfonyl fluoride, 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 in a SS-34 rotor. Soluble protein was applied to a 20-mL Affi-Gel Blue Gel column (12 cm × 1.77 cm², BioRad) equilibrated in HI buffer containing 50 mM KCl. The column was washed sequentially with 3 column volumes each of

HI buffer containing 50 mM KCl, 0.8 M KCl, 0.5 M NaSCN, 1.0 M NaSCN, and 1.5 M NaSCN. The peak of elution was determined by denaturing polyacrylamide gel electrophoresis, visualizing protein bands with Coomassie stain. Fractions containing the elution peak were pooled into 12,000-14,000 molecular weight cutoff tubing 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 cm²) 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 from 500 mM NaCl to 3 M NaCl. The peak of elution, also monitored by denaturing polyacrylamide gel electrophoresis, was pooled and concentrated by using a Centriprep-30 (Millipore) and the buffer was exchanged to HI buffer containing 200 mM KCl.

2.5.3 Protein Concentration Determinations

All protein concentrations were estimated by the Bradford assay (reagent from Bio-Rad) using BSA as the standard (16).

2.5.4 Electrophoretic Mobility Shift Assay

Varying amounts of hRPA (1 pmol to 50 pmol) were mixed with 10 pmol each of [^{32}P] 5'-end labeled oligonucleotide dT₃₀, each solution was adjusted to 20 mM HEPES-KOH (pH 7.0), 1 mM DTT, and 50 $\mu\text{g/mL}$ BSA, and incubated at room temperature for 30 min. Sucrose was added to each reaction mixture to a final concentration of 12% (final volume of 26 μL). The samples were loaded onto a 5% native polyacrylamide gel (BRL, 17cm $\times$ 15 cm $\times$ 0.5 cm), with TAE as the running buffer, and electrophoresis was carried out at room temperature and 100 V. The gel was then dried on DE81 (Whatman) backing paper and autoradiographed using X-OMAT film (Kodak). Quantitation of ^{32}P in the gel was conducted using a PhosphorImager (Molecular Dynamics) with ImageQuant 5.0 software (Molecular Dynamics).

2.5.5 UV-Catalyzed Crosslinking of hRPA

Samples of hRPA and oligonucleotide were diluted with HI buffer containing 50 mM KCl in Eppendorf tubes and allowed to equilibrate at room temperature for 30 min. Preparative samples (1 mL) were placed in a capped, 4-mL, quartz cuvette (NSG Precision Cells) or small volume samples (less than 200 μL) were placed in the cap from a 4-mL quartz cuvette (NSG Precision Cells). Crosslinking ($\lambda_{\text{max}}=254\text{nm}$) was conducted for the appropriate time in a Stratalinker 1800 (see Chapter 5, Section 6.1.5).

3. PURIFICATION OF NUCLEOPEPTIDES FROM THE UV-CROSSLINKED UNG × dT₂₀ NUCLEOPROTEIN COMPLEX FOR CHARACTERIZATION BY MALDI MS

The application of MALDI MS to characterize nucleoprotein and nucleopeptide complexes generated from UV-catalyzed crosslinking reactions has been demonstrated as a promising technique for studying protein-nucleic acid interactions. The main goal of this thesis was to extend the DNA-binding characterization of Ung using mass spectrometry by identifying specific amino acids within the protein taking part in crosslinking. In order to continue this work, sufficient quantities of isolated nucleopeptides were required for tentative identification by MALDI MS and definitive identification by ESI MS/MS. This means that the isolated nucleopeptide mixture must be sufficiently concentrated (~10 µM) and free from contaminants such as salts and surfactants. This chapter describes two procedures, the "in-pot" and "in-gel" procedures, that were used to isolate and tentatively identify nucleopeptides by MALDI MS analysis.

3.1 Results

3.1.1 "In-pot" Purification of Ung Nucleopeptides

3.1.1.1 Crosslinking of Ung to dT₂₀

In order to obtain sufficient quantities of nucleopeptides from the Ung × dT₂₀ nucleoprotein complex for mass analysis, a dose response experiment was performed to determine the optimum time of UV irradiation. This experiment consisted of mixing the crosslinking components of dT₂₀ and Ung in a 3:1 molar ratio (the molar ratio that would be used in future preparative crosslinking experiments) and then irradiating the mixture. At increasing times of irradiation, fractions were removed and analyzed by denaturing polyacrylamide gel electrophoresis with the protein bands being visualized by Coomassie staining (Figure 3.1). From this gel, one band corresponding to free Ung ($M_r \sim 25,000$) and the other band migrating with an apparent molecular weight of $\sim 32,000$ were identified. The molecular weights of Ung and dT₂₀ were determined by MALDI MS (Figure 3.2) to be 25,561 and 6,023, respectively, and therefore the predicted molecular weight of dT₂₀ covalently attached to Ung would be 31,584. The gel band of $M_r \sim 32,000$ (Figure 3.1) increased in intensity with increasing time of UV-irradiation. The increased intensity along with its estimated molecular weight suggested that the band corresponded to Ung crosslinked to dT₂₀. The optimum crosslinking time was determined to be that which produced the largest, well-defined, mobility-shifted band with minimum background signals. By using this criterion, the optimum time was selected to be 45 min (Figure 3.1, lane 7).

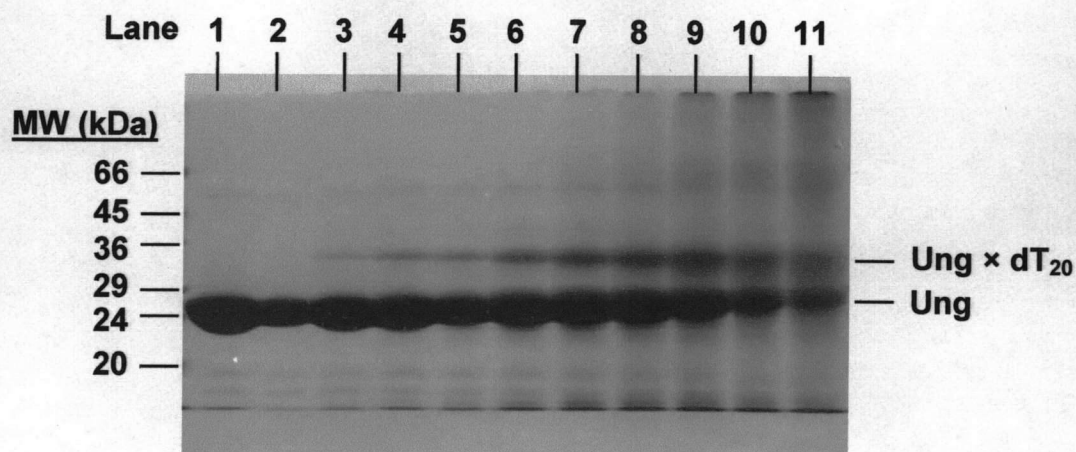


Figure 3.1. Effect of 254 nm irradiation time on the crosslinking of Ung to dT₂₀. A 1500- μ L reaction mixture containing 45 nmol of Ung and 134 nmol of dT₂₀ in buffer DAB was irradiated at 254 nm, as described under "Experimental Procedures." At the times indicated below, 50- μ L samples were removed and analyzed on a denaturing 12.5% polyacrylamide gel. Following electrophoresis, the gel was stained with Coomassie Brilliant Blue G-250 to visualize the protein bands. *Lane 1* corresponds to 3.3 nmol of unirradiated Ung and *lanes 2-11* correspond to 254 nm exposure times of 0, 5, 10, 20, 30, 45, 60, 90, 120, and 150 minutes, respectively. The location of molecular weight standards (kDa) are indicated to the left of the figure and the location of Ung and Ung $\times$ dT₂₀ are indicated to the right of the figure.

Figure 3.2. MALDI mass spectra of Ung and dT₂₀. *A*, a solution of 36 μ M of Ung was mixed with the matrix sinapinic acid in a 1:3 ratio of protein to matrix and analyzed. Matrix adduction to the protein is observed with ion species m/z 25,786 (mass sinapinic acid is 224 Da). *B*, a solution of 108 μ M of dT₂₀ was mixed with the maxtrix THAP in a 1:5 ratio of oligonucleotide to matrix and analyzed. Details of the MALDI analysis are described in more detail under "Experimental Procedures."

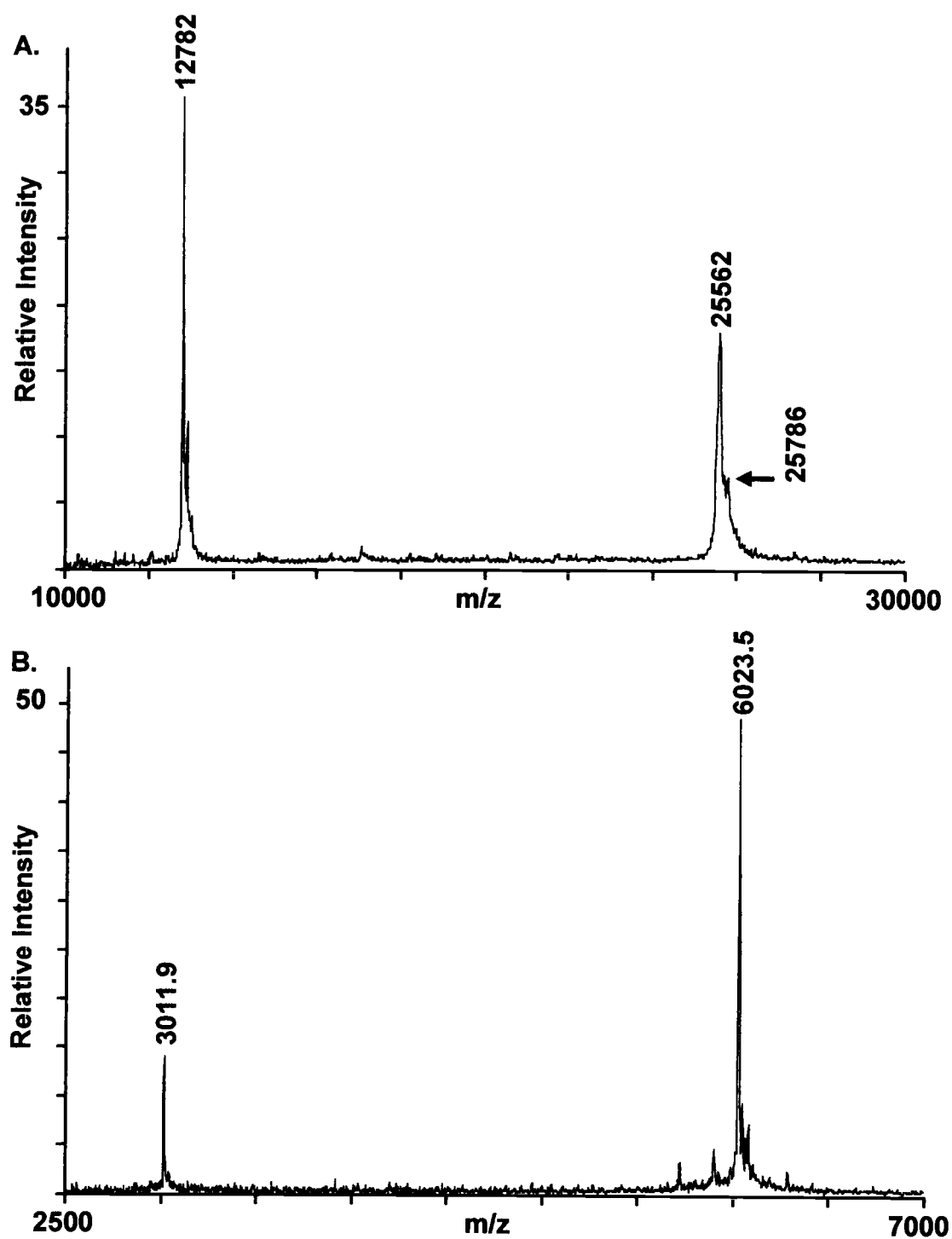


Figure 3.2. MALDI mass spectra of Ung and dT₂₀.

3.1.1.2 Isolation and MALDI Analysis of Ung Nucleopeptides from the Ung × dT₂₀ Nucleoprotein Complex

Preparative-scale crosslinking was performed with 180 nmol of Ung and 560 nmol of dT₂₀ to isolate and analyze by mass spectrometry Ung tryptic peptides crosslinked to dT₂₀. An "in-pot" purification scheme (Figure 3.3) was followed to isolate nucleopeptides from the nucleoprotein complex. In this scheme, the UV-irradiated mixture was treated with trichloroacetic acid to precipitate the protein. The precipitated protein was then suspended in 8 M urea, proteolytically digested, and peptide × dT₂₀ complexes were separated from non-crosslinked peptides by a DEAE-Sephadex column (Figure 3.4). The elution was monitored spectrophotometrically at 260 nm and 280 nm. A peak between fractions 150 and 180 had an elevated A_{260}/A_{280} absorbance ratio and eluted at a salt concentration of 460 mM (Figure 3.4). It was reasoned that elution peaks containing peptides would have A_{260}/A_{280} ratios of less than one and peaks containing nucleopeptides would have A_{260}/A_{280} ratios greater than one. Therefore, fractions 156 through 178 were pooled, desalted, concentrated, and then analyzed by MALDI MS (Figure 3.5). Five prominent ion species were observed in the m/z range of 6,000 to 9,500. The ion species at 6,025 corresponded to dT₂₀ while the ion species 7,187.4, 7,608.1, 8,078.7, and 9,405.2 had masses that were in the range of Ung peptides crosslinked to dT₂₀. Further analysis of this spectrum was not pursued since the signal to background ratio was very low and the peak widths were very large making it difficult to

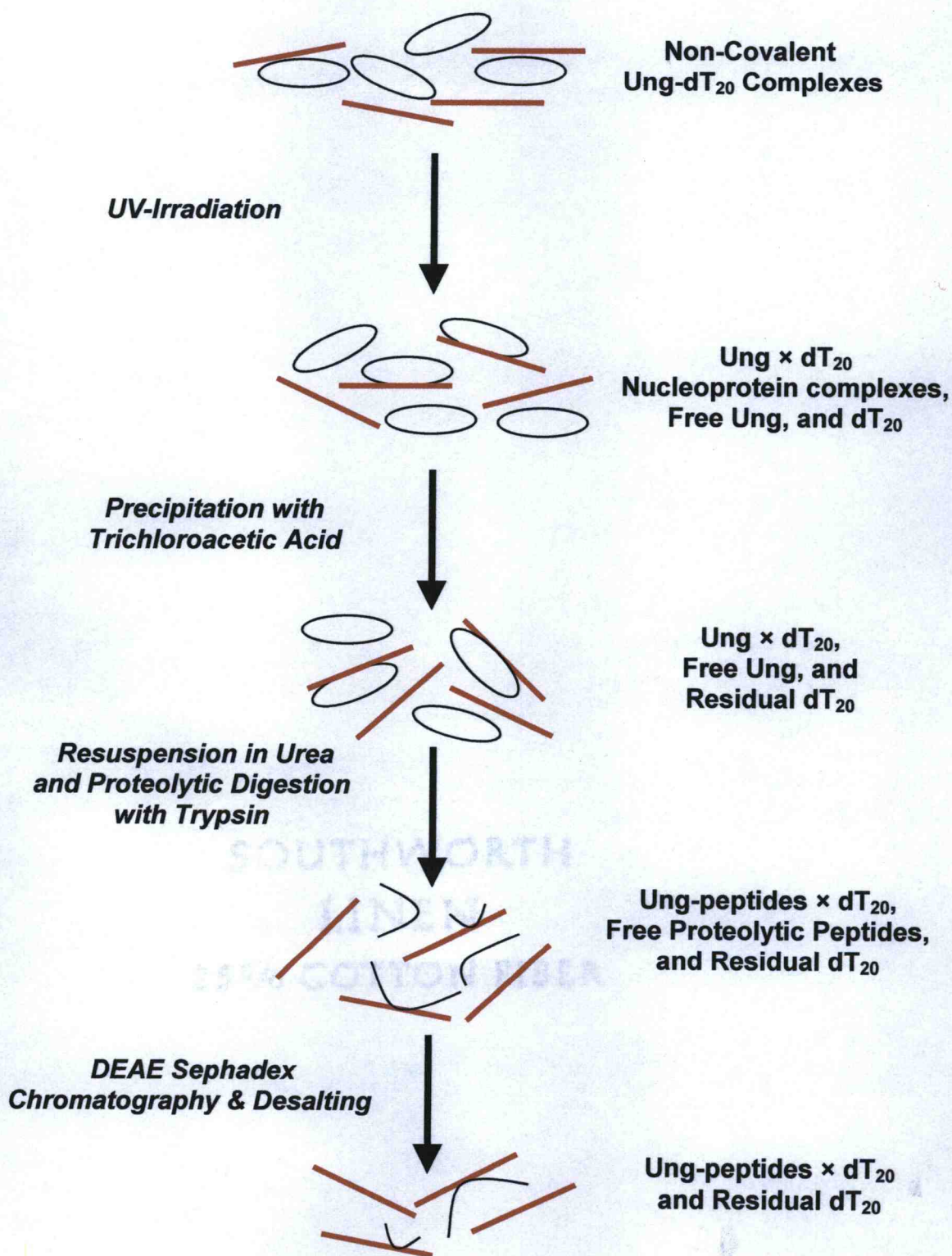


Figure 3.3. "In-pot" purification scheme for isoalting Ung-nucleopeptides from Ung $\times$ dT₂₀ nucleoprotein complexes.

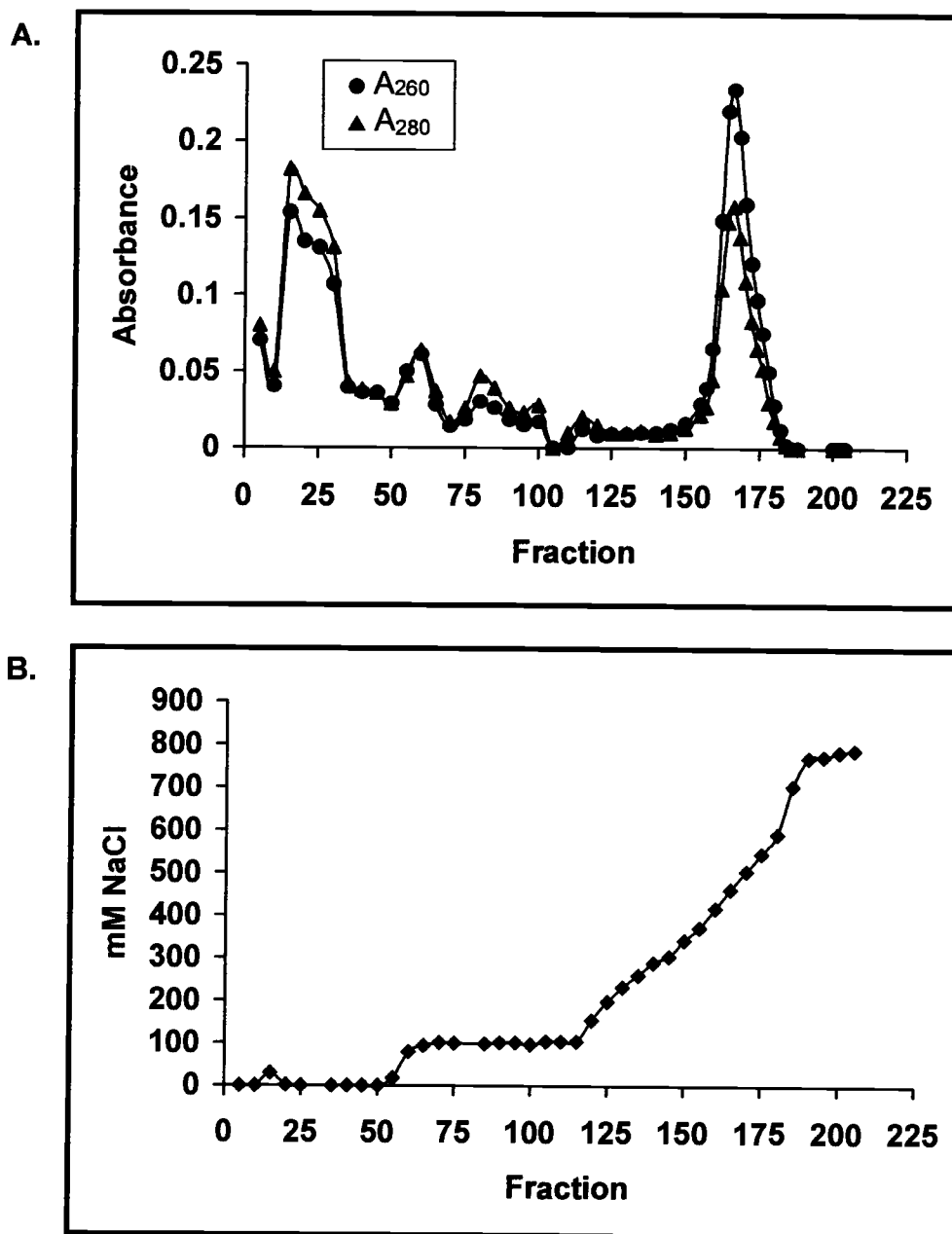


Figure 3.4. Anion exchange chromatography elution profile of Ung tryptic nucleopeptides. Six standard crosslinking mixtures (1 ml each) were UV irradiated, pooled, and treated using the “in-pot” analytical scheme as described under “Experimental Procedures.” Fractions collected from a DEAE-Sephadex A-50 column were monitored for absorbance at 260 nm (**A**) and 280 nm (**A**) and for conductivity (**B**).

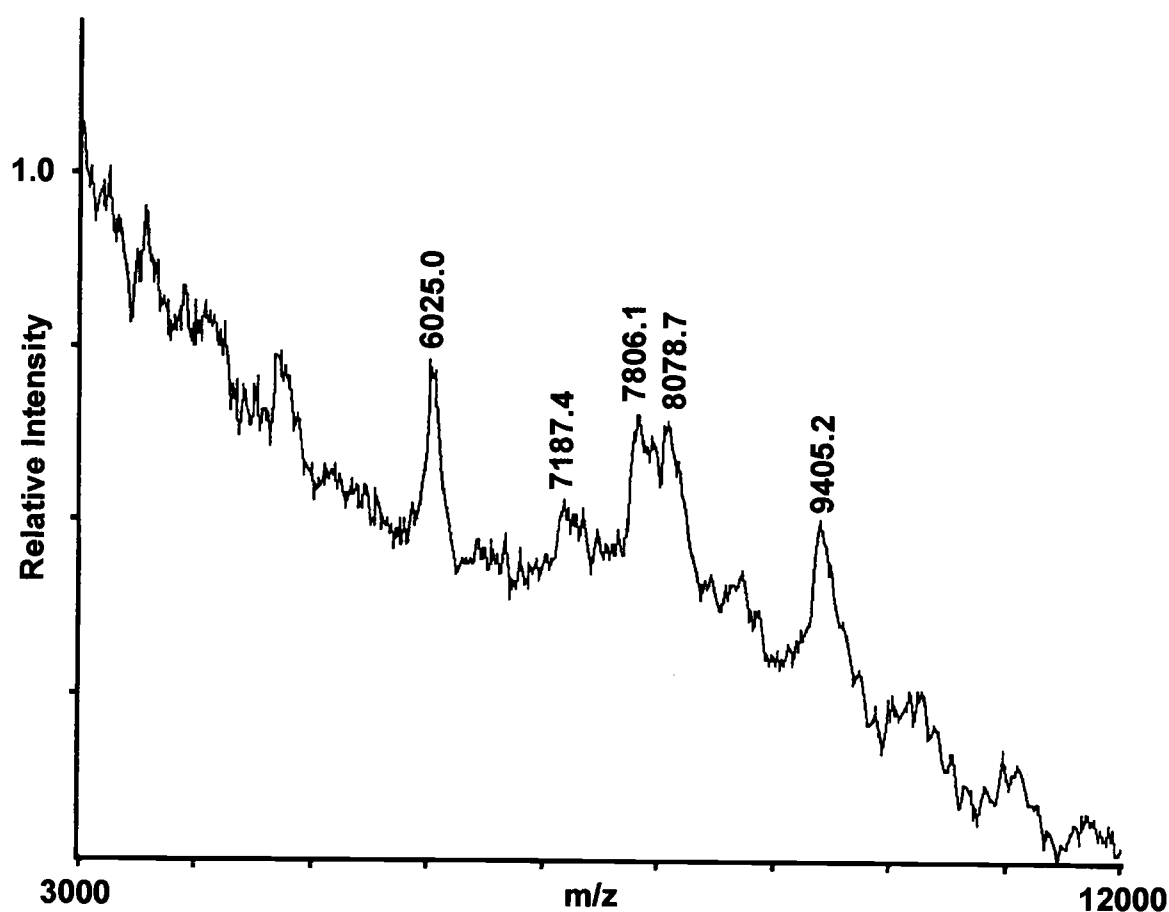


Figure 3.5. MALDI mass spectrum of purified Ung $\times$ dT₂₀ tryptic peptides isolated using the “in-pot” purification” scheme. Fractions 156-178 from the DEAE-Sephadex column (Figure 3.4) were pooled, desalted and concentrated by using a Nensorb-20 column, and then analyzed by MALDI-MS, as described under “Experimental Procedures.”

assign accurate masses. It was also found that reproducing this spectrum from samples produced via the “in-pot” separation procedure was challenging since most attempts yielded only ion signals of dT_{20} . The dT_{20} ion signal might have been a result of the oligonucleotide co-precipitating with Ung and Ung $\times$ dT_{20} during treatment with trichloroacetic acid. Any dT_{20} present at this stage of the purification scheme would have co-purified when attempting to isolate nucleopeptides. A disproportionate ratio of dT_{20} to nucleopeptides (higher concentration of dT_{20}) could have interfered with obtaining spectra of isolated nucleopeptides.

3.1.2 Gel-Based Purification of Ung Nucleopeptides

3.1.2.1 Crosslinking of Ung to dT_{20}

As a result of not being able to obtain sufficient quantities of nucleopeptide material using the “in pot” procedure, it was decided to reevaluate the effect of UV-light on the crosslinking reaction yield and its effects on the starting components of the reaction. The effect of UV light on the formation of the Ung $\times$ dT_{20} product was reevaluated by irradiating a mixture of 20 nmol of Ung and 40 nmol of $[^{32}P]dT_{20}$ for increasing times. Analysis of the reaction products on a SDS-polyacrylamide gel (Figure 3.6A) revealed a mobility-shifted band (relative to free Ung) of $M_r \sim 33,000$ that was consistent with the predicted molecular weight of Ung $\times$ $[^{32}P]dT_{20}$ (M.W. 31,664). Autoradiography showed that the mobility shifted band contained $[^{32}P]dT_{20}$ and corresponded to the covalent Ung $\times$

Figure 3.6. Effect of 254 nm irradiation time on the crosslinking of Ung to [^{32}P]dT₂₀. A 635- μl reaction mixture containing 20 nmol of Ung and 40 nmol of [^{32}P]dT₂₀ in buffer DAB was irradiated at 254 nm, as described under "Experimental Procedures." At the times indicated, 25- μl samples were removed and analyzed on a denaturing 12.5% polyacrylamide gel. Following electrophoresis, the gel was stained with silver (**A**), dried, and subjected to autoradiography (**B**). *Lane 1* contains 200 pmol of unirradiated Ung and *lanes 2-11* correspond to 254-nm exposure times of 0, 5, 10, 20, 30, 45, 60, 90, 120, and 180 minutes, respectively. The location of molecular weight standards (in kDa) and free Ung (*U*) are indicated to the left of *panel A* and the location of Ung $\times$ [^{32}P]dT₂₀ is indicated by an *asterisk* to the right of *panel B*. The amount of Ung $\times$ [^{32}P]dT₂₀ was quantified by cutting the bands from the dried gel and measuring the ^{32}P radioactivity by liquid scintillation counting (**C**).

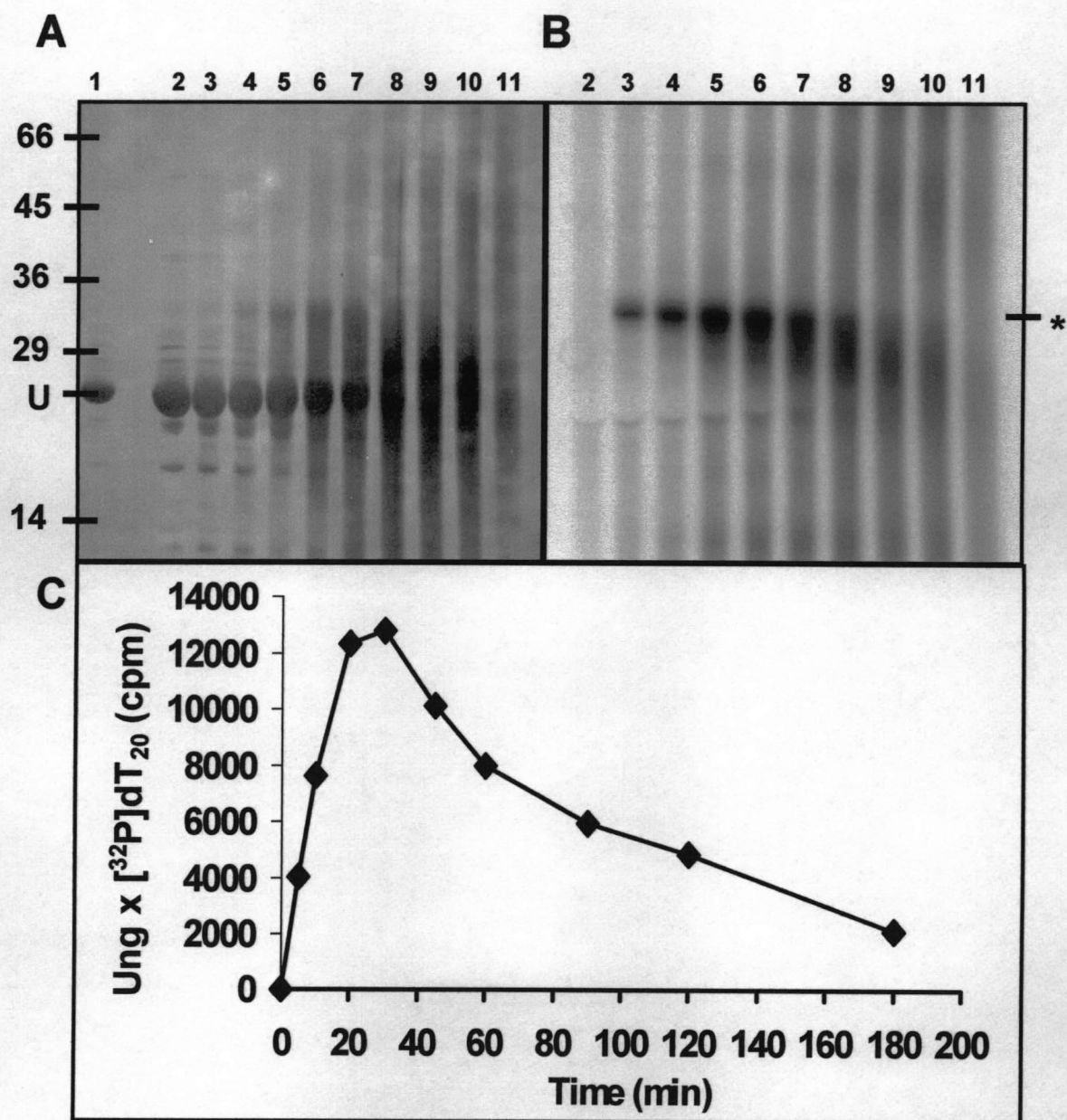


Figure 3.6. Effect of 254 nm irradiation time on the crosslinking of Ung to [^{32}P]dT₂₀.

$[^{32}\text{P}]\text{dT}_{20}$ nucleoprotein complex (Figure 3.6B). From measuring the amount of $\text{Ung} \times [^{32}\text{P}]\text{dT}_{20}$ in the crosslinked bands, the maximum crosslinking yield was determined to take place between 20 and 30 min of irradiation time before a reduction in the crosslinking product took place (Figure 3.6C). Closer analysis of the SDS-polyacrylamide gel showed that the reduction did not seem to be the result of a reversal in the crosslinking reaction, but a result of degradation in the nucleoprotein complex occurring at increased doses. The onset of degradation was observed between 10 and 20 min with a loss in the band shape of $\text{Ung} \times [^{32}\text{P}]\text{dT}_{20}$ and an increased background of the gel, especially in the lower molecular weight range relative to $\text{Ung} \times [^{32}\text{P}]\text{dT}_{20}$ (Figure 3.6A and 3.6B).

To assess the effect of UV light on the individual components of a crosslinking reaction, Ung or dT_{20} was exposed to 254-nm radiation and characterized by MALDI MS (Figure 3.7). For both protein and oligonucleotide, increased exposure to 254-nm radiation resulted in a decrease in the signal to background ratio. The reduced signal to background ratio was more pronounced for the protein than for the DNA, especially within the first 15 min of exposure. Also within the first 15 min of exposure, a distinct degradation product of the protein was observed occurring at $m/z \sim 10,500$. The effect of UV light on the protein and the oligonucleotide was wavelength dependent since irradiation at 312 nm showed virtually no loss in the signal to background ratio of dT_{20} and only a slight reduction in that of Ung after a 45-min dose (Figure 3.8). Combining the results of the two dose experiments (Figures 3.6 and 3.7), a 254 nm crosslinking

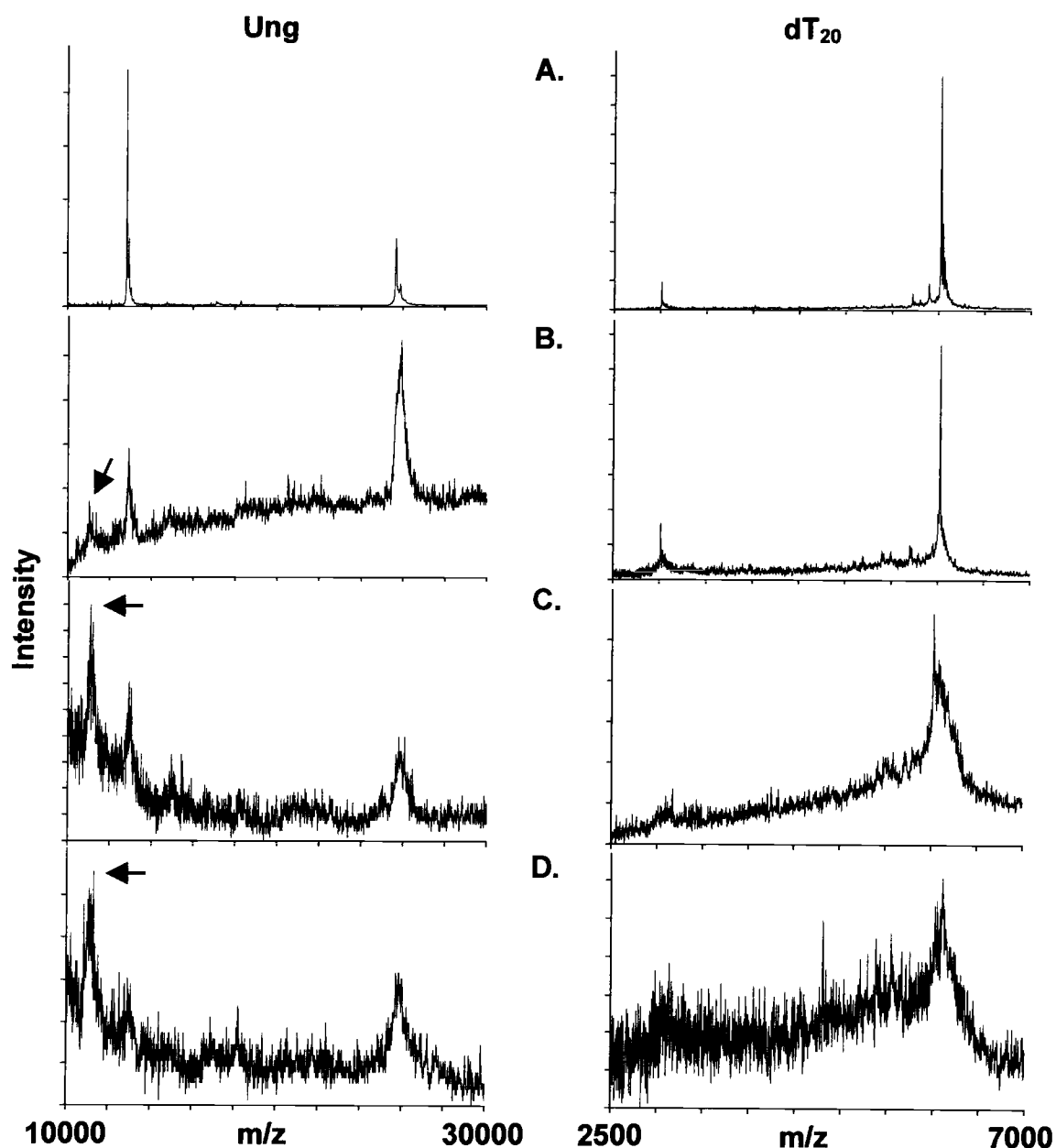


Figure 3.7. Effect of 254-nm light on MALDI-MS analysis of Ung and dT₂₀. Separate 1-mL solutions of Ung (36 μ M) or dT₂₀ (108 μ M) were exposed to 254-nm light for 0, 15, 30, or 45 min (A, B, C, and D, respectively) and analyzed by MALDI-MS, as described under "Experimental Procedures." Ung $[M+H^+] = 25,564$ and dT₂₀ $[M+H^+] = 6,024$, as measured at 0 min. An arrow indicates the position of a specific UV degradation product from Ung.

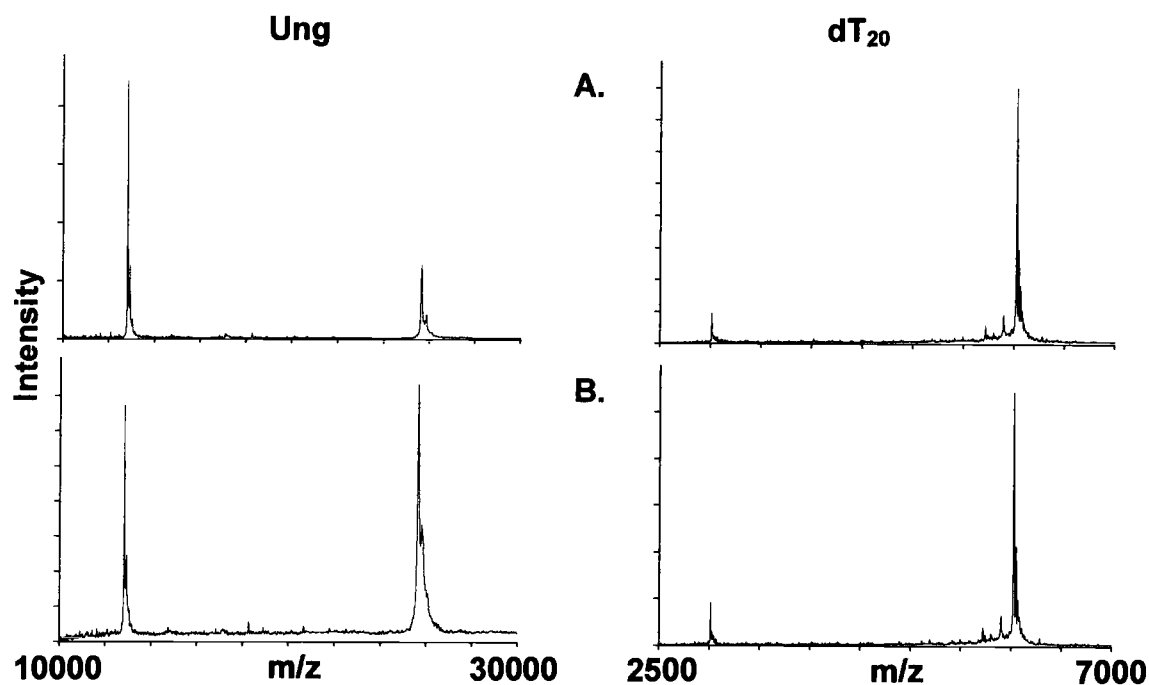


Figure 3.8. Effect of 312-nm light on MALDI MS analysis of Ung and dT₂₀. Separate 1-mL solutions of Ung (36 μ M) or dT₂₀ (108 μ M) were exposed to 312-nm light for 0 or 45 min (**A** and **B**, respectively) and analyzed by MALDI MS as described under "Experimental Procedures." Ung $[M+H^+] = 25,564$ and dT₂₀ $[M+H^+] = 6,024$, as measured at 0 min.

time of 15 min was used in preparative crosslinking experiments to obtain significant nucleoprotein complex yet minimize degradation to the nucleoprotein complex.

3.1.2.2 Isolation of Ung Nucleopeptides from the Ung $\times$ dT₂₀ Nucleoprotein Complex

A novel purification scheme was developed (Figure 3.9) to isolate sufficient quantities of Ung nucleopeptides for analysis by MALDI MS. In this scheme, the nucleoprotein complex was isolated from noncrosslinked Ung and dT₂₀ on a denaturing polyacrylamide gel, as a mobility shifted band, that was then excised and digested "in-gel" with trypsin. The products of the proteolytic digestion (tryptic peptides and tryptic peptides $\times$ dT₂₀) were extracted from the gel slice by using an organic extraction mixture and the extracted nucleopeptides were isolated using an anion exchange cartridge. After thorough desalting and concentration, the isolated peptides $\times$ dT₂₀ (termed nucleopeptides) were suitable for analysis by MALDI MS.

3.1.2.3 Tentative Identification of Peptide $\times$ dT₂₀ Complexes

Prior to the analysis of isolated Ung nucleopeptides, a tryptic peptide MALDI map of Ung was produced (Figure 3.10) to use as a guide for tentatively identifying peptide regions of the protein involved in crosslinking. In the map, 92% of the amino acid sequence was covered. The only regions of the protein

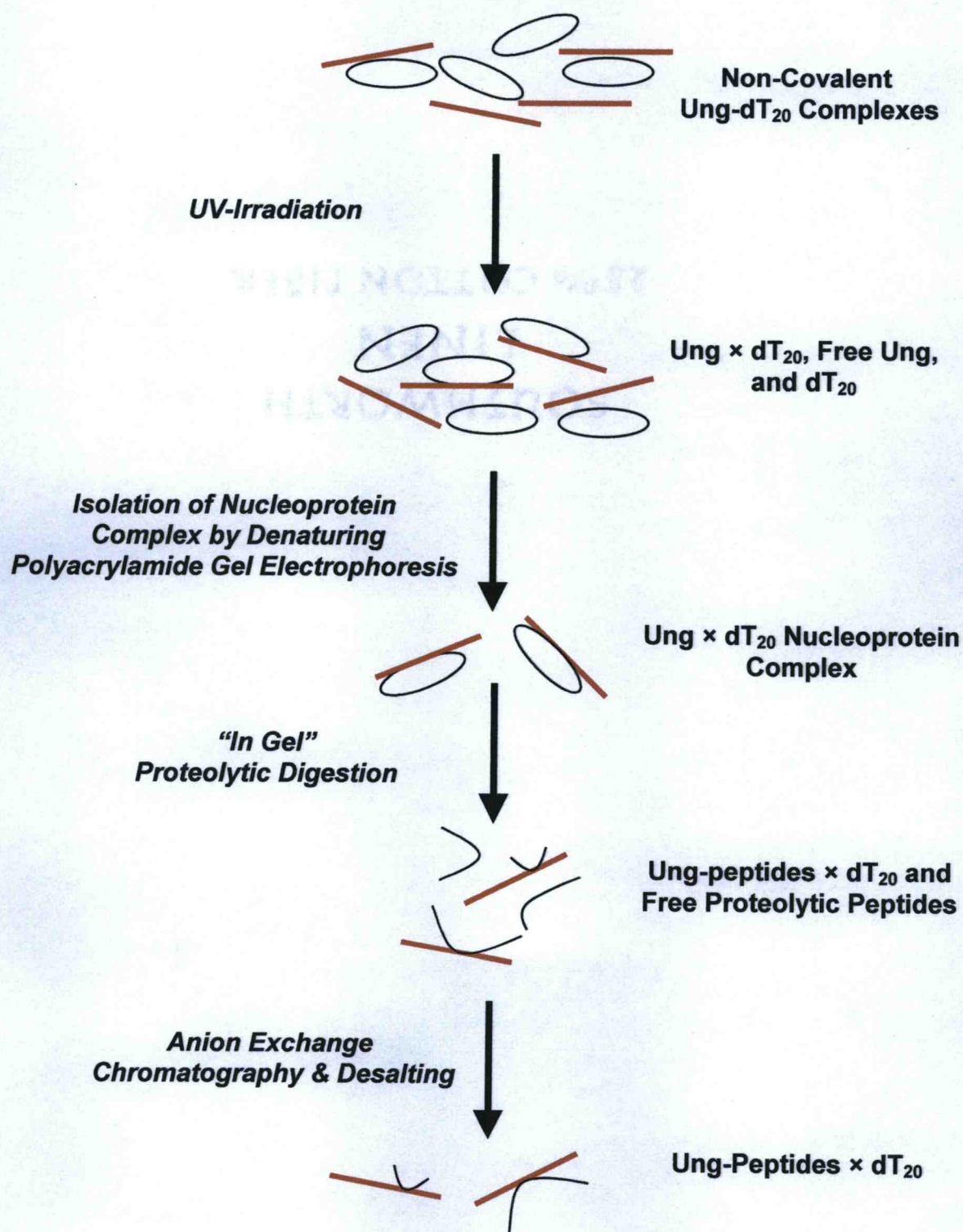


Figure 3.9. In-gel purification scheme for isolating nucleopeptides from nucleoprotein complexes.

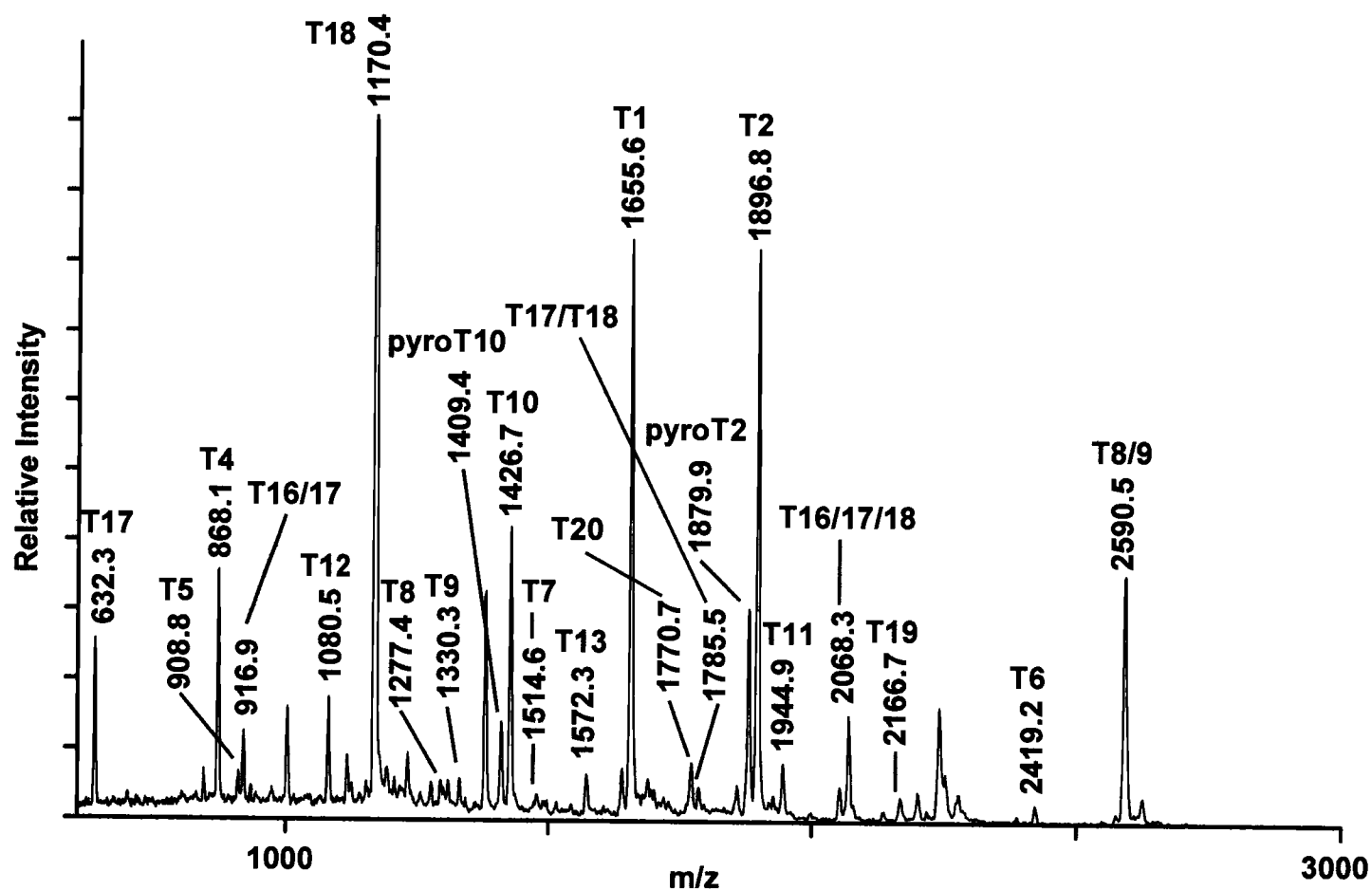


Figure 3.10. MALDI tryptic map of *E. coli* Ung. A band corresponding to Ung was cut from a SDS polyacrylamide gel and digested in-gel, as described under "Experimental Procedures." MALDI MS was conducted on the digestion products to produce a peptide map of the protein. The m/z range displayed is from 600 to 3000.

absent in the map were those spanning tryptic fragments T3, T14, T15, and T16 (Table 3.1). Fragments T14 and T15 were not expected to be observed in the map due to their small molecular weights, but T16 is accounted for in the incomplete digestion product T16/T17.

MALDI MS analysis of purified tryptic peptide $\times$ dT₂₀ complexes revealed eight distinct ion species (Figure 3.11), one of which corresponded to free dT₂₀ (m/z 6024.4). Subtracting the molecular weight of dT₂₀ from the seven other experimentally measured peptide $\times$ dT₂₀ complexes resulted in the calculated peptide molecular weight portion of the nucleopeptides. Comparison of the calculated peptide masses to the tryptic digestion map of the protein (Figure 3.10, Table 3.1) tentatively identified four of the peptide portions of ion species, 7191.7, 7352.8, 7963.9, and 8442.4 as T18 (185-APHPSPLSAHR-195), T11 (130-AGQAHSASLGWETFTDK-147), T9 (106-PNHGYLESWAR-116), and T6 (58-VVILGQDPYHGPGQAHLAFSVR-80), respectively. Ion species 8096.6, 7805.5, and 7711.7 did not directly correspond to the sum of a tryptic peptide and dT₂₀. The calculated peptide masses for these ion species were used in a mass search of the amino acid sequence of Ung. This indicated that ion species 8096.6 and 7711.7 were hydrolysis products of T6 $\times$ dT₂₀ in which three or seven C-terminal amino acids, respectively, had been lost during sample preparation. These truncated forms of T6 are hereafter designated as T6₍₋₃₎ and T6₍₋₇₎. The ion species 7805.5 was tentatively identified as T17/T18 (180-HHVLKAPHPSPLSAHR-195) covalently attached to dT₂₀. Collectively, these

Table 3.1. Tentative identification of Ung tryptic nucleopeptides measured by MADLI-MS.

Peptide	Predicted Mass ^a	Experimental Mass		Calculated Mass ^d Ung × dT ₂₀ Peptide – dT ₂₀
		Ung Peptide Mass ^b	Ung × dT ₂₀ Peptide ^c	
met-T1 ^e	1786.0			
T1	1654.8	1654.6		
T2	1895.1	1895.8		
pyroT2 ^f	1878.1	1878.9		
T3	1217.4	N.D. ^g		
T4	868.0	867.1		
T5	908.0	907.8		
T6	2418.7	2419.2	8441.2	2418.1
T6 ₍₋₃₎ ^h	2075.1	N.D.	8095.6	2072.5
T6 ₍₋₇₎ ⁱ	1686.9	N.D.	7710.7	1687.6
T7	1513.9	1513.6		
T8	1276.4	1276.4		
T8/T9	2587.8	2589.5		
T9	1329.4	1329.3	7351.8	1328.7
T10	1425.7	1425.7		
pyroT10 ^f	1408.7	1408.4		
T11	1943.1	1943.9	7962.9	1939.8
T12	1079.3	1079.5		
T13	1570.8	1571.3		
T14	146.2	N.D.		
T15	615.7	N.D.		
T16	302.3	N.D.		
T17	632.8	631.3		
T16/T17	917.1	915.9		

Table 3.1 (Continued)

Peptide	Predicted Mass ^a	Experimental Mass		Calculated Mass ^d Ung × dT ₂₀ Peptide – dT ₂₀
		Ung Peptide ^b	Ung × dT ₂₀ Peptide ^c	
T16/17/18	2067.1	2067.3		
T17/T18	1784.1	1784.5	7805.5	1782.4
T18	1169.3	1169.4	7190.7	1167.6
T19	2165.4	2165.7		
T20	1771.0	1769.7		
dT ₂₀	6023.1	6023.1		

^a Predicted average masses were calculated from a predicted tryptic digestion map of *E. coli* Ung using the computer program GPMW, version 3.13, Lighthouse Data.

^b Experimental Ung tryptic peptide masses were calculated from Figure 3.10.

^c Experimental Ung peptides × dT₂₀ were calculated from Figure 3.11.

^d The molecular weight of crosslinked peptides were calculated by subtracting the mass of dT₂₀ from the experimentally measured peptide × dT₂₀ masses.

^e met-T1 contains the N-terminal methionine that is post-translationally removed *in vivo* (138).

^f Peptides T2 and T10 contain N-terminal glutamine that can lose NH₃ to form pyroglutamic acid.

^g N.D. refers to not detected.

^{h, i} Represents truncated forms of the T6 nucleopeptide that have lost 3 or 7 C-terminal amino acids, respectively.

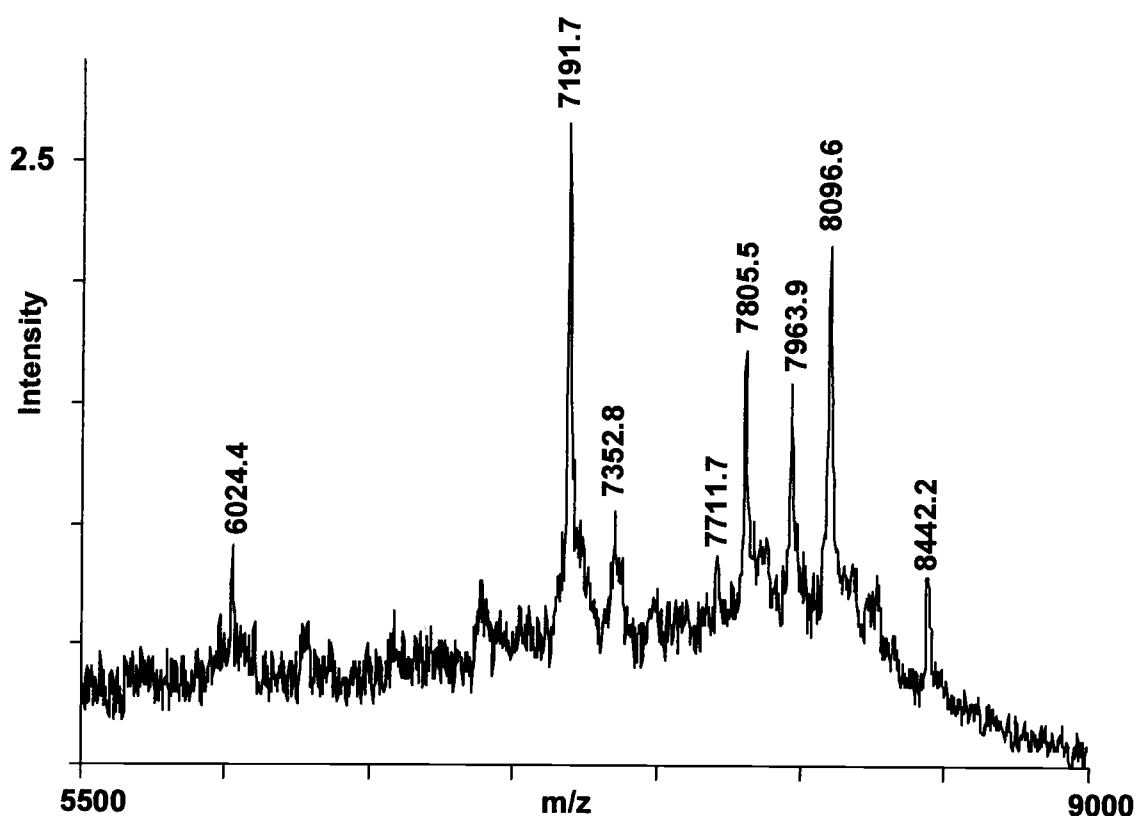


Figure 3.11. MALDI mass spectrum of purified Ung $\times$ dT₂₀ tryptic peptides.

A 1-mL reaction mixture containing 36 nmol of Ung and 108 nmol of dT₂₀ was crosslinked and nucleopeptides were purified by using the in-gel purification procedure. Tryptic peptides crosslinked to dT₂₀ were eluted from a NAC-52 anion exchange cartridge, concentrated and desalted with a Centricon-3, and further concentrated by vacuum centrifugation to approximately 10 μ L as described under "Experimental Procedures." The ion signal corresponding to free dT₂₀ is observed at m/z 6024.4. The remaining ion signals correspond to Ung tryptic peptides covalently attached to dT₂₀.

results suggest that amino acids within regions 58-80, 106-116, 130-147, and 180-195 of the protein are crosslinked to dT₂₀ (Figure 3.12).

3.2 Discussion

The ultimate goal of this work was to obtain sufficient quantities of nucleopeptides from UV-catalyzed Ung × dT₂₀ nucleoprotein complexes for mass spectral analysis. The important step in reaching this goal was establishing an optimum irradiation time to minimize protein and oligonucleotide damage. The hazards of over irradiating crosslinking samples was well-understood (37, 73, 146). However, the analytical method that would best determine the optimum crosslinking time was not. Two optimization experiments were performed. The first experiment used Coomassie blue staining of a polyacrylamide gel to monitor the formation and destruction of the Ung × dT₂₀ complex with increasing UV doses (Figure 3.1). The second experiment used silver staining and autoradiography to monitor the formation and destruction of the Ung × [³²P]dT₂₀ complex with increasing UV doses (Figure 3.6). It was found that the Coomassie blue visualization of the crosslinking products estimated the optimum crosslinking time to be 45 min (Figure 3.1, lane 7) while visualization by silver staining and autoradiography estimated the optimum time to be 20 min (Figure 3.6, lane 4). An additional experiment was performed in which Ung and dT₂₀ were independently exposed to UV light for increased doses and monitored by MALDI MS (Figure 3.7). This experiment showed that doses above 15 min were severely detrimental to the protein, and to a lesser extent the oligonucleotide.

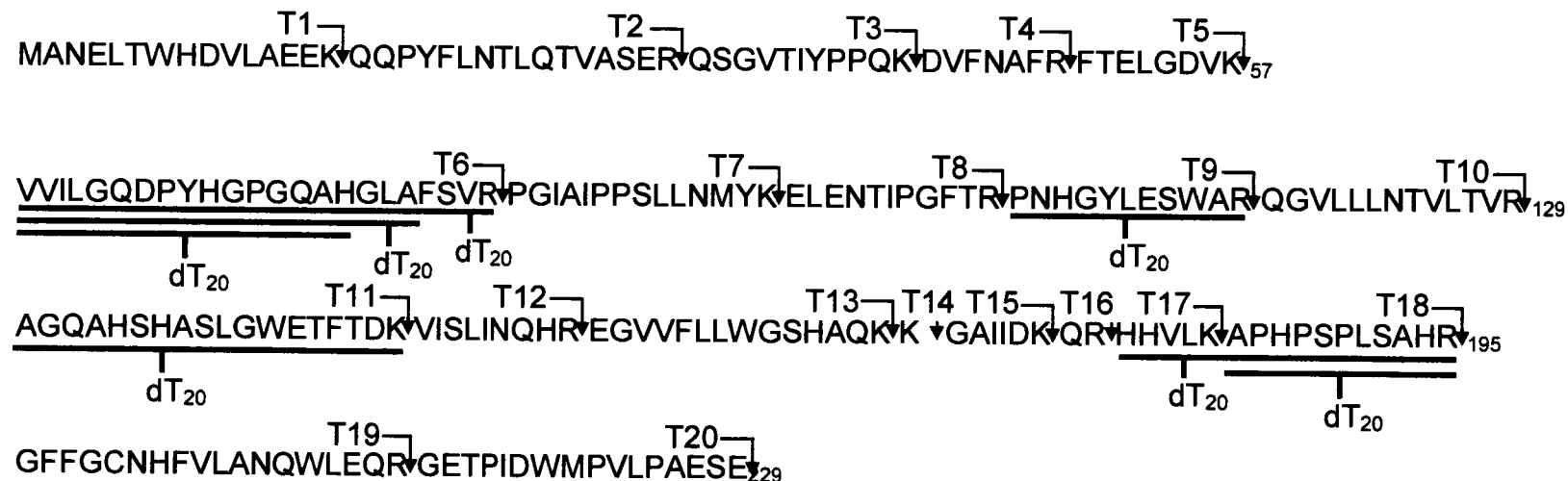


Figure 3.12. *E. coli* uracil-DNA glycosylase peptides tentatively identified by MALDI MS as being crosslinked to dT₂₀. Arrows indicate the possible sites of cleavage by trypsin and the corresponding peptide fragments are numbered (T1-T20). Regions of the protein that were identified by MALDI peptide mapping as being crosslinked to dT₂₀ are underlined.

This corroborated the optimum crosslinking time of 20 min as determined by the silver staining and autoradiography method. The use of Coomassie blue stain in these types of experiments, which is about 10 times less sensitive than silver stain (86, 95), grossly overestimated the optimum crosslinking time.

Since previous work by Bennett *et al.* (6) had successfully isolated tryptic peptides from the UV-catalyzed Ung $\times$ dT₂₀ complex for subsequent peptide sequencing and MALDI MS analysis, the method in that study ("in-pot" method) was initially followed. Multiple attempts to isolate Ung nucleopeptides using the "in-pot" method showed two characteristic types of spectra. The first were spectra with very weak signals from the nucleopeptides (Figure 3.5). These spectra made mass assignments difficult and the amount of sample available for further mass analysis (ESI MS/MS) was insufficient. The second type of spectra were those that only produced strong signals from dT₂₀ (data not shown). Based on the purification scheme it was suspected that during TCA precipitation a significant amount of dT₂₀ precipitated with the protein and this residual dT₂₀ co-purified with the nucleopeptides. From these results it was reasoned that a new purification method was needed. It is important to note that crosslinking reactions used in the "in-pot" purification method were irradiated for 45 min. This overestimated crosslinking time may have produced significant photodegradation to the nucleoprotein complex thus contributing to the poor nucleopeptide MALDI spectra.

Besides the "in-pot" purification method, other procedures for isolating crosslinked peptides from nucleoprotein complexes have been reported (4, 28,

144, 146). In each case, the purification procedures were specific for isolating nucleopeptides that would be sequenced by gas phase sequencing. It was felt that any newly developed procedure should be designed such that the isolated nucleopeptide sample would be suitable for mass spectrometric analysis. The method developed here was based fundamentally on the scheme used for mass spectrometric identification of proteins from 2D polyacrylamide gels (125, 126). The "in-gel" method took advantage of the differing electrophoretic mobilities of the crosslinking products, $\text{Ung} \times \text{dT}_{20}$, Ung , and dT_{20} , to isolate crosslinked protein on a polyacrylamide gel for further processing. Upon discovering the optimum crosslinking time as 15 min, the newly developed "in-gel" purification method reproducibly yielded sufficient quantities of nucleopeptides for mass spectral analysis.

Nucleobase-amino acid crosslinking chemistry of protein-nucleic acid complexes is not well understood, making *a priori* predictions about the expected mass of isolated nucleopeptide species difficult. Varying masses for a given nucleopeptide species were expected based on the potential for differing points of attachment between the oligonucleotide and the peptide portion of the protein. However, analysis of the seven nucleopeptides indicated that their masses were very close to the sum of the neutral dT_{20} mass, the neutral peptide mass, and one proton (Table 3.1) which is consistent with other crosslinking studies that used dT-containing substrates (27, 28, 120). Even though the individual nucleopeptides identified appear to be homogeneous in mass, each nucleopeptide may not be homogeneous in structure. The mass analysis

performed here cannot distinguish between a peptide crosslinking to identical amino acids in different positions along the peptide or crosslinking to differing amino acids within the peptide resulting in the same mass. In either of these cases therefore, there may be multiple structural populations within a single mass population that would require further elucidation. It is for this reason that the results shown in this study can only be characterized as tentative.

The DNA binding region of *E. coli* Ung was first mapped by Bennett *et al.* (6). The UV-catalyzed Ung × dT₂₀ complex was digested with trypsin and crosslinked tryptic products were identified using a combination of MALDI MS and gas-phase sequencing. Four tryptic peptides, T6, T18, T19, and T18/T19, were identified as being crosslinked to dT₂₀ and thus the regions of the protein that these peptides covered were suggested to make up the DNA-binding region of Ung. From results shown here, seven peptides from Ung, T6, T6₍₋₃₎, T6₍₋₇₎, T9, T11, T18, and T17/18, were tentatively identified as being crosslinked to dT₂₀. Tryptic peptides T6 and T18 were also identified by the original work of Bennett *et al.* (6), but peptides T6₍₋₃₎, T6₍₋₇₎, T9, T11, and T17/18 were only identified in the current work. One reason for identifying the additional nucleopeptides may be due to the enhanced MALDI MS detection brought on by the different purification procedures used. The "in-gel" procedure developed here could be more thorough in removing contaminants such as denaturants and salts that interfere with the MALDI analysis. A second reason may be the different irradiation times used in the two procedures. In this work an irradiation time of 15 min was used whereas in the work of Bennett *et al.* (6) an irradiation time of 30

min was used. The extended irradiation time of 30 min may have been more detrimental to the nucleopeptides of T6₍₋₃₎, T6₍₋₇₎, T9, T11, and T17/18, preventing their detection. Also from comparing the results of Bennett *et al.* (6) and the work conducted here, nucleopeptides T18/T19 and T19, were not identified in the present work. These nucleopeptides may have contained labile crosslinks that were cleaved using the "in-gel" purification protocol.

The crosslinked species identified by Bennett *et al.* (6) defined two DNA binding regions within *E. coli* Ung corresponding to amino acids 58-80 (T6) and 185-213 (T18/T19). Since the publication of that work, the three-dimensional crystal structures of uracil-DNA glycosylases from *E. coli* (90, 108, 153), herpes simplex virus type-1 (116-118), and human (90, 91, 96, 128) have been reported. The two identified DNA-binding regions were directly compared to the crystal structure of *E. coli* Ung and to the corresponding homologous regions of the crystal structure of human Ung co-crystallized with DNA (Figure 3.13). The comparison of these structures revealed that these regions are involved in contacting double-stranded DNA and make up the catalytic region of the protein. From the present work, the regions of the protein covering amino acids 106-116 (T9), 130-147 (T11), and 180-184 (T17 of the T17/18 peptide) within the protein were also identified as being crosslinked to dT₂₀. Of the three newly identified regions, the N-terminal portion of T11 appears to be part of the DNA-binding region when compared to the crystal structures of *E. coli* Ung and human Ung co-crystallized with double-stranded DNA (Figure 3.14). The location of T9 within the protein does not appear to be close enough to the DNA-binding domain for

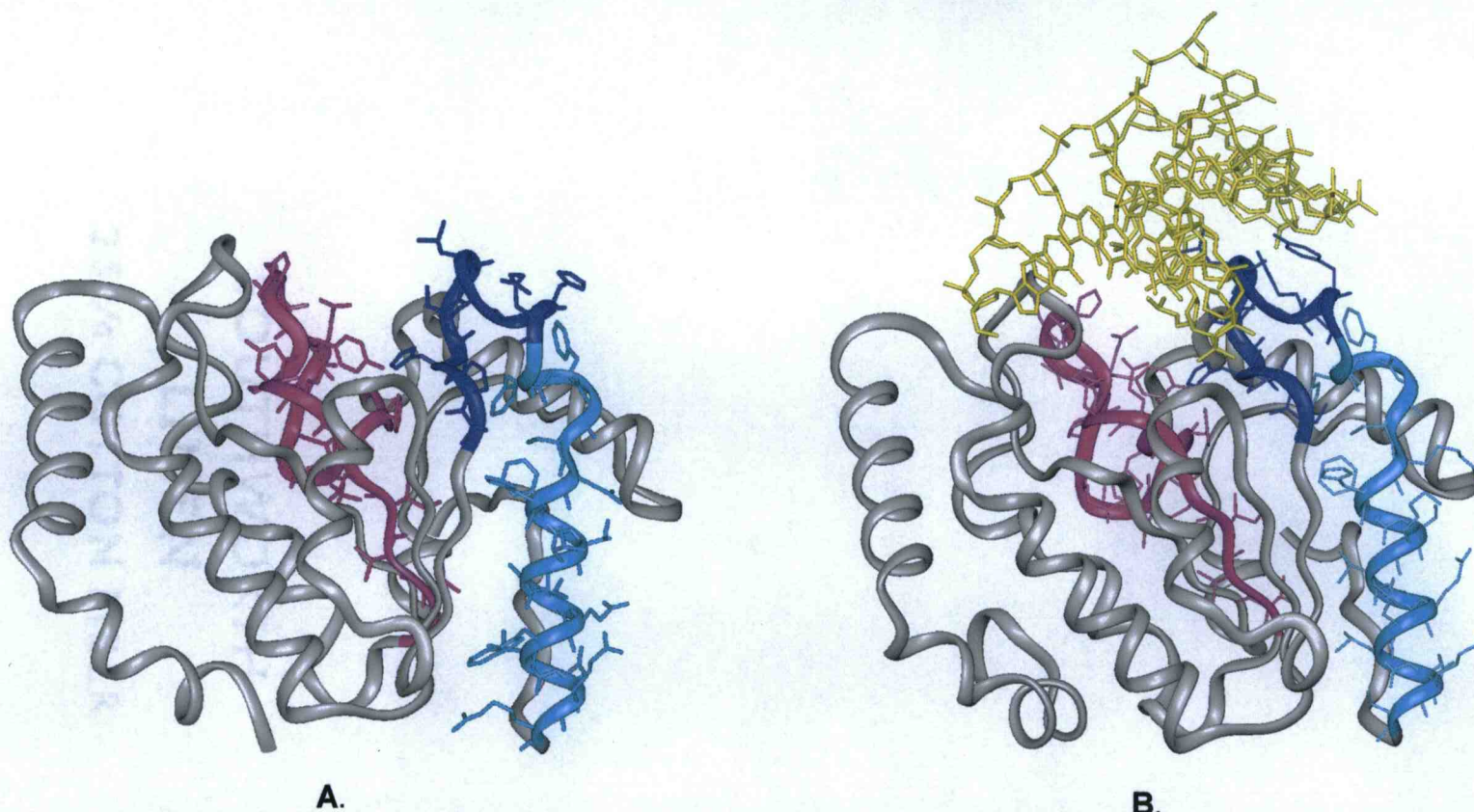
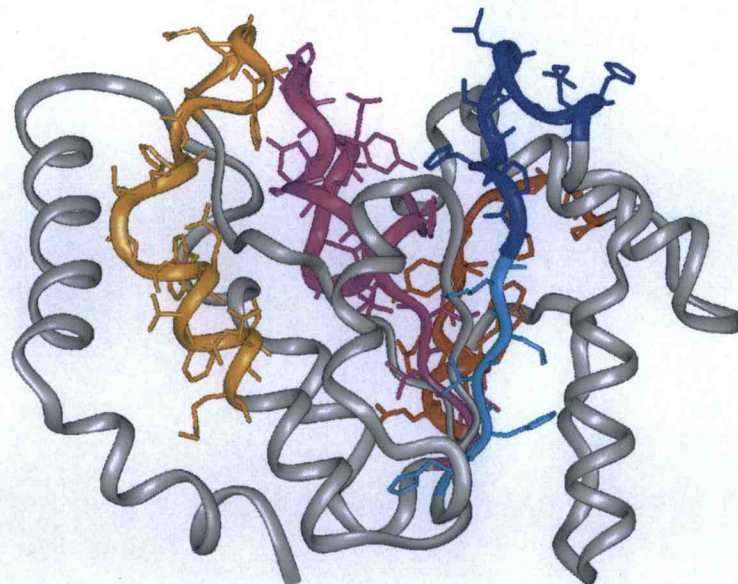
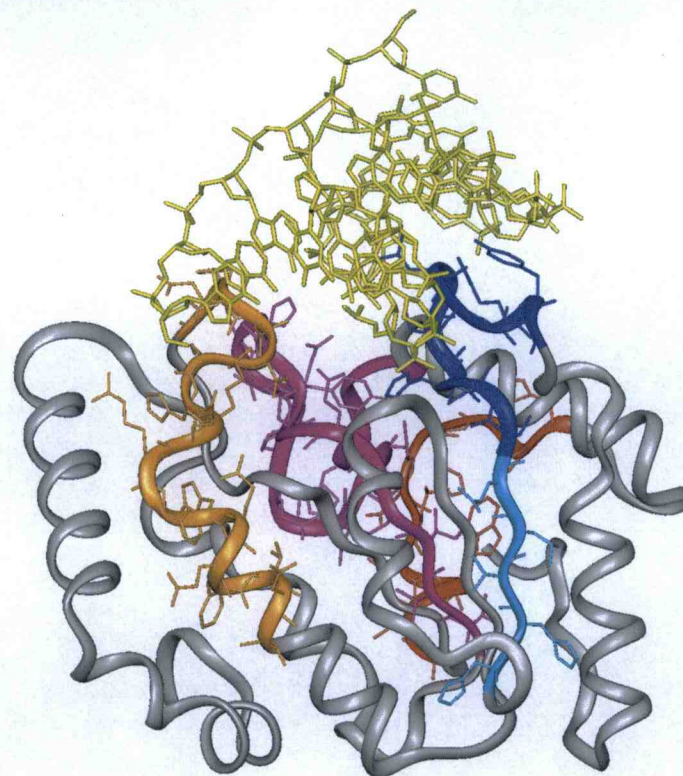


Figure 3.13. Correlation of Edman sequencing and MALDI MS data with the crystal structures of uracil-DNA glycosylase. Peptides of *E. coli* uracil-DNA glycosylase identified by Edman sequencing and MALDI MS as being crosslinked to dT₂₀ (6) are outlined in the crystal structure of the enzyme (**A**). Based on the sequence homology between *E. coli* Ung and human Ung, regions of the *E. coli* crystal structure that were identified as being crosslinked are highlighted in the crystal structure of the human enzyme co-crystallized with DNA (96) (**B**). Residues corresponding to T6 are highlighted in purple, T18 residues are in blue, T19 residues are in turquoise, and DNA is in yellow.



A.



B.

Figure 3.14. Correlation of MALDI MS data with the crystal structures of uracil-DNA glycosylase. Peptides of *E. coli* uracil-DNA glycosylase tentatively identified by MALDI MS as being crosslinked to dT₂₀ are outlined in the crystal structure of the enzyme (**A**). Based on the sequence homology between *E. coli* Ung and human Ung, regions of the *E. coli* crystal structure that were observed by MALDI MS as being crosslinked to dT₂₀ were highlighted in the crystal structure of the human enzyme co-crystallized with DNA (96) (**B**). Residues corresponding to T6 are colored in purple, T9 residues are in red, T11 residues are in orange, T17 residues are in turquoise, T18 residues are in blue, and the DNA is in yellow.

any of these amino acids to take part in. Possibly this nucleopeptide was mistakenly identified or the static crystal structures of the protein may not accurately represent the Ung:dT₂₀ complex present during crosslinking.

Likewise, the T17 portion of the T17/T18 nucleopeptide does not appear to be part of the DNA-binding region when compared to the crystal structures of the protein. This nucleopeptide may have been formed by either T17 or T18 being crosslinked to dT₂₀ and trypsin not cleaving the site joining the two peptides or by both peptides being crosslinked to one dT₂₀ molecule and proteolysis not taking place at the joining tryptic site. Further studies to sequence the amino acids of the isolated nucleopeptides, such as Edman degradation or tandem mass analysis (Chapter 4), would definitively identify the nucleopeptides and provide the identification of specific crosslinked residues.

4. AMINO ACID SEQUENCING OF PURIFIED UNG NUCLEOPEPTIDES USING ELECTROSPRAY IONIZATION MASS SPECTROMETRY

4.1 Results

4.1.1 Analytical Outline for Mass Spectrometric Analysis of Ung Nucleopeptides

After initial characterization of the isolated Ung nucleopeptides by MALDI MS (Table 4.1), nucleolytic digestion of the nucleopeptides was carried out to reduce the oligonucleotide portion of the molecules and simplify subsequent ESI MS/MS spectra. The nucleolytic sample was then analyzed by one of three analytical routes (Figure 4.1). The first route was analysis by MALDI MS to characterize the resulting digestion products. The second route was separation of the reaction products by HPLC and then amino acid sequencing by nanospray/ESI MS/MS. The third route was in-line separation and amino acid sequencing of the nucleopeptides by LC/ESI MS/MS. The results from each of these routes are reported below.

4.1.2 MALDI MS Characterization of Nucleolytically Digested Ung Nucleopeptides

Characterization by MALDI MS of the nuclease digestion products showed five series of ions (Figure 4.2) with each ion species within a series increasing in mass by one deoxyribothymidylic acid (MW=304.4). The molecular weight of mono-, di-, tri-, tetra-, and pentanucleotides were subtracted from the

Table 4.1. Summary of tentatively identified Ung × dT₂₀ tryptic peptides. The molecular weight of dT₂₀ ([M + H⁺] = 6024.1) was subtracted from the experimentally measured Ung × dT₂₀ m/z values (Figure 3.11). The resulting calculated peptide masses were compared to a peptide map for tentative identification of the Ung tryptic peptides that take part in crosslinking.

Experimental [peptide × dT ₂₀]	Calculated peptide mass [peptide × dT ₂₀] – dT ₂₀	Predicted peptide mass ^a	Tryptic peptide
8442.4	2418.1	2418.7	T6
8096.6	2072.5	2076.3	T6 ₍₋₃₎ ^b
7963.9	1939.8	1942.1	T11
7805.5	1782.4	1784.1	T17/T18
7711.7	1687.6	1687.9	T6 ₍₋₇₎ ^c
7352.6	1328.7	1328.6	T9
7191.7	1167.6	1169.3	T18

^a Predicted average peptide masses were calculated from a tryptic peptide map of *E. coli* Ung generated by the computer program GPMW, version 3.1, Lighthouse Data.

^{b,c} Comparison of these masses to a mass search of *E. coli* Ung indicated tryptic peptide T6 without three and seven C-terminal amino acids, respectively.

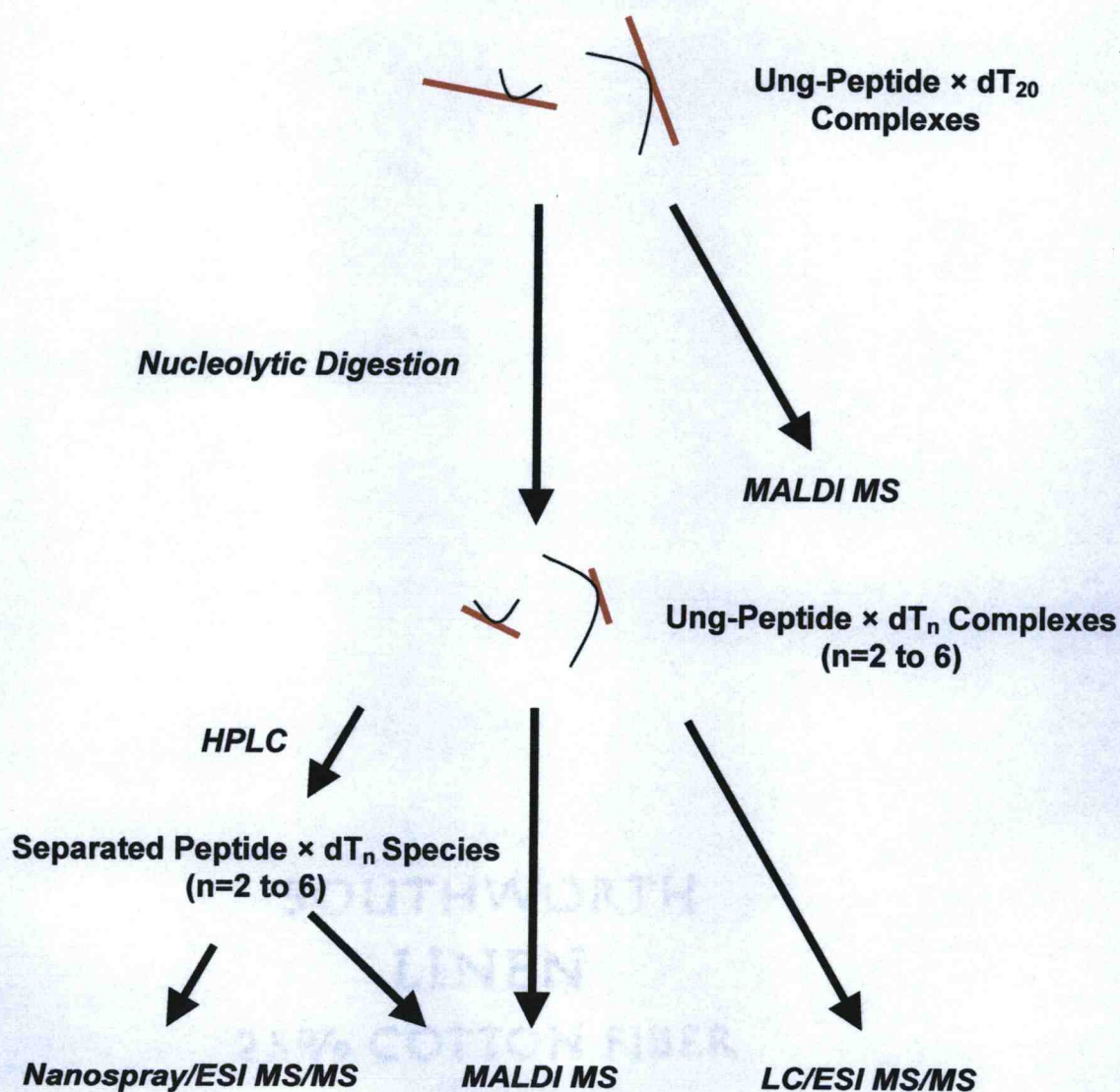


Figure 4.1. Analysis scheme for the identification of covalently attached amino acids by mass spectrometry. After characterization by MALDI MS, Ung peptide × dT₂₀ complexes were nucleolytically digested and the resulting peptide × dT_n complexes were first characterized by MALDI MS. The peptide × dT_n mixture was then separated by HPLC and collected fractions were analyzed by nanospray ESI MS/MS or the peptide × dT_n mixture was analyzed by LC/ESI MS/MS.

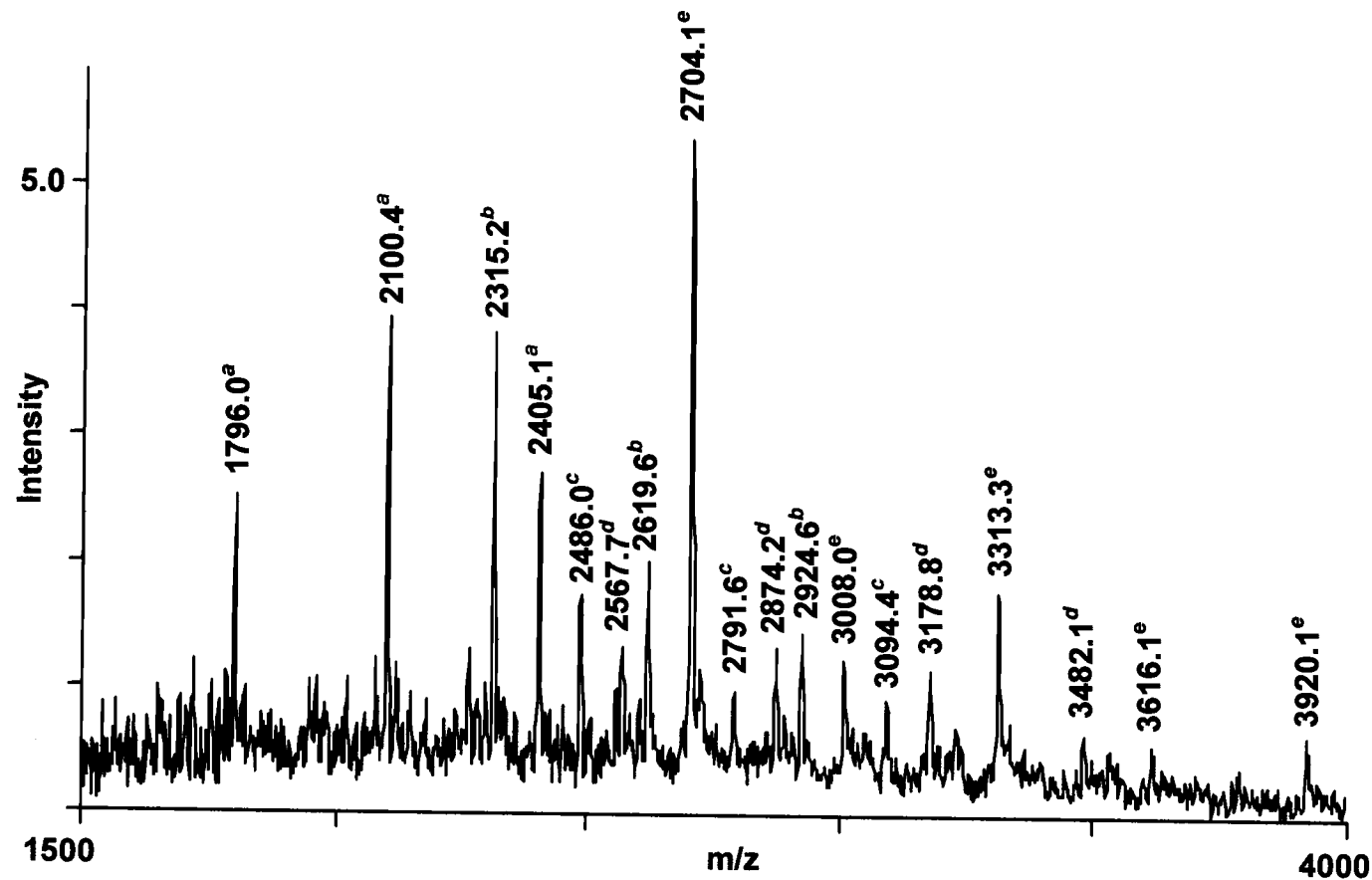


Figure 4.2. MALDI mass spectrum of purified tryptic Ung $\times$ dT₂₀ peptides after digestion with nuclease P1. Ung-peptide $\times$ dT₂₀ complexes isolated by the "in-gel" purification scheme (Figure 3.9) were digested with nuclease P1 and then analyzed by MALDI MS as, described under "Experimental Procedures." The resulting ion signals corresponded to singly protonated forms of Ung tryptic peptides crosslinked to varying sizes of polydeoxyribothymidylic acid. Five series of digestion products, as signified by **a**, **b**, **c**, **d**, and **e**, were observed.

experimentally measured peptide $\times$ dT_n ion species within a series and the resulting peptide masses were compared to a digestion map of Ung (Table 4.2). This analysis tentatively identified T18 (m/z 1796.0, 2100.4, and 2405.1), T11 (m/z 2567.7, 2874.2, 3178.8, 3482.1) T6₍₋₇₎ (m/z 2315.2, 2619.6, and 2924.6) and T6₍₋₃₎ (m/z 2704.1, 3008.0, 3313.3, 3616.1, and 3920.1) as peptides present in the nucleopeptide complexes, corroborating the presence of those peptides identified in the MALDI MS analysis of the peptide $\times$ dT₂₀ complexes (Table 4.1). One other ion series, m/z 2486.0, 2791.6, and 3094.4, did not directly identify a peptide that was identified in the peptide $\times$ dT₂₀ analysis. Using the calculated peptide mass of 1858.6 in a mass search of Ung indicated that the peptide was a truncated form of T6 that had lost five C-terminal amino acids, hereafter designated T6₍₋₅₎. Nucleopeptide ion series for T9 $\times$ dT_n and T17/T18 $\times$ dT_n, which were identified in the MALDI spectrum of peptide $\times$ dT₂₀ complexes, were not identified in the nuclease digestion mass spectrum. The quantity of each nucleopeptide species in this spectrum was estimated by absorbance to be in the range of 0.5 pmol to 5.0 pmol.

4.1.3 MALDI MS Characterization of HPLC-Purified Ung $\times$ dT_n Nucleopeptides

An Ung peptide $\times$ dT_n nuclease digestion mixture was separated by HPLC and two elution fractions were collected (Figure 4.3). The fractions were characterized by MALDI MS revealing that fraction A (Figure 4.4) contained the tentatively identified nucleopeptide series of T18 (m/z 1795.0, 2098.7,

Table 4.2. Analysis of nuclease P1 digestion products.

Nucleopeptide	-	[Predicted DNA mass + H⁺]^a	=	Peptide mass	Predicted peptide mass	Peptide I.D.
1796.0	-	627.4	=	1168.6	1169.3	T18
2100.4	-	931.6	=	1168.8	1169.3	T18
2405.1	-	1235.8	=	1169.3	1169.3	T18
2315.2	-	627.4	=	1687.8	1687.9	T6 ₍₋₇₎ ^b
2619.6	-	931.6	=	1688.0	1687.9	T6 ₍₋₇₎ ^b
2924.6	-	1235.8	=	1688.8	1687.9	T6 ₍₋₇₎ ^b
2486.0	-	627.4	=	1858.6	1858.1	T6 ₍₋₅₎ ^c
2791.6	-	931.6	=	1860.0	1858.1	T6 ₍₋₅₎ ^c
3094.4	-	1235.8	=	1858.6	1858.1	T6 ₍₋₅₎ ^c
2567.7	-	627.4	-	1940.3	1943.1	T11
2874.2	-	931.6	=	1942.6	1943.1	T11
3178.8	-	1235.8	=	1943.0	1943.1	T11
3482.1	-	1540.0	=	1942.1	1943.1	T11
2704.1	-	627.4	=	2076.7	2076.4	T6 ₍₋₃₎ ^d
3008.0	-	931.6	=	2076.4	2076.4	T6 ₍₋₃₎ ^d
3313.3	-	1235.8	=	2077.5	2076.4	T6 ₍₋₃₎ ^d
3616.1	-	1540.0	=	2076.1	2076.4	T6 ₍₋₃₎ ^d
3920.1	-	1844.2	=	2075.9	2076.4	T6 ₍₋₃₎ ^d

Table 4.2 (continued)

^a The protonated molecular weights of di-, tri-, tetra-, penta-, and hexadeoxyribothymidylic acid are 627.4, 931.6, 1235.8, 1540.0, and 1844.2, respectively.

^{b,c,d} Digestion products corresponding to tryptic fragment T6 less seven, five, and three C-terminal amino acids, respectively.

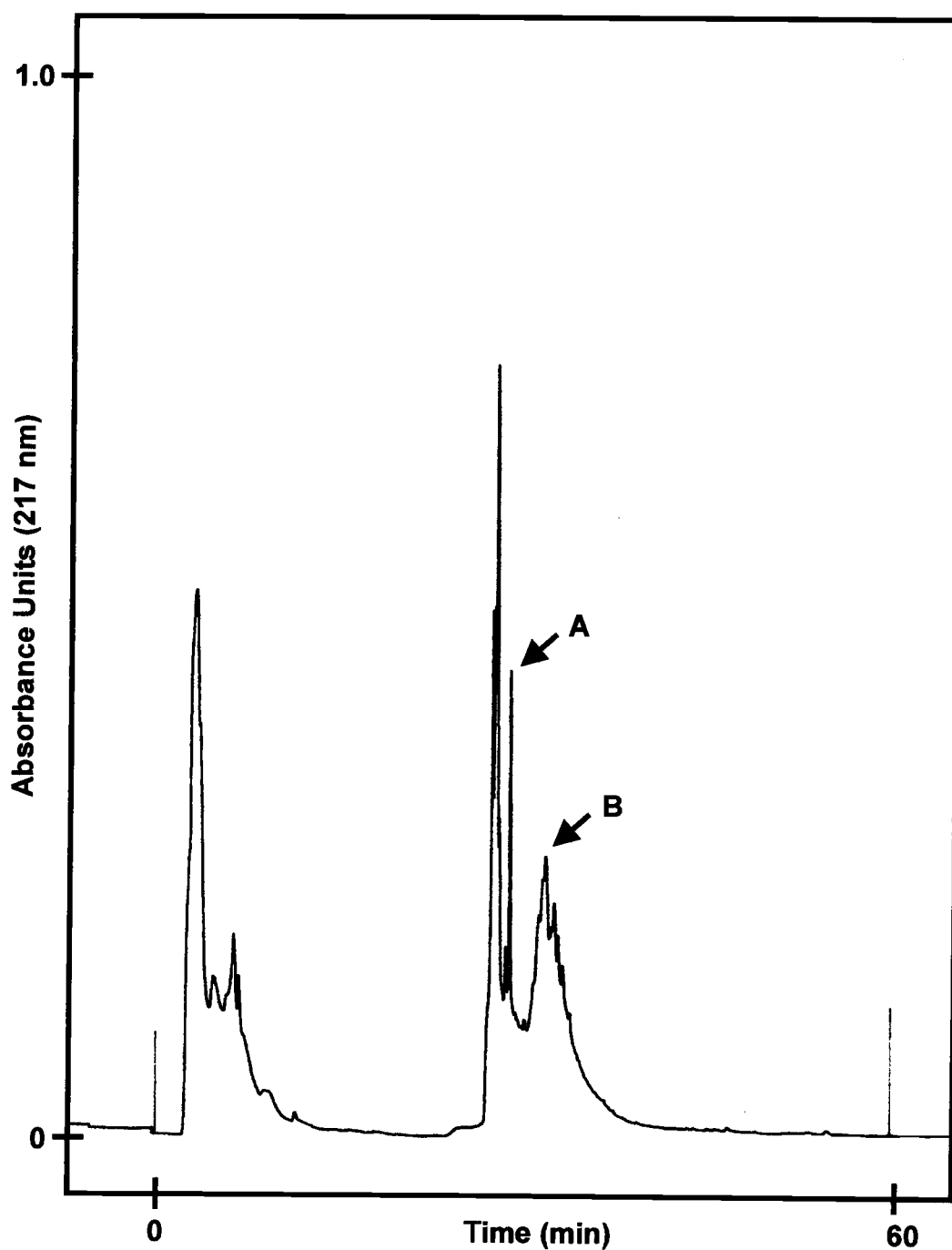


Figure 4.3. HPLC chromatogram of nuclease P1-digested nucleopeptides. Ung peptide $\times$ dT₂₀ complexes were nucleolytically digested and the resulting peptide $\times$ dT_n complexes (Figure 4.2) were separated by HPLC, as described under "Experimental Procedures." Two fractions, labeled **A** and **B**, were collected and subsequently analyzed by MALDI MS and ESI MS/MS.

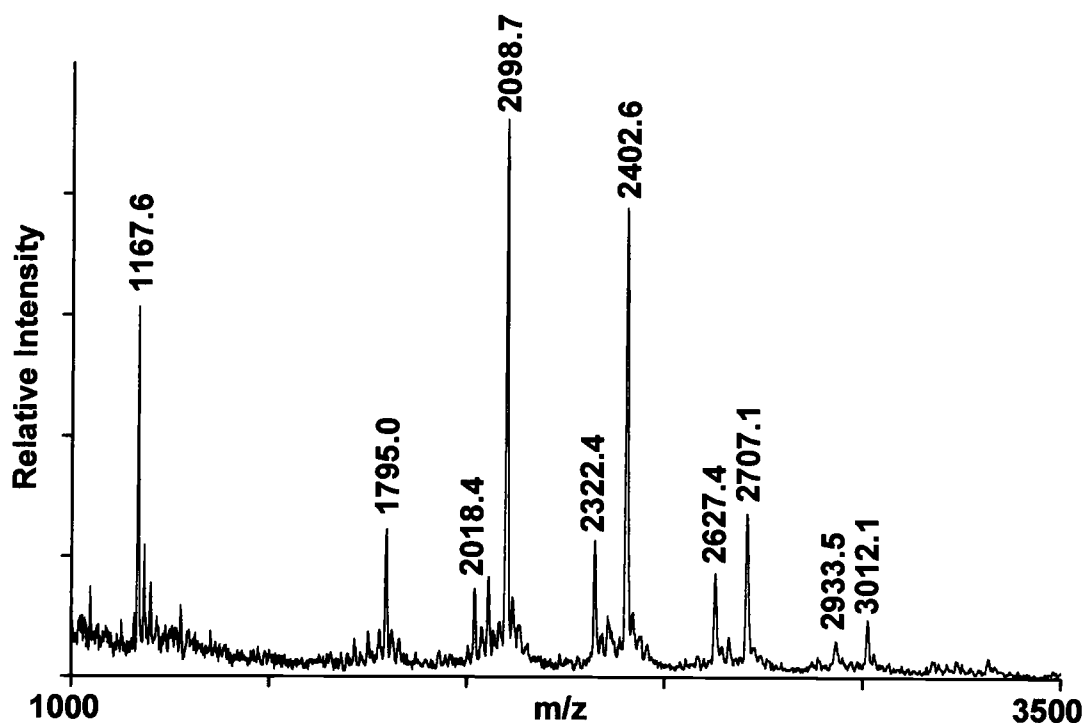


Figure 4.4. MALDI MS spectrum of HPLC fraction A. An Ung-peptide $\times$ dT_n mixture was separated by HPLC and a collected elution fraction, fraction A (Figure 4.3), was analyzed by MALDI MS, as described under "Experimental Procedures."

2402.6, 2707.1, and 3012.1). A second ion series was present (m/z 2018.4, 2322.4, 2627.4 and 2933.5) that differed from the first ion series by 80 mass units, the mass of a phosphate group. Nuclease P1 digestion produces 5'-phosphorylated nucleotides during digestion; therefore, the mass difference observed in this second ion series may have occurred if the nucleopeptide digestion products contained the initial 5'-terminus of dT₂₀ that was lacking a phosphate group. An ion species corresponding to free T18 (m/z 1167.6) was also observed. Fraction B (Figure 4.5) was tentatively identified as a mixture of T6₍₋₇₎, T6₍₋₅₎, and T6₍₋₃₎ crosslinked to dT₂ (m/z 2316.6, 2487.1, and 2704.8, respectively), T6₍₋₇₎ crosslinked to dT₄ (m/z 3314.5), and T11 crosslinked to dT₂, dT₃, and dT₄ (m/z 2571.7, 2875.9, and 3181.6).

4.1.4 Nanospray/ESI MS/MS Amino Acid Sequencing of HPLC Purified Ung-peptide × dT_n Nucleopeptides

4.1.4.1 Nanospray/ESI MS/MS of T18 × dT₄

A nanospray/ESI Q1 (MS) spectrum was obtained for HPLC fraction A (Figure 4.6). In comparing this spectrum to the MALDI MS spectrum for fraction A (Figure 4.4), the ions observed in the Q1 spectrum corresponded to triply charged T18 × dT_n nucleopeptides, where n increased from 3 to 6 (m/z 700.8, 802.4, 903.6, and 1005.2). Also associated with the triply charged nucleopeptide T18 × dT_n ions, where n increased from 4 to 6, were triply charged species corresponding to a difference of a phosphate group (m/z 775.6, 876.8, and 978.4). From this spectrum, ion species 802.4, tentatively identified as triply

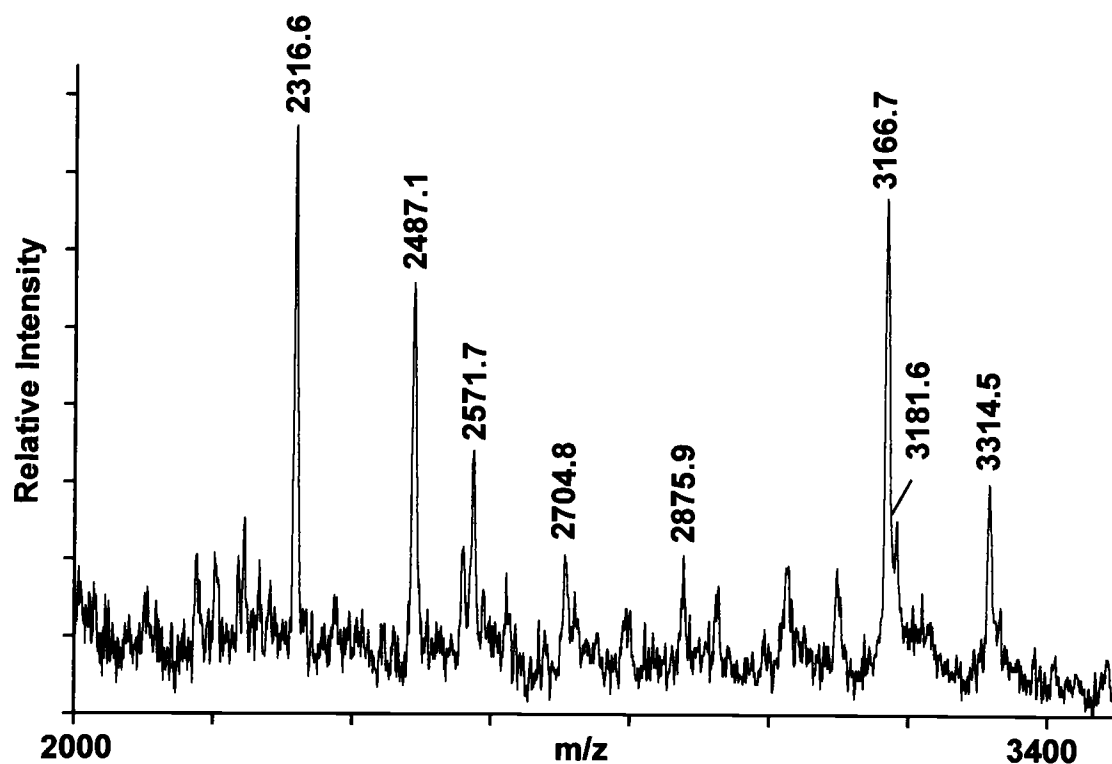


Figure 4.5. MALDI MS spectrum of HPLC fraction B. An Ung-peptide $\times$ dT_n mixture was separated by HPLC and a collected elution fraction, fraction B (Figure 4.3), was analyzed by MALDI MS, as described under "Experimental Procedures."

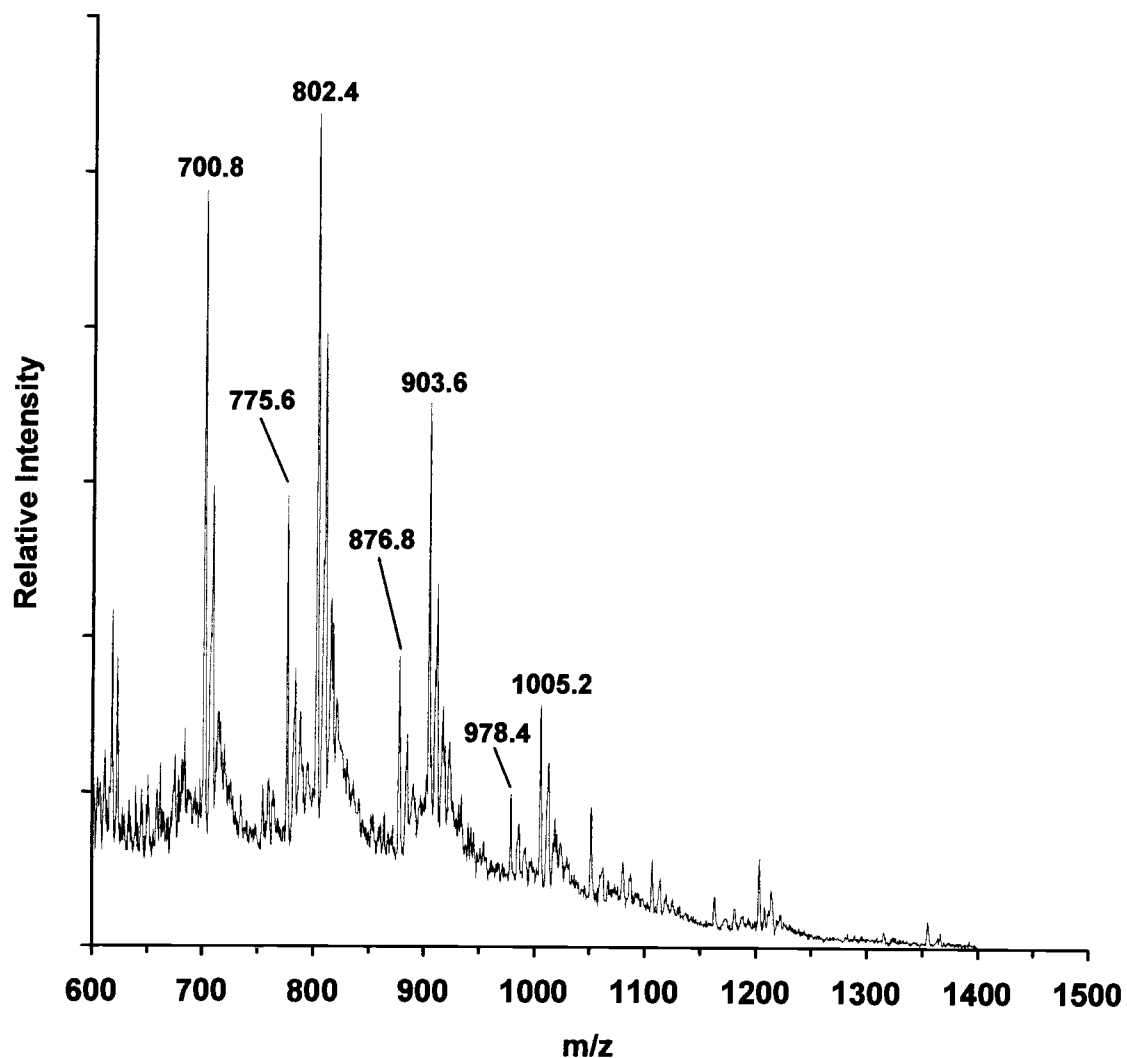


Figure 4.6. Nanospray ESI MS spectrum of HPLC elution fraction A. HPLC purified fraction A (Figures 4.3 and 4.4) was analyzed by nanospray/ESI MS, as described under "Experimental Procedures." The ion signals adjacent to each labeled peak on the high m/z -side are due to sodium adduction.

charged $T18 \times dT_4$, was selected as a precursor ion for tandem mass analysis since it was the most intense ion signal in the Q1 analysis. The resulting tandem mass spectrum (Figure 4.7) produced weak b_n - ($n=4-7,9,10$) and y_n -series ions ($n=1-4,6,8$) that confirmed the identity of the nucleopeptide as being $T18 \times dT_4$, as well as strong ion signals that resulted from fragmentation taking place along the oligonucleotide portion of the nucleopeptide (see Appendix for fragmentation nomenclature). The fragmentation pattern did not indicate any amino acid(s) containing the site(s) of covalent attachment.

4.1.4.2 Nanospray/ESI MS/MS of $T6_{(-7)} \times dT_2$ and $T6_{(-5)} \times dT_2$

A nanospray Q1 spectrum was also obtained for HPLC fraction B (Figure 4.8) and compared to the corresponding MALDI MS spectrum for this fraction (Figure 4.5). The ions observed in the Q1 spectrum corresponded to triply charged nucleopeptides which had been tentatively identified as $T6_{(-7)} \times dT_2$, $T6_{(-5)} \times dT_2$, $T11 \times dT_3$, and $T11 \times dT_4$ (m/z 772.4, 829.2, 958.8, and 1060.4, respectively). Tandem mass spectra of $T6_{(-7)} \times dT_2$ (Figure 4.9) showed both extensive b_n - ($n=2-7$) and y_n -ion series ($n=1-6$) that clearly identified this nucleopeptide species as $T6_{(-7)} \times dT_2$. A y_n^* -ion series ($n=9-15$) was also present contributing to the conformation of the nucleopeptide. The combination of b_n -, y_n -, and y_n^* -ions observed in this spectrum localized the site(s) of crosslinking within the Ung $T6_{(-7)}$ peptide to Pro₆₅, Tyr₆₆, and His₆₇. A similar interpretation was also made for the spectrum of $T6_{(-5)} \times dT_2$. In this spectrum,

Figure 4.7. Nanospray/ESI MS/MS spectrum of triply charged T18 × dT₄. Ion m/z 802.4 from the nanospray/ESI MS spectrum of HPLC elution fraction A (Figure 4.6) was selected as a precursor ion for MS/MS analysis, as described under "Experimental Procedures." Fragment ions for the peptide portion of the nucleopeptide are labeled as b_n (n=4-7, 9,10) and y_n (1-4, 6 to 8). Fragment ions from the oligonucleotide part of the nucleopeptide are labeled with italicized letters and fragment ions containing the point of covalent attachment are indicated with an asterisk.

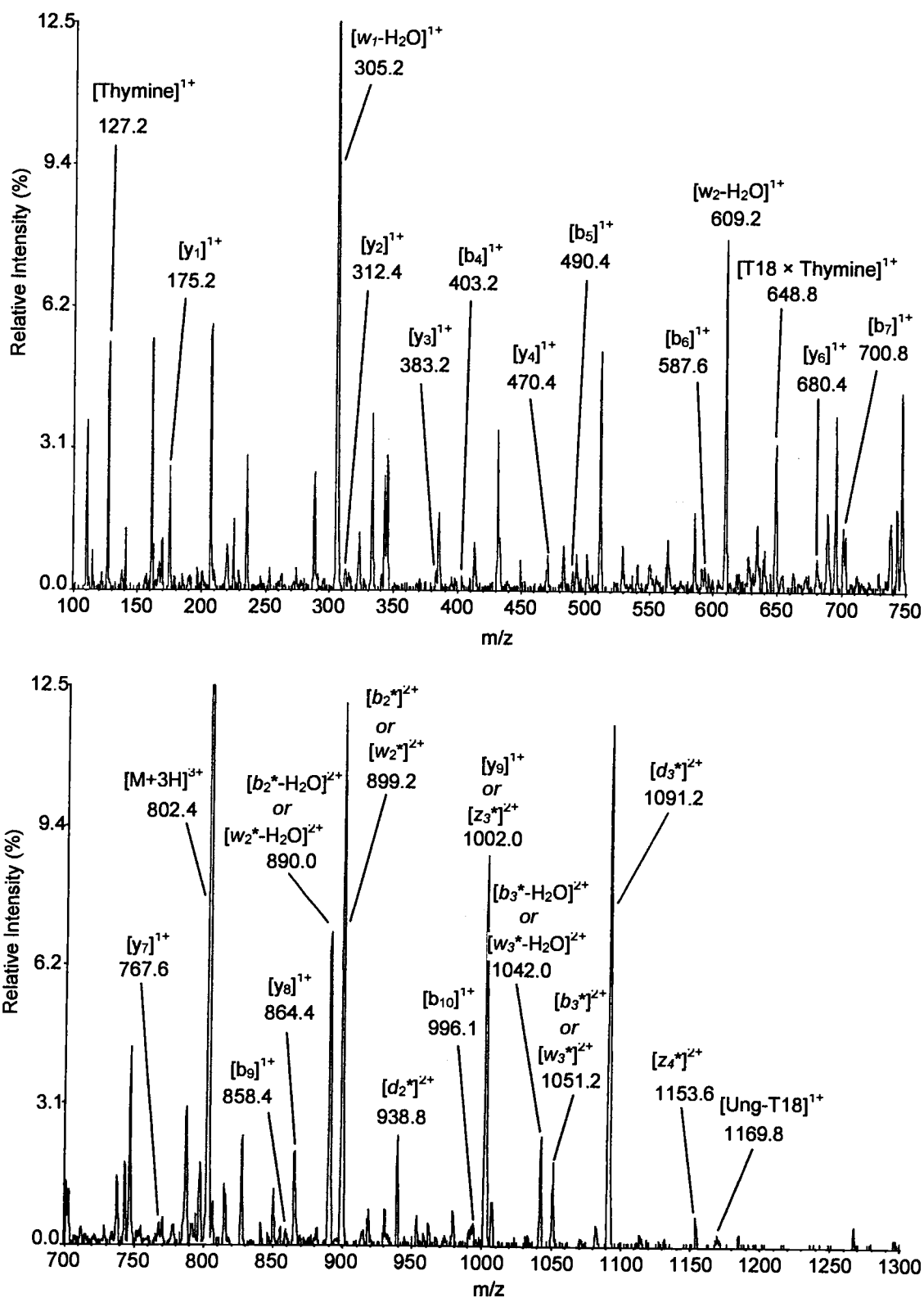


Figure 4.7. Nanospray/ESI MS/MS spectrum of triply charged T18 × dT4.

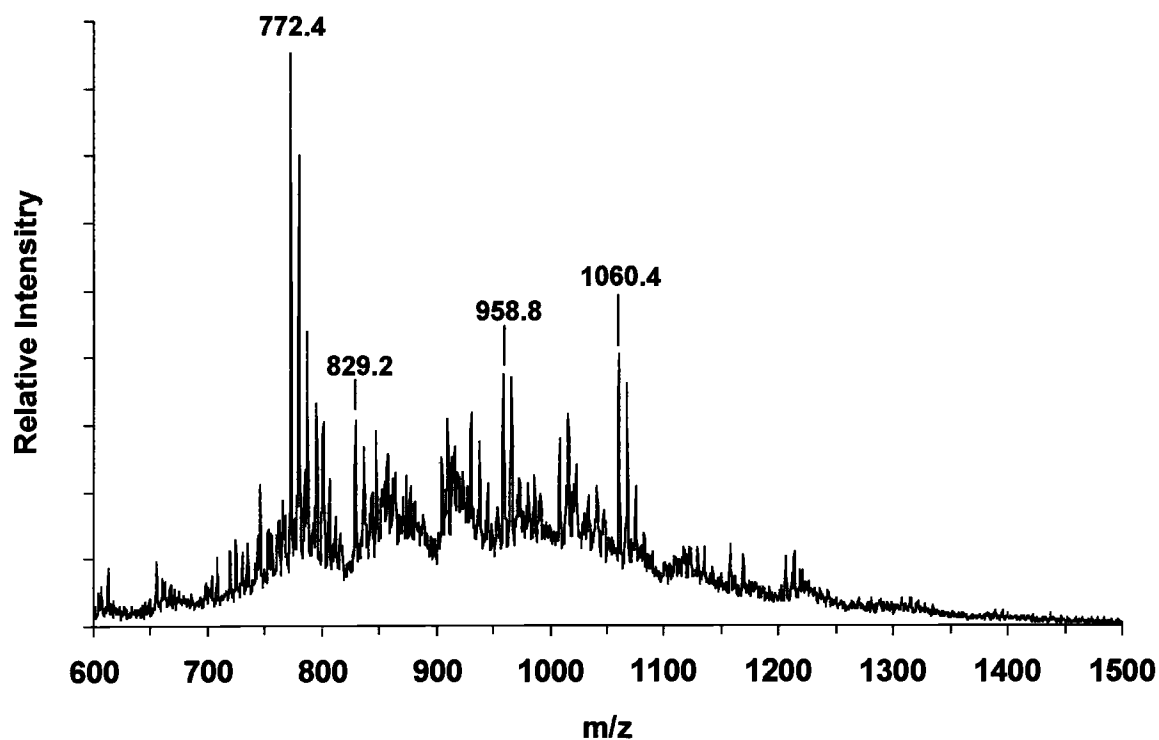


Figure 4.8. Nanospray/ESI MS spectrum of HPLC elution fraction B. HPLC purified fraction B (Figure 4.3) was analyzed by nanospray ESI MS as described under "Experimental Procedures." The ion signals adjacent to each peak on the high m/z-side are due to sodium adduction.

Figure 4.9. Nanospray/ESI MS/MS spectrum of triply charged $T6_{(-7)} \times dT_2$. Ion m/z 772.4 from the nanospray/ESI MS spectrum of HPLC elution fraction B (Figure 4.8) was selected as a precursor ion for MS/MS analysis, as described under "Experimental Procedures." Fragment ions from the peptide portion of the nucleopeptide are labeled as b_n ($n=2$ to 7), y_n ($n=1$ to 6), and y_n^* ($n=9$ to 15). Fragment ions containing the oligonucleotide portion of the nucleopeptide are indicated with an asterisk.

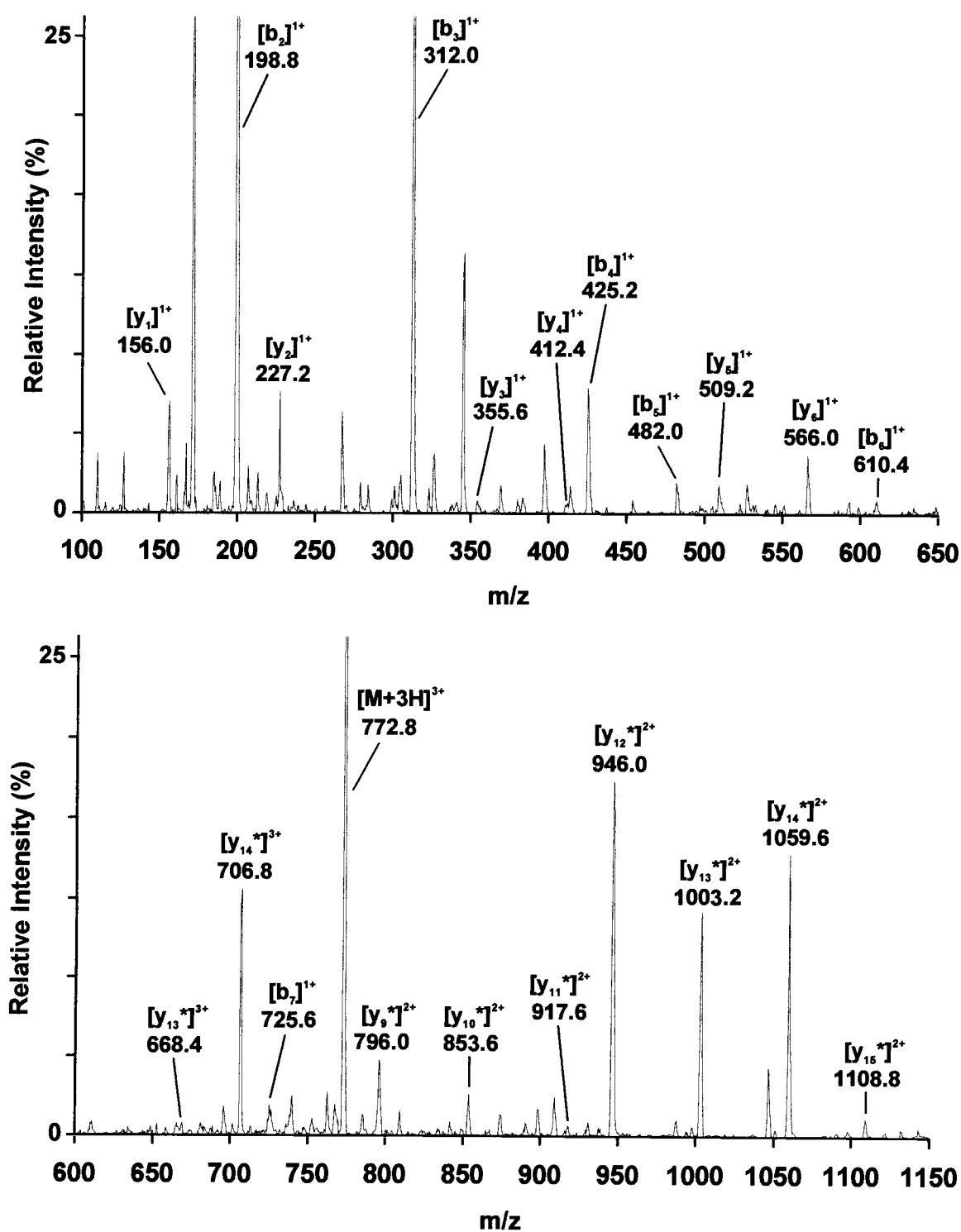


Figure 4.9. Nanospray/ESI MS/MS spectrum of triply charged $T6_{(-7)} \times dT_2$.

fragmentation corresponding to b_n^- (2-6,8), y_n^- (3,4,7,8), and y_n^{*-} ions ($n=10-12,14-16$) identified this nucleopeptide. The fragment ions of $[b_7]^{1+}$ and $[y_{15}]^{3+}$ had predicted m/z of 725.9 and 725.2, respectively. Thus ion species 725.6 observed in the spectrum could not be interpreted due to inadequate resolution of the instrument. Analysis of this spectrum also localized the site(s) of crosslinking to Pro₆₅, Try₆₆, and His₆₇ (Figure 4.10). It is interesting to note that b^* -series ions for either T6 nucleopeptide species were absent. Also absent in these spectra were fragmentation products resulting from cleavage along the oligonucleotide portion of the nucleopeptide.

4.1.5 LC MS Analysis of Ung- peptide $\times$ dT_n Complexes

An Ung-peptide $\times$ dT_n nuclease digestion mixture was injected into a LC system and chromatographic separation of the nucleopeptide mixture was conducted in-line with an ESI ion-trap mass spectrometer. Analysis of a base peak ion chromatogram (Figure 4.11A) of the nucleopeptide mixture showed multiple elution peaks. The tentatively identified nucleopeptide species T18 $\times$ dT₂, T6₍₋₇₎ $\times$ dT₂, T11 $\times$ dT₂, and T6₍₋₃₎ $\times$ dT₂ were located from extracted ion chromatograms (Figure 4.11 B through E). Analysis of the extracted ions showed $M+2H^+$ and $M+3H^+$ ion species (Figure 4.12) that were selected for collisionally induced dissociation (CID) experiments.

Figure 4.10. Nanospray/ESI MS/MS spectrum of triply charged $T6_{(-5)} \times dT_2$. Ion m/z 829.2 from the nanospray/ESI MS spectrum of HPLC elution fraction B (Figure 4.8) was selected as the precursor ion for MS/MS analysis, as described under "Experimental Procedures." Fragment ions from the peptide portion of the nucleopeptide are labeled as b_n ($n=2$ to 6, 8), y_n ($n=3,4,7,8$), and y_n^* ($n=10-12, 14,16$). Fragment ions containing the oligonucleotide portion of the nucleopeptide are indicated with an asterisk.

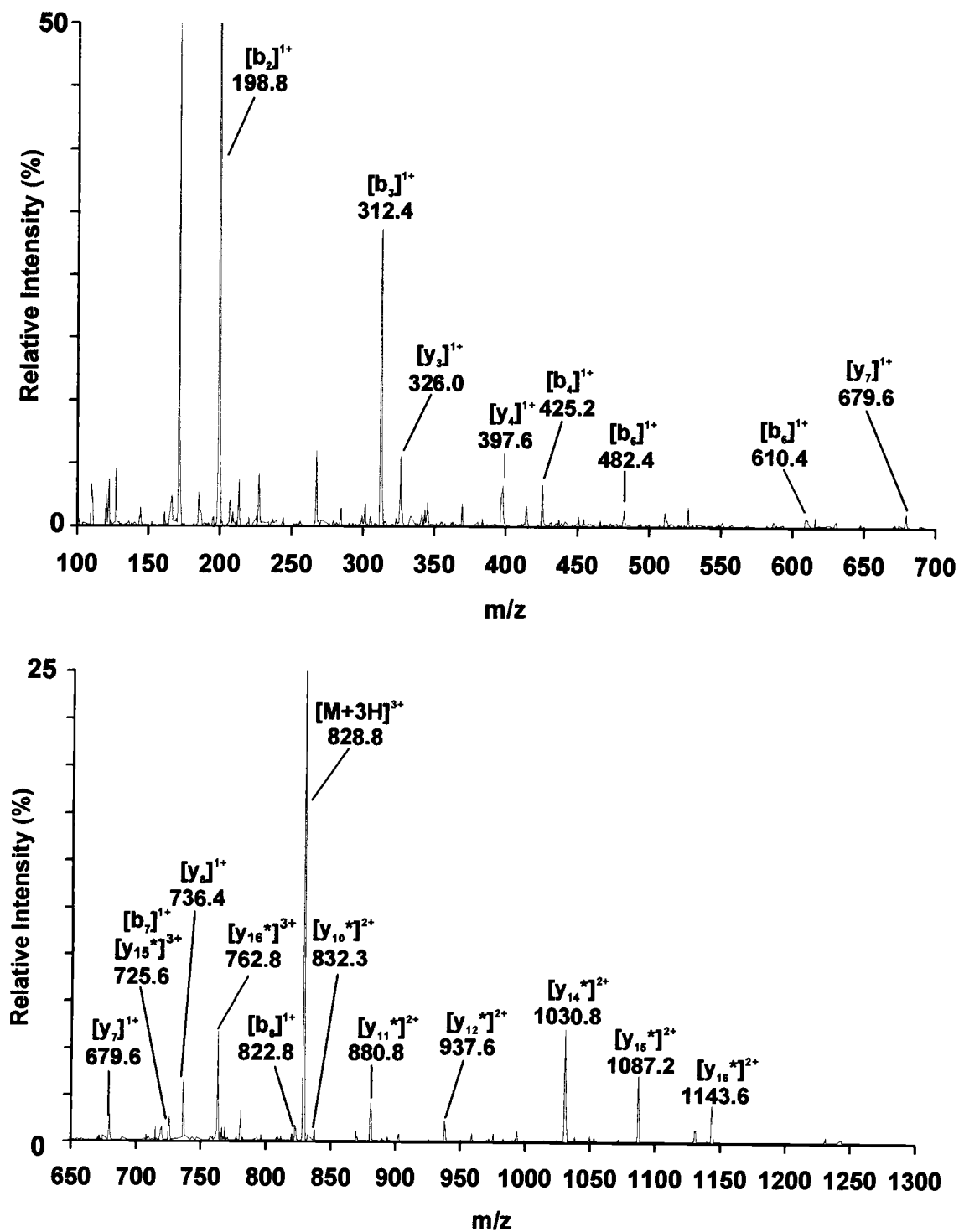


Figure 4.10. Nanospray/ESI MS/MS spectrum of triply charged T6₍₋₅₎ × dT₂.

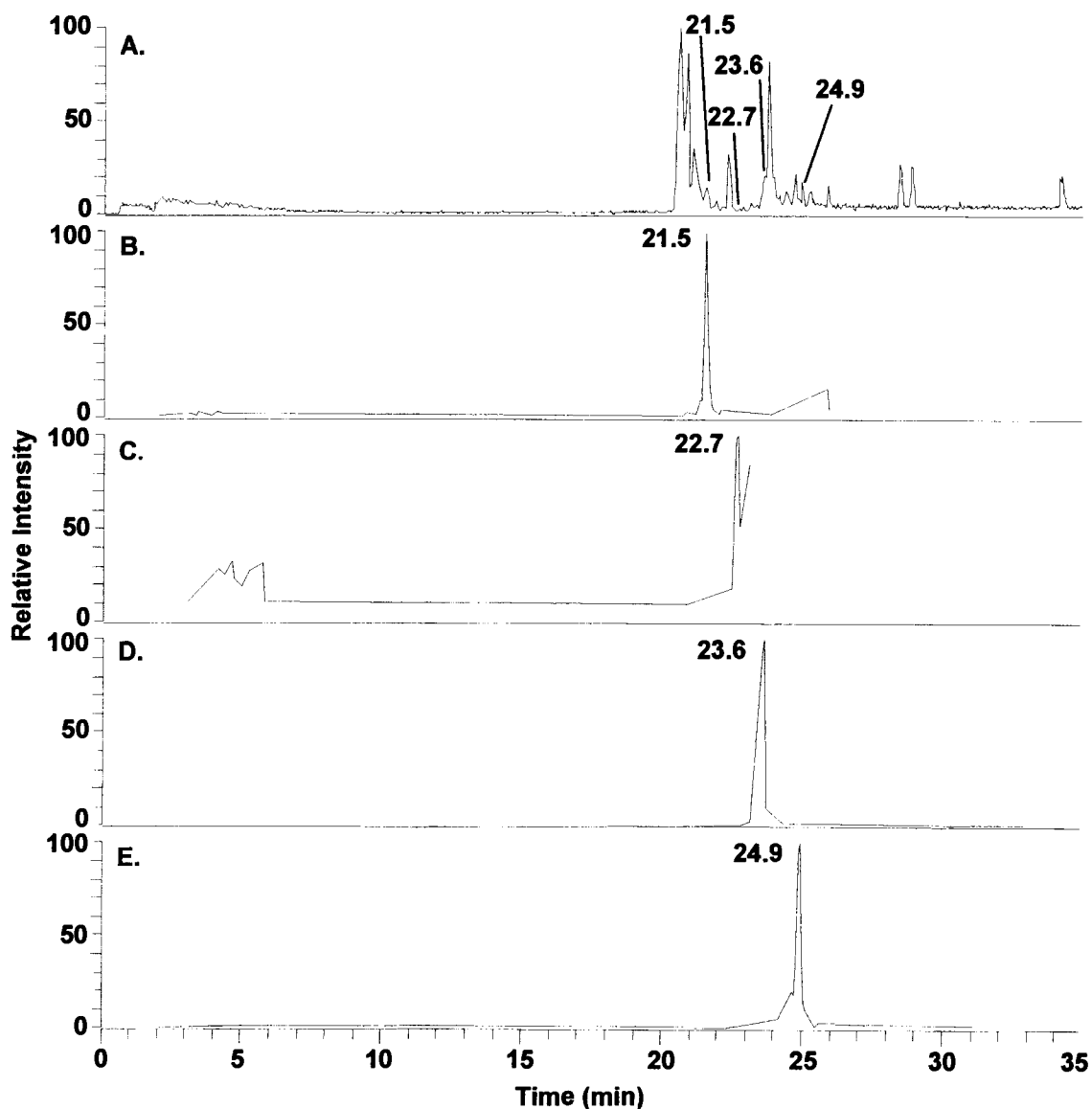


Figure 4.11. LC/ESI MS base peak ion chromatogram of Ung-peptide $\times$ dT_n complexes. A mixture of Ung-peptide $\times$ dT_n complexes produced from a nuclease P1 digestion (Figure 4.2) were directly analyzed by LC/ESI tandem mass spectrometry as described under "Experimental Procedures." A base peak ion chromatogram is shown for the entire chromatographic separation (**A**) and extracted ion chromatograms are shown for elution peaks corresponding T18 $\times$ dT₂ (**B**), T6₍₋₇₎ $\times$ dT₂, (**C**), T11 $\times$ dT₂ (**D**), and T6₍₋₃₎ $\times$ dT₂ (**E**).

Figure 4.12. LC/ESI MS spectra of Ung-peptide $\times$ dT_n complexes. Ung-peptide $\times$ dT_n complexes were chromatographically separated and the elution peaks (Figure 4.11B through E) were analyzed in-line by ESI MS as described under "Experimental Procedures." Doubly and triply charged ion species are labeled for T18 $\times$ dT₂ (**A**), T6₍₋₇₎ $\times$ dT₂ (**B**), T11 $\times$ dT₂ (**C**), and T6₍₋₃₎ $\times$ dT₂ (**D**).

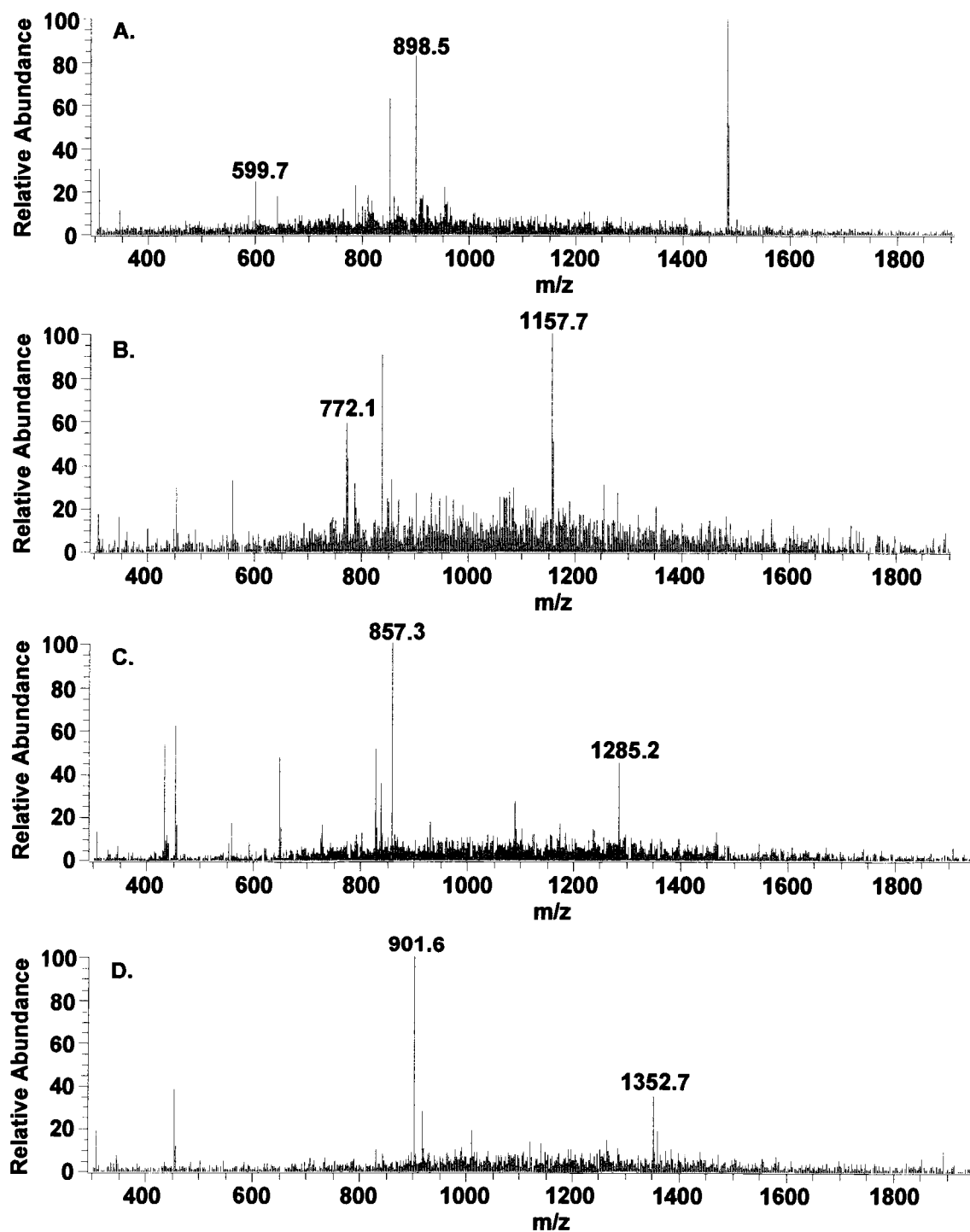


Figure 4.12. LC/ESI MS spectra of Ung-peptide $\times$ dT_n complexes.

4.1.6 LC MS/MS Peptide Sequencing of Ung-peptide $\times$ dT_n Complexes

4.1.6.1 LC MS/MS of T18 $\times$ dT₂

The CID mass spectrum of doubly charged T18 $\times$ dT₂ (Figure 4.13) revealed fragmentation taking place along the peptide backbone to produce y-series ions (y₄ to y₁₀) that confirmed the identity of this ion species. Fragmentation taking place along the peptide backbone to produce y*-ions was also present (y₃* to y₉*), as was fragmentation taking place along the oligonucleotide portion of the nucleopeptide. Three ion species from this mass spectrum, m/z 745.7, 1393.4, and 1490.5, corresponded to nucleopeptide fragments that could have been produced from cleavage of either the peptide backbone (y_n*) or the DNA backbone (*italicized lettering*). The ion species m/z 1490.5 and 745.7 could be predicted to be [y₈*]¹⁺ (calculated M+H⁺=1490.6) and [y₈*]²⁺ (calculated M+2H⁺=745.8), respectively, or could be predicted to be [w₁*]¹⁺ (calculated M+H⁺=1490.7) and [w₁*]²⁺ (calculated M+2H⁺=745.8), respectively. Similarly, ion species m/z 1393.4 could be predicted to be [y₇*]¹⁺ (calculated M+H⁺=1393.5) or [z₁*]¹⁺ (calculated M+H⁺=1393.7), respectively. The mass correlations between these predicted fragmentation products are all within 0.2 mass units, thus making identification of these ion species ambiguous. The CID fragmentation pattern produced for this nucleopeptide was extensive enough to confirm its identity, but the overlap in y- and y*-ions and the lack of b- and b*-ions prevented the identification of any amino acids within T18 $\times$ dT₂ that contained a covalent crosslink.

Figure 4.13. LC/ESI MS/MS spectrum of T18 × dT₂. The doubly charged ion m/z 898.5 was observed in the ESI MS spectrum (Figure 4.12A) as well as observed as a singly charged ion in the MALDI MS spectrum (Figure 4.2) of the nucleopeptide mixture. To confirm the identity of this ion species and to structurally characterize the nucleopeptide, it was selected as a precursor for LC/ESI MS/MS, as described under "Experimental Procedures." Fragmentation peaks m/z 1571.4, 1490.5, 1393.8, 889.5, 849.4, 745.7, 737.3, and 648.3 were consistent with predictions of cleavage along the DNA backbone of the nucleopeptide. The identity of this ion species was confirmed by y_n^- ($n=2,4-10$) b_n^- ($n=3,5,7,9$), y_n^* -ion ($n=2-6, 9$), and b_n^* -ion series ($n=9,10$). ♦ indicates ion species that can not be differentiated from fragmentation of the peptide backbone (y_7^* , y_8^* , and y_8^{*2+}) or fragmentation of the DNA backbone (z_1^* , w_1^* , and w_1^{*2+}).

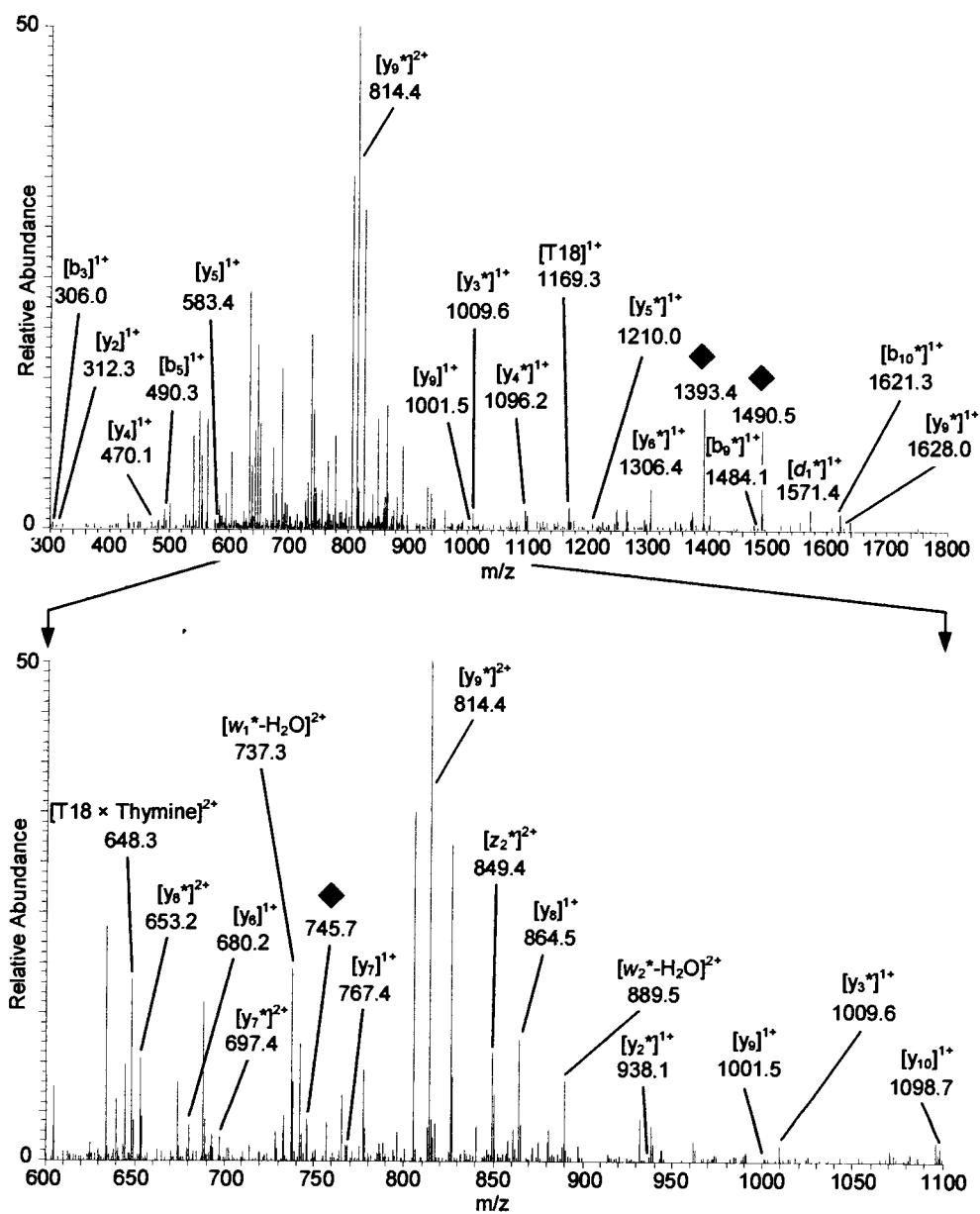


Figure 4.13. LC/ESI MS/MS spectrum of T18 x dT₂.

4.1.6.2 LC MS/MS of $T6_{(-7)} \times dT_2$

The doubly and triply charged precursor ions of $T6_{(-7)} \times dT_2$ were selected for CID analysis. Analysis of the fragmentation pattern for the triply charged precursor (Figure 4.14) revealed a series of six b-ions (b_3 through b_7 , and b_9), a series of four b^* -ions (b_{12}^* through b_{15}^*), a series of two y-ions (y_5 and y_6), and a series of eight y^* -ions (y_7^* , y_9^* through y_{15}^*). The difference between y_7^* and y_6 results in the combined mass of dT_2 and histidine, identifying His_{67} as a site of crosslinking. A continuous conversion of b- to b^* -ions was not observed, which would have also identified His_{67} as a site of crosslinking, but the b- and b^* -ion fragments that were present localized the site of crosslinking between His_{67} and Pro_{69} . The CID spectrum of the doubly charged precursor of $T6_{(-7)} \times dT_2$ (Figure 4.15) was about an order of magnitude less intense than for the triply charged precursor, but still yielded significant structural information about the nucleopeptide. This spectrum showed three y-series ions (y_4 to y_6) and seven y^* -series ions (y_7^* to y_{10}^* and y_{12}^* to y_{14}^*) that definitively identified this ion species. The mass difference between y_7^* and y_6 indicated His_{67} as the site of crosslinking, thus supporting the fragmentation pattern interpretation for the triply charged ion species of $T6_{(-7)} \times dT_2$.

Figure 4.14 LC/ESI MS/MS spectrum of triply charged $T6_{(-7)} \times dT_2$. The triply charged ion of m/z 772.5 was observed in the ESI MS spectrum (Figure 4.12B) as well as observed as a singly charged ion in the MALDI MS spectrum (Figure 4.2) of the nucleopeptide mixture. To confirm the identity of this ion species and to structurally characterize the nucleopeptide, it was selected as a precursor for LC/ESI MS/MS, as described under "Experimental Procedures." The identity of this ion species was confirmed by b_n - ($n=3-7,9$), y_n -($n=5,6$), and y_n^* -ion series ($n=7,9-15$). Fragment ions containing the dinucleotide portion of the nucleopeptide are indicated with an asterisk. The location of the covalent crosslink to His₆₇ is determined from the difference in mass between y_7^* and y_6 .

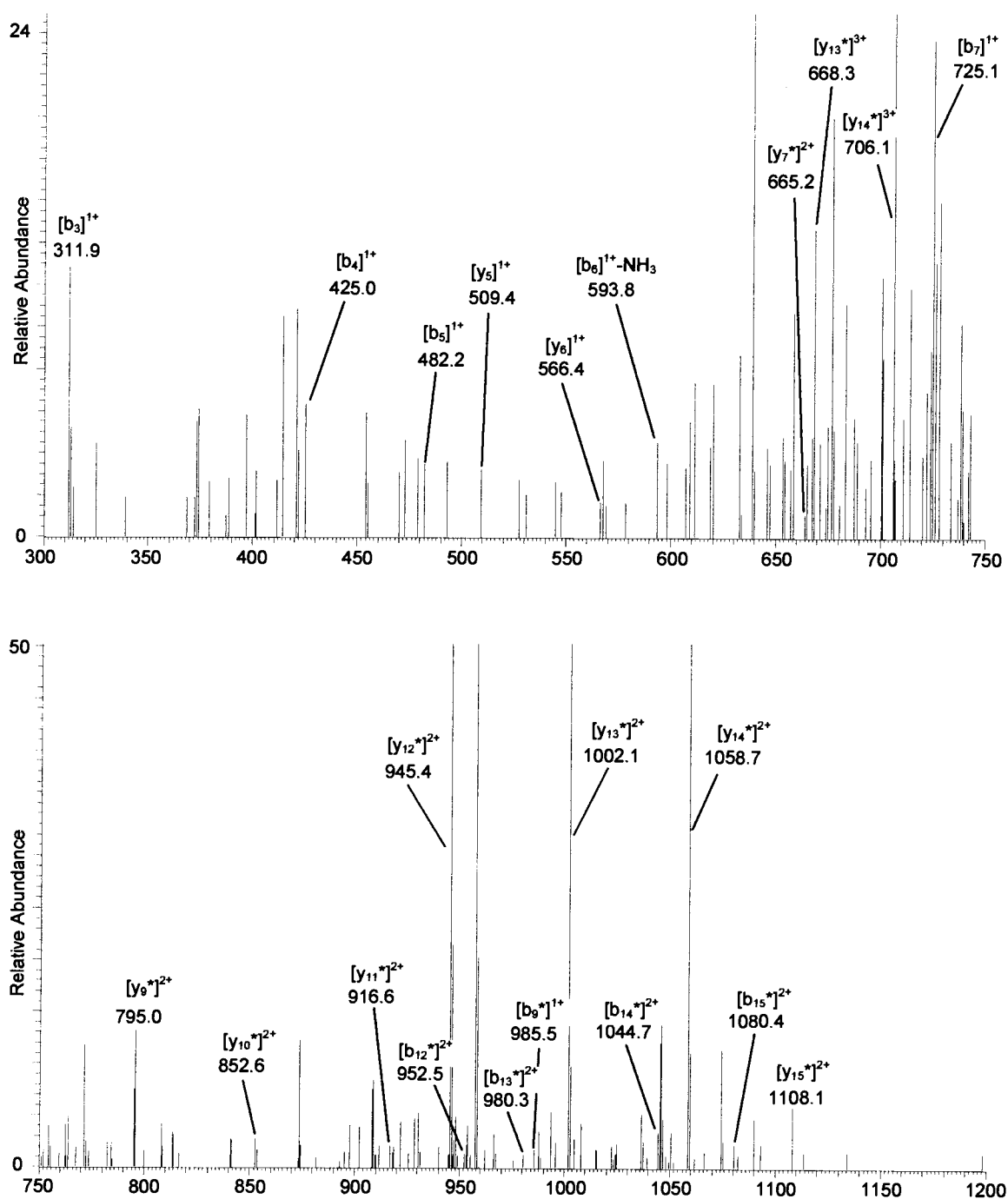


Figure 4.14 LC/ESI MS/MS spectrum of triply charged T6₍₋₇₎ × dT₂.

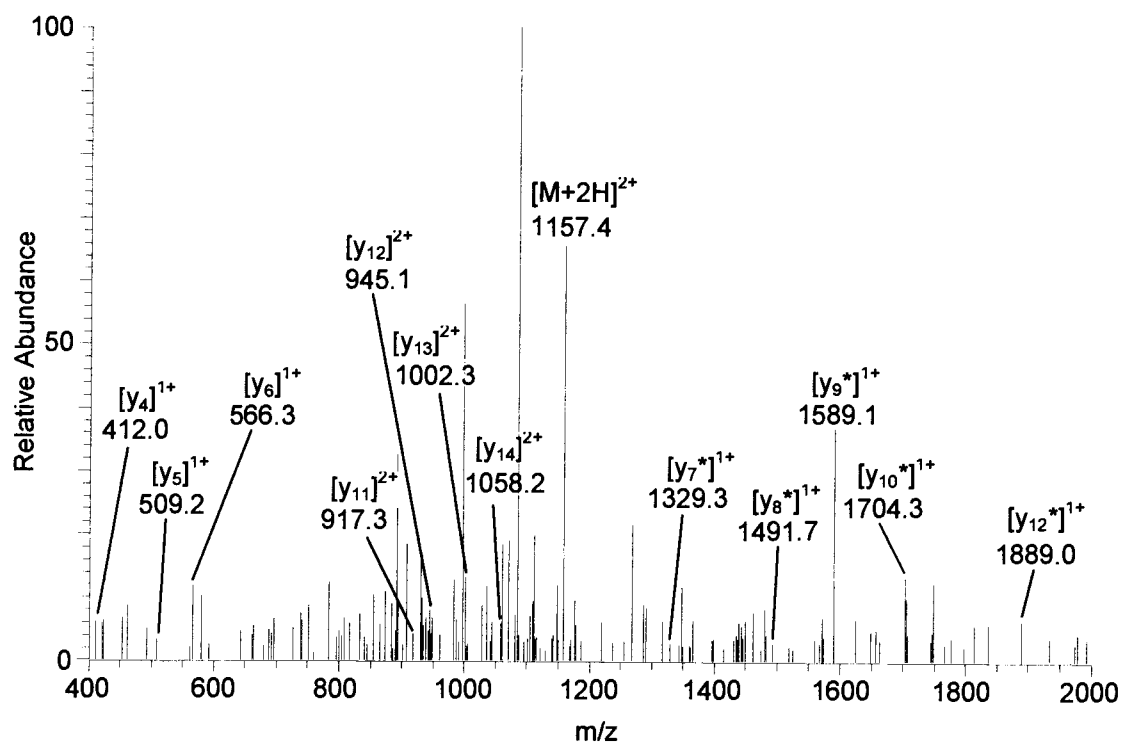


Figure 4.15. LC/ESI MS/MS spectrum of doubly charged $T6_{(-7)} \times dT_2$. Peptide sequencing of isolated nucleopeptide species was conducted by LC/ESI MS/MS, as described under "Experimental Procedures." The doubly charged ion species of putatively identified $T6_{(-7)} \times dT_2$ ($M+2H^+=1157.4$) was selected as a precursor ion for tandem mass analysis. The resulting mass spectrum produced y_n - ($n=4-6$) and y_n^* -ions ($n=7-9, 11-14$) ions that confirmed the identity of the nucleopeptide. The mass difference between y_7^* and y_6 indicated His₆₇ as the site of crosslinking.

4.1.6.3 LC MS/MS of T11 × dT₂

The CID spectrum of the triply charged ion species of T11 × dT₂ revealed series of y-ions (y₃ through y₁₁, and y₁₃), y*-ions (y₁₄* and y₁₅*), b-ions (b₄ to b₆, b₈ to b₁₁, b₁₃, and b₁₅) and b*-ions (b₇* to b₁₃*) that identify this ion species as T11 × dT₂ (Figure 4.16). In addition to these peptide fragments, intense fragmentation peaks corresponding to cleavage along the dinucleotide backbone were present. Cleavage at the covalent crosslink bond(s) that produced a doubly charged T11 fragment (m/z 972.1) was also present. The mass difference between y₁₄* and y₁₃ results in the combined mass of dT₂ and histidine indicating His₁₃₃ as being crosslinked to dT₂, but there is no corresponding conversion of b- to b*-ions to support this amino acid as being crosslinked. Similarly, the mass difference between b₇* and b₆ indicates His₁₃₅ as an amino acid crosslinked to dT₂, but there are no corresponding y- and y*-ions to support this identification. The overlap of b- and b*-ions also suggests the presence of additional crosslinked amino acids between Ala₁₃₈ and Trp₁₄₂.

4.1.6.4 LC MS/MS of T6₍₋₃₎ × dT₂

Both doubly and triply charged precursor ions of tentatively identified T6₍₋₃₎ × dT₂ were selected for CID experiments. The triply charged precursor (Figure 4.17) showed a series of seven b-ions (b₃ through b₇, b₉, and b₁₀) and a series of seven y-ions (y₅ through y₁₁) that definitively identified the peptide portion of this nucleopeptide. Series of y*-ions (y₁₁* through y₁₉*) and b*-ions (b₉*, b₁₀*, and

Figure 4.16. LC/ESI MS/MS spectrum of T11 × dT₂. The triply charged ion of m/z 857.3 was observed in the ESI-MS spectrum (Figure 4.12C) as well as the singly charged ion in the MALDI MS spectrum (Figure 4.2) of the nucleopeptide mixture. To confirm the identity of this ion species and to structurally characterize the nucleopeptide, it was selected as a precursor for LC/ESI MS/MS, as described under "Experimental Procedures." The major fragmentation peaks m/z 1173.1, 1164.4, 1124.0, 1035.3, 750.0, and 690.4 correspond to cleavage along the DNA backbone and are labeled with italicized letters. The ion at m/z 972.1 corresponds to doubly charged T11 (less the nucleotide). The peptide sequence is represented by b_n - ($n=4-6, 8-11, 13, 15$), b_n^* - ($n=7-13$), y_n - ($n=3-11, 13$), and y_n^* - ion series ($n=14, 15$). An asterisk indicates fragment ions containing the covalent attachment.

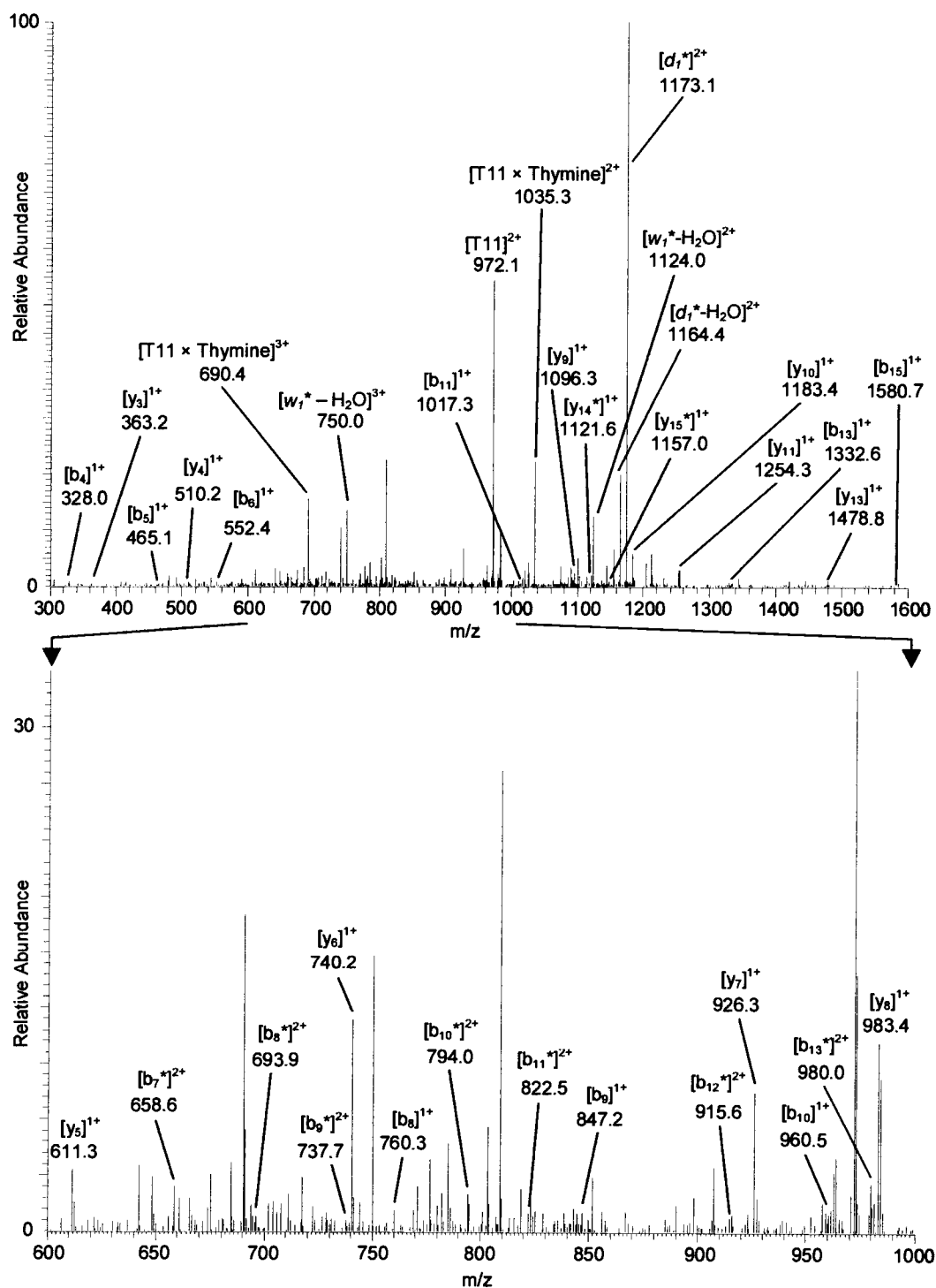


Figure 4.16. LC/ESI MS/MS spectrum of T11 $\times$ dT₂.

Figure 4.17. LC/ESI MS/MS spectrum of $T6_{(-3)} \times dT_2$. The triply charged ion of m/z 901.9 was observed in the ESI-MS spectrum (Figure 4.12D) and was observed as a the singly charged ion form in the MALDI-MS spectrum (Figure 4.2) of the nucleopeptide mixture and was putatively identified as $T6_{(-3)} \times dT_2$ (Table 4.2). To confirm the identity of this ion species and structurally characterized the nucleopeptide, it was selected as a precursor for LC/ESI MS/MS, as described under "Experimental Procedures." The identity of this ion species was confirmed by b_n^- ($n=3-7,9,10$), b_n^* - ($n=9,10,16-19$), y_n^- ($n=5-11$), and y_n^* -ion series ($n=11-19$). Fragment ions containing the dinucleotide portion of the nucleopeptide are indicated with an asterisk. The location of the covalent crosslink to Tyr₆₆ and His₆₇ is determined from the difference in mass between y_{11}^* and y_{10} and the difference in mass between y_{12}^* and y_{11} , respectively.

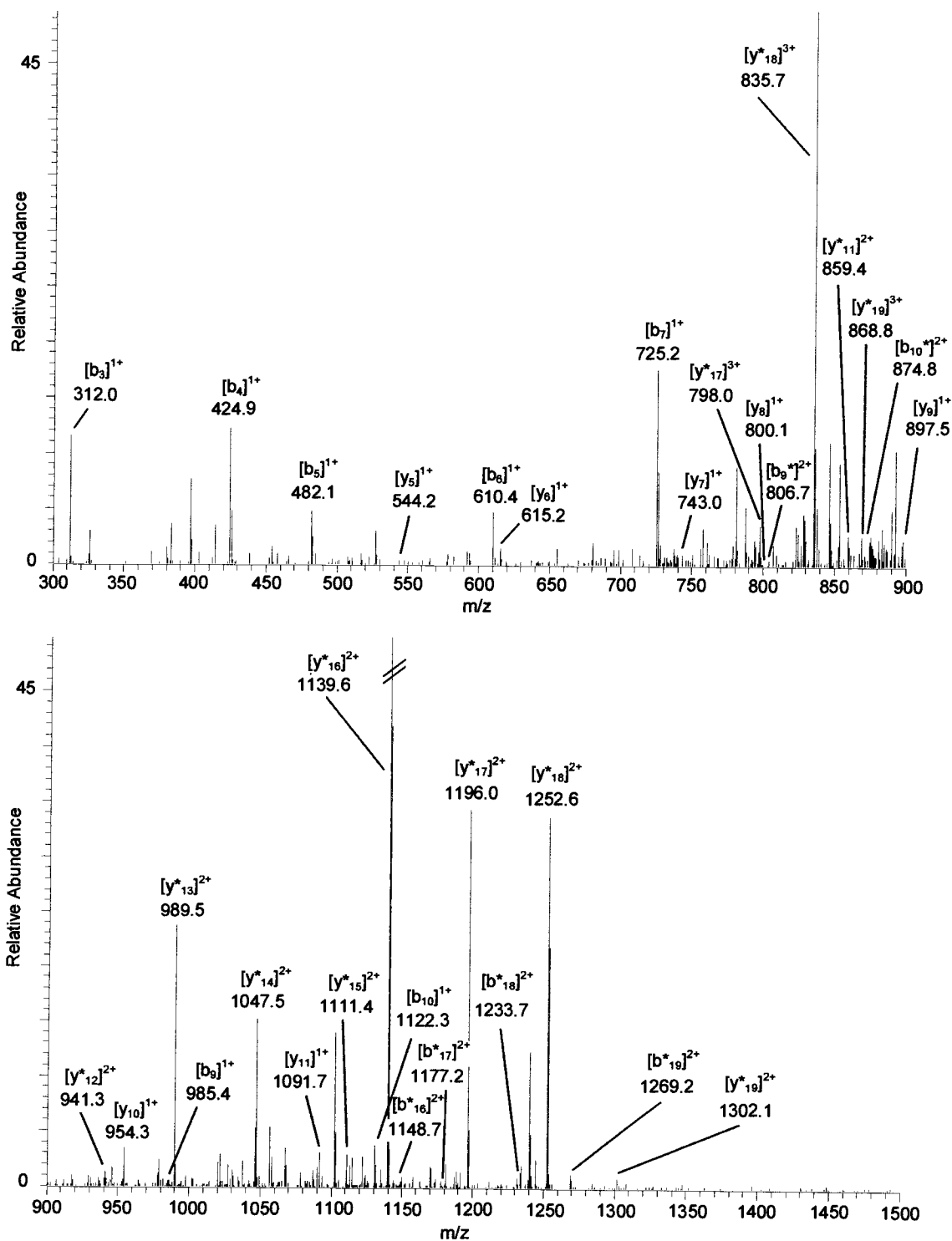


Figure 4.17. LC/ESI MS/MS spectrum of $T6_{(-3)} \times dT_2$.

b_{16}^* through b_{19}^*), which represent those ion fragments containing the crosslinked dinucleotide, were consistent with the nucleopeptide complex being $T6_{(-3)} \times dT_2$. The positions of covalent crosslinks within the peptide were determined from the difference in mass between y_{11}^* and y_{10} and the difference in mass between y_{11} and y_{12}^* . These mass differences, respectively, corresponded to Tyr₆₆ and His₆₇ crosslinked to dT_2 . The mass difference between b_{10}^* and b_9 also confirmed the location of a crosslink to His₆₇ as identified in the y - and y^* -ion fragmentation series. The b_7 and b_9^* ions suggested Tyr₆₆ as a site of crosslinking, but the lack of a b_8 ion prevented confirmation. The doubly charged fragmentation spectrum of $T6_{(-3)} \times dT_2$ was almost an order of magnitude less intense than the triply charged fragmentation spectrum (Figure 4.18). This spectrum revealed three y -ions (y_6 , y_9 , y_{10}), seven y^* -ions (y_{11}^* to y_{13}^* and y_{16}^* to y_{19}^*), three b -ions (b_6 , b_7 and b_9) and three b^* -ions (b_{10}^* , b_{18}^* , and b_{19}^*), which confirmed the identity of the nucleopeptide. The mass difference between y_{11}^* and y_{10} , as well as the mass difference between b_{10}^* and b_9 , identified the site of crosslinking to His₆₇. The fragmentation pattern interpretation of the doubly charged ion species of $T6_{(-3)} \times dT_2$ supports the fragmentation pattern interpretation of the triply charged ion species in locating His₆₇ as the site of crosslinking. The data from the doubly charged ion species, however, do not support the identification of Tyr₆₆ as a crosslinked amino acid within $T6_{(-3)} \times dT_2$, as was suggested by the fragmentation pattern of the triply charged ion species.

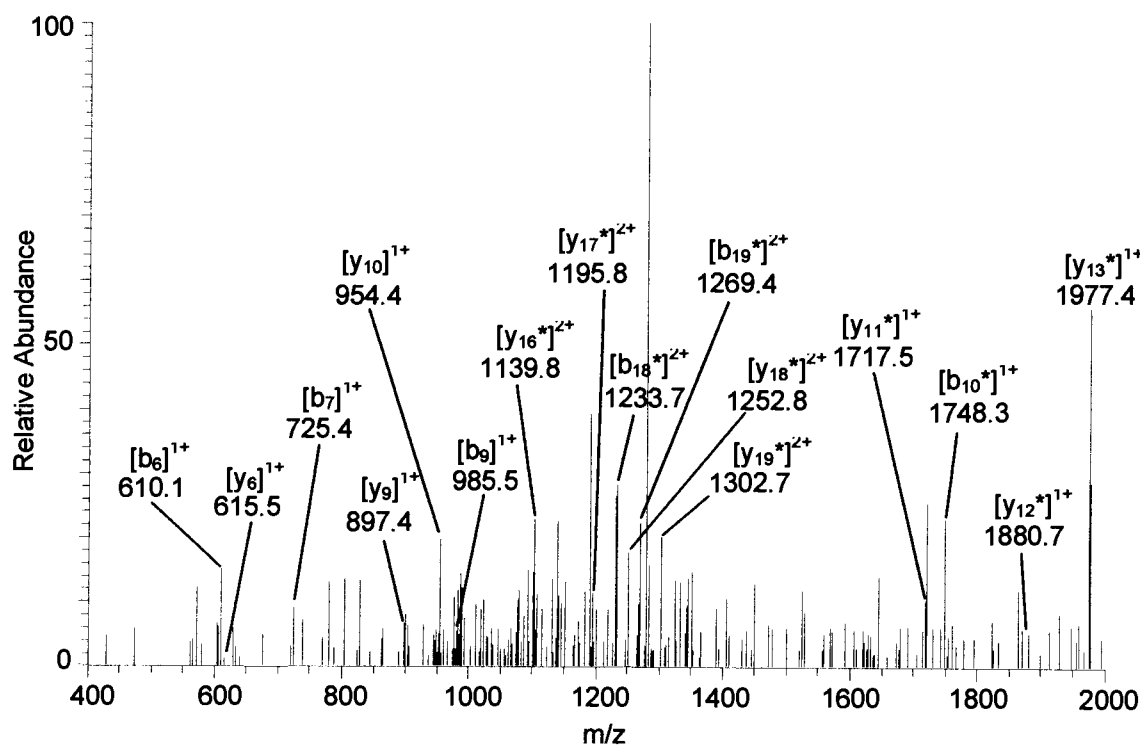


Figure 4.18. LC/ESI MS/MS spectrum of doubly charged $T6_{(-3)} \times dT_2$. Peptide sequencing of the isolated nucleopeptide species was conducted by LC/ESI MS/MS, as described under "Experimental Procedures." The doubly charged ion species of putatively identified $T6_{(-3)} \times dT_2$ ($M+2H^+=1352.7$) was selected as a precursor for tandem mass analysis. The resulting mass spectrum produced y_n - (y_6, y_9, y_{10}), y_n^* - (y_{11}^* to y_{13}^* and y_{16}^* to y_{19}^*), b_n - (b_6, b_7 , and b_9) and b_n^* -ions (b_{10}^*, b_{18}^* , and b_{19}^*) that confirmed the identity of the nucleopeptide. The mass difference between y_{11}^* and y_{10} and the mass difference between b_{10}^* and b_9 indicated the site of crosslinking as His₆₇.

4.2 Discussion

The sensitivity, the short analysis time, and the ability to simultaneously analyze complex mixtures of nucleopeptides made MALDI MS an invaluable tool over the course of these experiments. MALDI MS was used as a rapid screen of purified material to verify the presence of peptide $\times$ dT₂₀ complexes and to monitor the nuclease P1 digestion products for optimizing and determining the extent of digestion. MALDI MS was also used for the tentative identification of peptide $\times$ dT_n samples (Table 4.1). With seven ion species observed from the MALDI MS analysis of the peptide $\times$ dT₂₀ mixture, it seemed likely that a complete digestion of the nucleopeptides by concerted endo- and exonucleolytic action of nuclease P1 would have produced seven ion species; each peptide with a deoxyribothymidylic acid covalently attached. This scenario was not observed, but instead after nucleolytic digestion produced five ion series (Figure 4.2). Each ion species within a series increased in mass by one deoxyribothymidylic acid clearly indicating that these ion series were peptides crosslinked to oligonucleotides of increasing sizes. The smallest digestion product observed was a peptide crosslinked to dinucleotide. It is likely that steric interference between the peptide portion of the nucleopeptide and the nuclease may have prevented the digestion from reaching the level of a peptide crosslinked to a mononucleotide. In fact, others have reported this observation (38). The extent of nuclease digestion may have also been limited by the presence of cyclobutane pyrimidine dimers and/or 6-4 photoproducts in the dT₂₀ portion of the isolated nucleopeptides in which these UV-induced photoproducts interfered with the

action of nuclease P1 (145). Even with the heterogeneity of the nuclease digestion products, MALDI MS still provided a thorough analysis of the nucleopeptide species. Recent advances in the development of MALDI tandem time of flight (TOF/TOF) instruments will potentially allow for the complete characterization of nucleopeptide species (tentative identification, definitive identification, and locating covalent crosslinks) with a single MALDI analysis (62, 84, 102).

The data collected from MALDI MS provided *a priori* knowledge about the nucleopeptide species for nanospray/ESI MS/MS and LC/ESI MS/MS analysis. This *a priori* knowledge greatly facilitated interpreting tandem mass spectra for definitive identification of the nucleopeptide species. For this project, two types of tandem mass analyses were conducted, nanospray/ESI MS/MS and LC/ESI MS/MS. Initial tandem mass spectrometry experiments were conducted by nanospray/ESI MS/MS using a triple quadrupole instrument. This analytical methodology was tedious since the complexity of the peptide $\times$ dT_n mixtures (Figure 4.2) required prior separation and concentration by micro-HPLC before analysis. The resulting separations were never extensive enough to isolate a single nucleopeptide. At best, they were capable of producing fractions that were enriched for nucleopeptides that contained the same peptide portion, but varying sizes of oligonucleotide (Figures 4.4 and 4.6). At worst, the HPLC separations produced fractions enriched for multiple peptide species with varying lengths of oligonucleotides (Figures 4.5 and 4.8). Regardless of the degree of separation,

micro-HPLC elution fractions could be analyzed by nanospray tandem mass spectrometry.

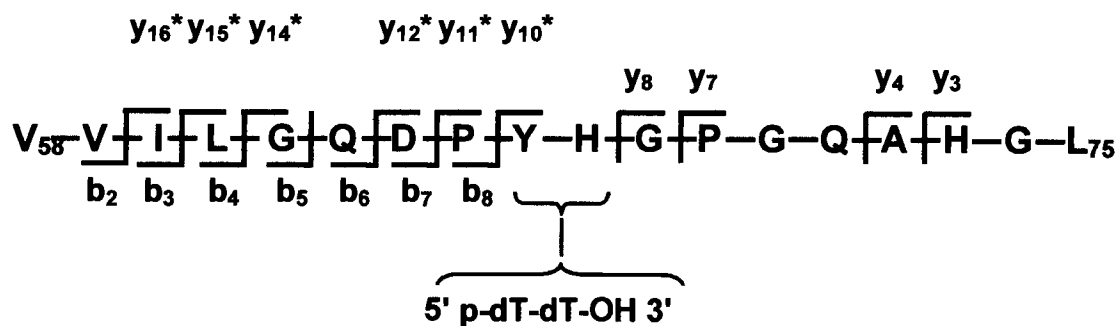
The reported virtues of nanospray tandem mass spectrometry, primarily the sensitivity of the technique towards analyzing peptides at sub-pmol quantities and sub- μ L volumes of sample (147, 148), made this an attractive form of mass spectrometry for analyzing isolated Ung-nucleopeptides. From the experiments performed here, it was found that HPLC-enriched fractions of nucleopeptides produced very simple MALDI MS spectra, but very complicated nanospray Q1 spectra (compare Figure 4.4 vs. 4.6 and Figure 4.5 vs. 4.8). The complexity of the nanospray Q1 spectra arose partly from the multiple charge states of the individual nucleopeptides that is characteristic of electrospray ionization. In general, the charge states of a nucleopeptide were observed between +2 and +4, usually with the +3 charge state being most abundant. Additional complexity in these spectra came from impurities in the sample, primarily cations of sodium and potassium. Since the HPLC-purified fractions analyzed by MALDI MS prior to nanospray analysis revealed no significant cation adduction, the strong metal adduction observed in the nanospray Q1 spectra was most likely a result of contamination of the sample while in the nanospray needle. Metal ions could have been leached from the needle by the acidified solvent into the sample. The combination of multiple charge states and cation adduction greatly reduced the effective concentration of a nucleopeptide species that could be selected as a precursor ion for nanospray tandem mass spectrometry experiments.

Nanospray/ESI MS/MS spectra were collected and interpreted for the nucleopeptides $T6_{(-7)} \times dT_2$, $T6_{(-5)} \times dT_2$, and $T18 \times dT_4$. Comparing these three spectra to one another shows that the information content of the spectra for $T6_{(-7)} \times dT_2$ and $T6_{(-5)} \times dT_2$ is distinctly different than that of the spectrum of $T18 \times dT_4$ (Figure 4.19). Triply charged precursors of both $T6_{(-7)} \times dT_2$ and $T6_{(-5)} \times dT_2$ produced b_n^- , y_n^- , y_n^* -ions during nanospray tandem mass spectrometry experiments (Figures 4.9, 4.10, 4.19A, and 4.19B). This allowed for the definitive identification of these nucleopeptides and localized the position(s) of crosslinking to Tyr₆₆ and His₆₇ for $T6_{(-7)} \times dT_2$ and Pro₆₅, Tyr₆₆, and His₆₇ for $T6_{(-5)} \times dT_2$. Oddly, b_n^* -ions were absent from both spectra, for which no explanation can be given. In contrast to these two nucleopeptides of T6, the doubly charged ion species of $T18 \times dT_4$ produced the strongest ion signals from fragmentations taking place along the oligonucleotide portion of the nucleopeptide (Figures 4.7 and 4.19C). Of the diagnostic ions required to identify the peptide portion of the nucleopeptide and the site(s) of crosslinking, only y_n^- and b_n^- -ions were present. These fragmentation ions were very weak in intensity, but were adequate to definitively identify this nucleopeptide. The lack of b_n^* - and y_n^* -ions prevented the identification of crosslinked amino acids within the nucleopeptide.

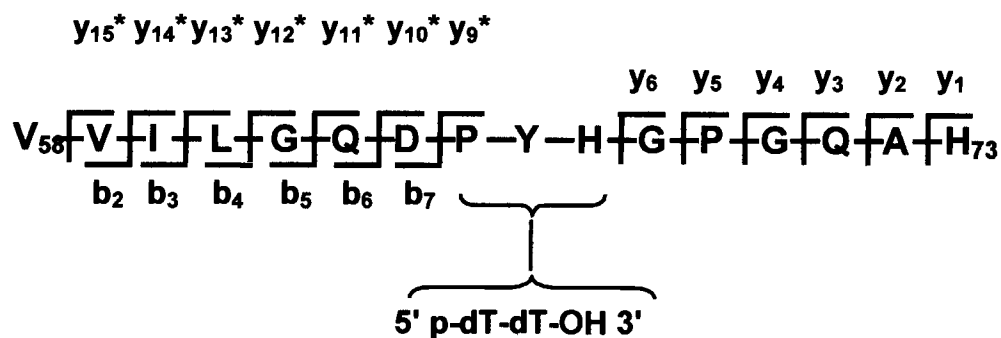
It became obvious from conducting multiple nanospray experiments that tandem mass spectrometry should be conducted via LC/ESI MS. In-line HPLC purification and mass spectrometric analysis of the nucleopeptides would be considerably less tedious and time consuming and elution peaks would be directly eluted into the mass spectrometer to minimize cation adduction.

Figure 4.19. Summary of fragmentation patterns observed in nanospray/ESI MS/MS spectra of Ung nucleopeptides. Peptide (b- and y-ions) and nucleopeptide (b*- and y*-ions) fragmentation patterns of Ung-peptide $\times$ dT_n complexes from Figures 4.7, 4.9, and 4.10 are summarized. For clarity, fragmentation taking place along the oligonucleotide portion of the nucleopeptide has been omitted. **(A)** The fragmentation pattern of triply charged T6₍₋₅₎ $\times$ dT₂ showed a series of b-ions (b₂ to b₈) that identified the nucleopeptide species. A continuous conversion of a y- to y*-ion (y₈ to y₁₀*) was observed localizing the position of the crosslink(s) between Try₆₆ and His₆₇. Fragmentation corresponding to b*-ions were not observed. **(B)** The fragmentation pattern of the triply charged T6₍₋₇₎ $\times$ dT₂ produced b- (b₂ to b₇) and y-ions (y₁ to y₆) that identified this nucleopeptide. A non-continuous conversion of y- to y*-ions (y₆ to y₉*) localized the sites of crosslinking to Pro₆₅, Try₆₆, and His₆₇. **(C)** Analysis of the y- and b-ion series of doubly charged T18 $\times$ dT₄ confirmed the identity of the nucleopeptide, but did not locate the position of crosslinked amino acid(s). Fragmentations producing b- and b*-ions were not observed.

A.



B.



C.

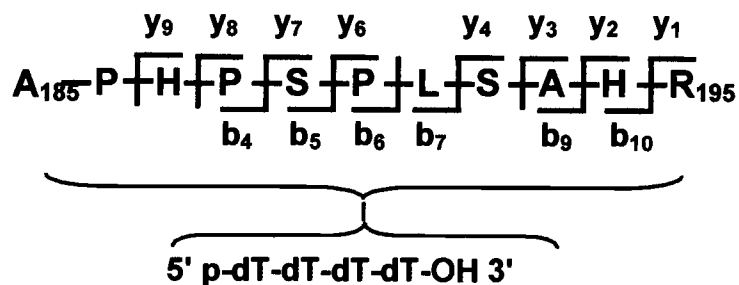


Figure 4.19. Summary of fragmentation patterns observed in nanospray/ESI MS/MS spectra of Ung nucleopeptides.

Moreover, the LC/ESI configuration would allow for multiple tandem mass spectrometric analyses of various nucleopeptides during a single chromatographic separation. Upon installation of an LCQ ion-trap mass spectrometer at the OSU Mass Spectrometry Core Facility, nucleopeptide mixtures (similar to Figure 4.2) were analyzed on this instrument in the LC tandem mass spectrometry configuration.

The complexity of the peptide $\times$ dT_n mixture is underscored by the base peak chromatogram of this sample (Figure 4.11A). The majority of the ion chromatogram peaks observed did not correspond to predicted nucleopeptides, but peaks corresponding to predicted nucleopeptides could be extracted from this chromatogram (Figure 4.11B through E) for subsequent MS and MS/MS analysis. The charge state distribution of the nucleopeptide species observed during LC/ESI MS experiments were similar to those observed for nanospray/ESI experiments (charge states distributed over +2 to +4, with the +3 charge state being most abundant), but cation adduction to the nucleopeptide species was not observed (Figure 4.12). Of the five peptide species observed in the peptide $\times$ dT_n MALDI MS spectrum, four of them were characterized during a single LC MS/MS analysis. The missing nucleopeptide, T6₍₋₅₎ $\times$ dT₂, may have been of low enough quantity that it was not observed during analysis. The ease of conducting MS/MS analysis and the ability to analyze an entire mixture of nucleopeptides in a single LC MS/MS analysis made the LC format of tandem mass spectrometry the preferred method for characterizing nucleopeptides.

The LC tandem mass spectrometry format was used in the analysis of doubly and triply charge nucleopeptides of $T6_{(-3)} \times dT_2$, $T6_{(-7)} \times dT_2$, $T11 \times dT_2$, and $T18 \times dT_2$ (Figure 4.20). The identity of these nucleopeptides was confirmed by b_n - and y_n -ions series sequence tags. Locating amino acid(s) containing a covalently attached oligonucleotide was performed by looking for the conversion of y - to y^* -ions or b - to b^* -ions that would represent the mass addition of the oligonucleotide to an amino acid. Sites of crosslinking could only be definitively identified in the nucleopeptides of T6. Sites of crosslinking in the triply charged ion species of $T6_{(-3)} \times dT_2$ were identified as Tyr₆₆ and His₆₇ while a site of crosslinking in the doubly charged ion species of $T6_{(-3)} \times dT_2$ was identified as His₆₇. Both the doubly and triply charged ion species of $T6_{(-7)} \times dT_2$ showed His₆₇ as a site of crosslinking. The triply charged ion species of $T11 \times dT_2$ strongly suggested His₁₃₄ and His₁₃₆ as sites of crosslinking, but the complexity of this spectrum prevented definitive identification. A conversion of b -to b^* -ions or a conversion of y - to y^* -ions were absent from the spectra of $T18 \times dT_2$, thus preventing the localization of crosslinked amino acids.

Comparing the nucleopeptides analyzed by LC/ESI MS/MS shows that the major fragmentation pathway for the T6 nucleopeptides took place along the peptide backbone while the major fragmentation pathway for the nucleopeptides of T11 and T18 took place along the DNA backbone. Fragmentation of the DNA backbone yields no useful information in identifying the amino acid(s) involved in crosslinking, but could prove useful in identifying the DNA sequence context in which crosslinking takes place. The CID spectra of $T11 \times dT_2$ and $T18 \times dT_2$ are

Figure 4.20. Summary of fragmentation patterns observed in LC/ESI MS/MS analysis of Ung nucleopeptides. Peptide (b- and y-ions) and nucleopeptide (b*- and y*-ions) LC/ESI MS/MS fragmentation of Ung-peptide $\times$ dT_n complexes from Figures 4.13 through 4.18 are summarized. For clarity, fragmentation taking place along the oligonucleotide portion of the nucleopeptide has been omitted. **(A)** The fragmentation pattern of triply charged T6₍₋₃₎ $\times$ dT₂ showed two continuous conversions of y- to y*-ions (y₁₀ to y₁₁* and y₁₁ to y₁₂*) locating the position of two crosslinks to Tyr₆₆ and His₆₇. Fragments of b- and b*-ions showed a noncontinuous conversion of the ion series (b₇ to b₉*) that indicated the position of a covalent crosslink between Pro₆₅ and Tyr₆₆. A continuous conversion of b- to b*-ions was observed (b₉ to b₁₀*), supporting the y-ion series identification of His₆₇ as a crosslinked amino acid. **(B)** The doubly charged T6₍₋₃₎ $\times$ dT₂ ion species showed a conversion of y- to y*-ions (y₁₀ to y₁₁*) and b- to b*-ions (y₉ to y₁₀*) in which the site of crosslinking was located to His₆₇. An additional site of crosslinking to Tyr₆₆ could not be identified. **(C)** The fragmentation pattern of the triply charged T6₍₋₇₎ $\times$ dT₂ ion species showed a single continuous conversion of y- to y*-ions (y₆ to y₇*) locating the site of crosslinking to His₆₇. A noncontinuous conversion of b- to b*-ions (b₉ to b₁₂*) supports His₆₇ as the site of crosslinking, but does not directly identify this amino acid as being crosslinked. **(D)** The fragmentation pattern of the doubly charged T6₍₋₇₎ $\times$ dT₂ ion species also showed a single continuous conversion of y- to y*-ions (y₆ to y₇*) locating the site of crosslinking to His₆₇, but b-series fragmentation ions were absent from the spectrum. **(E)** The triply charged T11 $\times$ dT₂ fragmentation pattern showed one continuous conversion of y- to y*-ions (y₁₃ to y₁₄*) and one continuous conversion of b- to b*-ions (b₆ to b₇*). These conversions, respectively, suggested His₁₃₄ and His₁₃₆ as the sites of crosslinking. The combination of b- and b*-ions (b₈ to b₁₁ and b₈* to b₁₂*) also indicates another crosslinked amino acid(s) between Ala₁₃₇ and Trp₁₄₁. **(F)** Analysis of fragments of T18 $\times$ dT₂ confirmed the identity of the nucleopeptide and overlap of y- and y*-ions suggests multiple sites of crosslinking across the peptide. ♦ in the positions for y₇* and y₈* indicate that the assignments of these ions is uncertain because of possible isobaric conflicts.

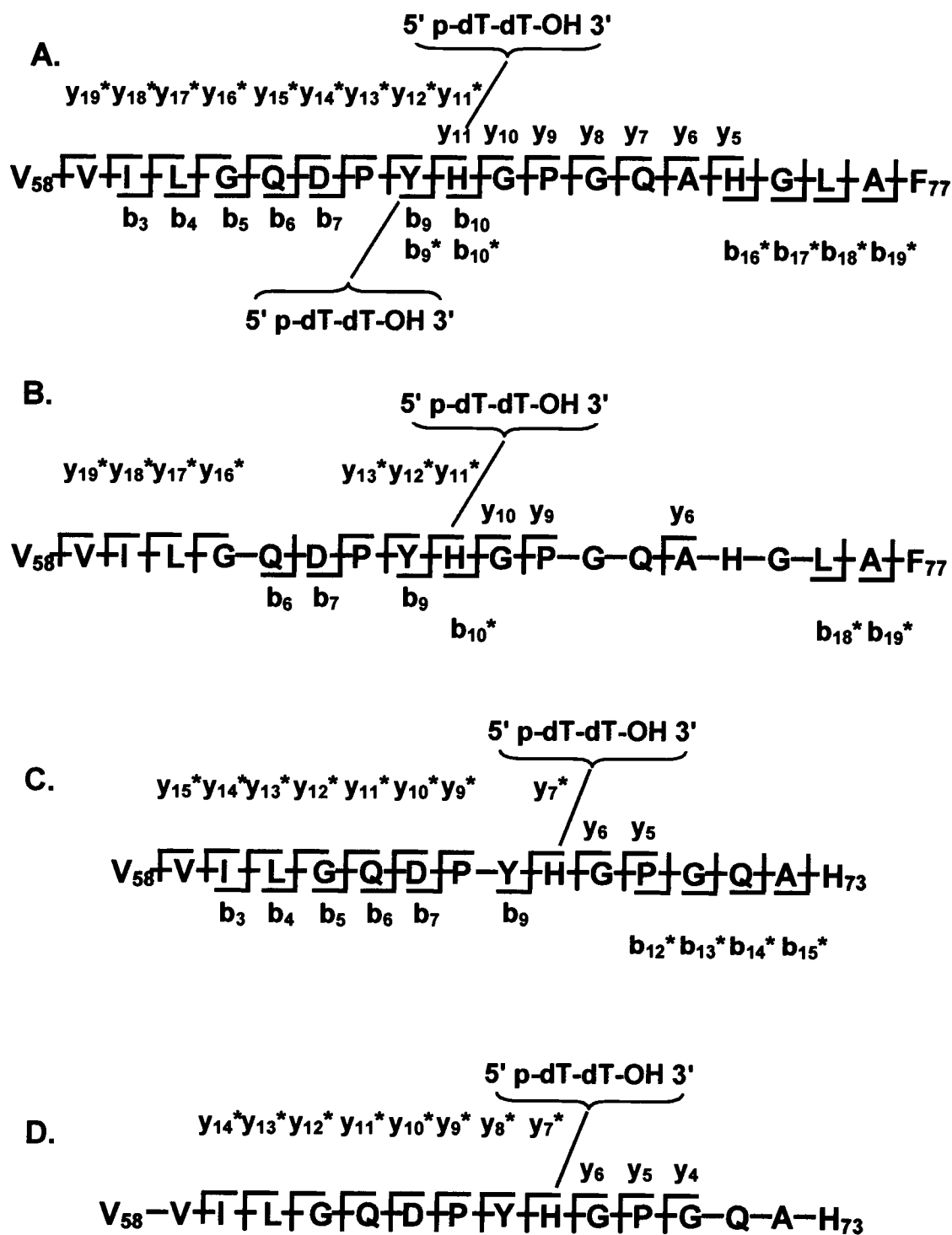


Figure 4.20. Summary of fragmentation patterns observed in LC/ESI MS/MS analysis of Ung nucleopeptides.

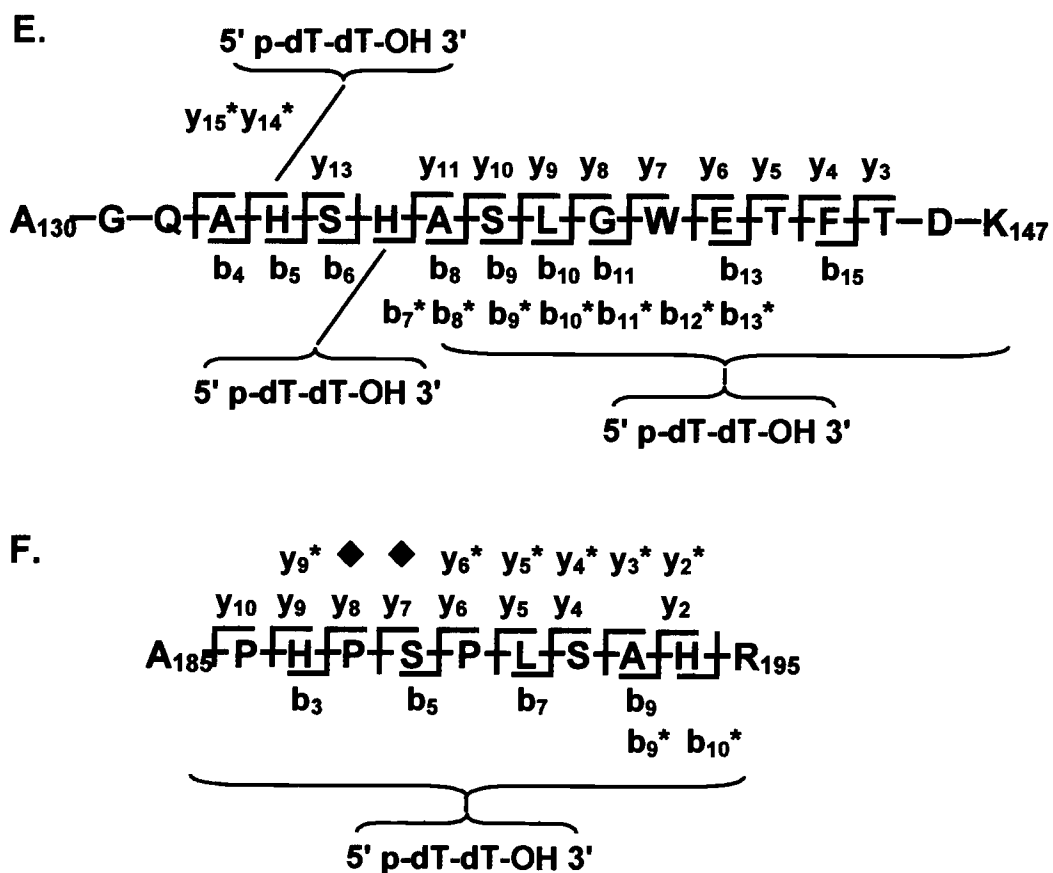


Figure 4.20 (continued). Summary of fragmentation patterns observed in LC/ESI MS/MS analysis of Ung nucleopeptides.

considerably more difficult to interpret since peptide ion series (b- and y-ions) overlap crosslink ion series (b*- and y*-ions). Three possible routes may have produced this overlap. First, varying populations of a nucleopeptide may have undergone internal fragmentation in which one cleavage takes place at the covalent crosslink and another takes place along the peptide backbone. Resulting fragment ions from the population of crosslink ions added to the population of peptide ions would yield the observed overlap. A second possible route is that multiple sites of crosslinking are present within a nucleopeptide population such that a site of crosslinking might be present at various amino acids. The resulting CID spectrum for this heterogeneous structural population would have the potential to produce overlaps in the peptide and crosslink ion series. A third possible route is that a nucleopeptide contained multiple sites of crosslinking and a population of the nucleopeptide underwent cleavage at the crosslink with subsequent cleavage of the peptide while a population of the nucleopeptide only underwent cleavage along the peptide backbone of the nucleopeptide.

It is interesting to note that $T11 \times dT_2$ and $T18 \times dT_2$ are true tryptic peptides in the sense that they contain a C-terminal arginine or lysine. The T6 nucleopeptides were generated from a tryptic digestion of Ung, but lost three, five, or seven C-terminal amino acids during work-up of the sample and therefore do not contain a fixed charged amino acid at the C-terminus. The two nucleopeptides that contained C-terminal-fixed charged amino acids showed directed fragmentation towards the DNA backbone whereas the nucleopeptides

that had no fixed charges at the C-terminus showed directed fragmentation towards the peptide backbone. The preference of fragmentation towards the oligonucleotide portion versus the peptide portion of a nucleopeptide has been reported previously (38). A likely explanation for the observed nucleopeptide fragmentation pathways comes from CID studies of multiply-protonated peptides (21, 30, 31, 132, 133). From these studies it has become generally accepted that highly basic sites (such as arginine and lysine) act in immobilizing protons on the peptide. Mobilization of protons to less basic sites along the peptide backbone is required in order to promote peptide fragmentation mechanisms under low-energy CID. In the case of nucleopeptides of T11 and T18, it is possible that the highly basic C-terminal amino acids were able to anchor an ionizing proton causing Coulombic repulsion of additional protons to be directed from the peptide portion to the oligonucleotide portion of the nucleopeptides. The repelled protons become mobile along the oligonucleotide inducing fragmentation (43). Without the highly basic amino acids of arginine or lysine present in the nucleopeptides of T6, ionizing protons are mobile to protonate along the peptide backbone of these nucleopeptides to induce the fragmentation pathways that produced b-, y-, b*- and y*-ion series.

The complexity of the CID spectra observed in this work is inherent to the nature of a nucleopeptide. The peptide portion of a nucleopeptide has three potential sites of cleavage per peptide bond that can produce six types of fragments, namely a, b, c, x, y, and z ions (see Appendix). Fragmentation can also take place along the side chains of the peptide. The oligonucleotide portion

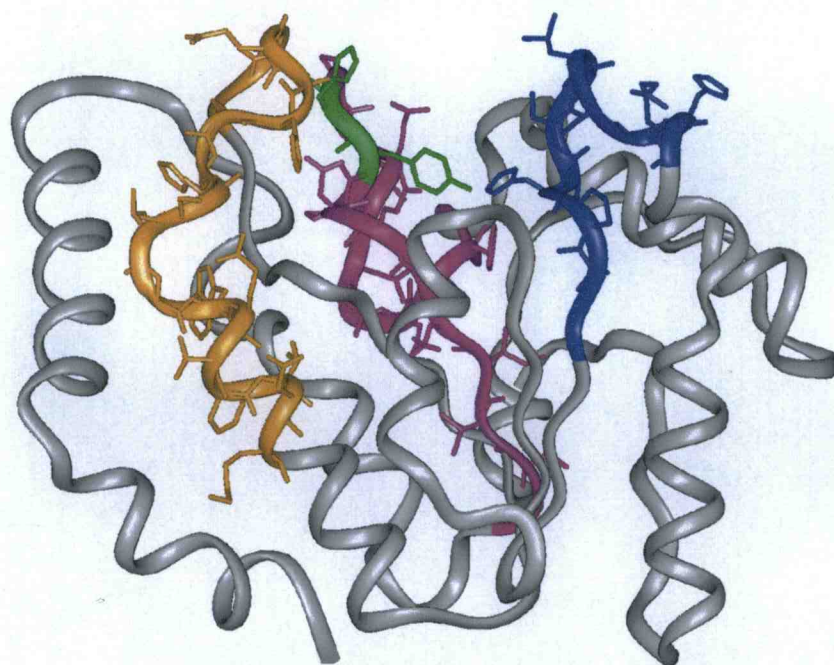
of the nucleopeptide has four potential sites of cleavage per phosphodiester bond that can produce eight types of fragments, namely *a*, *b*, *c*, *d*, *w*, *x*, *y*, and *z* ions. The multiple types of fragments within a nucleopeptide lead to a greater probability of producing isobaric fragment ions during the CID process. This was seen during the LC MS/MS analysis of T18 × dT₂ with *m/z* ion species 1393.4 and 1490.5. Ion species 1393.4 could be interpreted as either singly charged *y*₇⁺ (*MH*⁺=1393.519) or singly charged *z*₁⁺ (*MH*⁺= 1393.697). Likewise, ion species 1490.5 could be interpreted as either singly charged *y*₈⁺ (*MH*⁺=1490.571) or singly charge *w*₁⁺ (*MH*⁺=1490.666). To distinguish between these isobaric fragments or resolve these fragments, a mass accuracy of about 10 parts per million and resolving powers of about 10,000 would be required. These specifications are outside the capabilities of the LCQ ion-trap instrument, but could be achieved with either MALDI FTICR or ESI Q-TOF instruments. Additional ambiguity can arise in the nucleopeptide CID spectra from the sequence context of the peptide. Two of the nucleopeptides analyzed in this study contain multiple proline residues, which can directly influence the CID fragmentation of peptides. Cleavage of the proline amide bond is highly favorable and often dominates the CID spectra of peptides that contain this amino acid, while *b*-ions formed from the cleavage at the C-terminal side of the proline are generally weak or absent (2, 54, 106, 133). The absence of *b*₈-ions for the nucleopeptides of T6 (Figures 4.19B and 4.20A-C) is consistent with this expectation. Interestingly, the nucleopeptide of T18 contains three prolines that

alternate from the N-terminus. The lack of b-ions in this spectrum may be a result of the "proline effect" (Figure 4.20F).

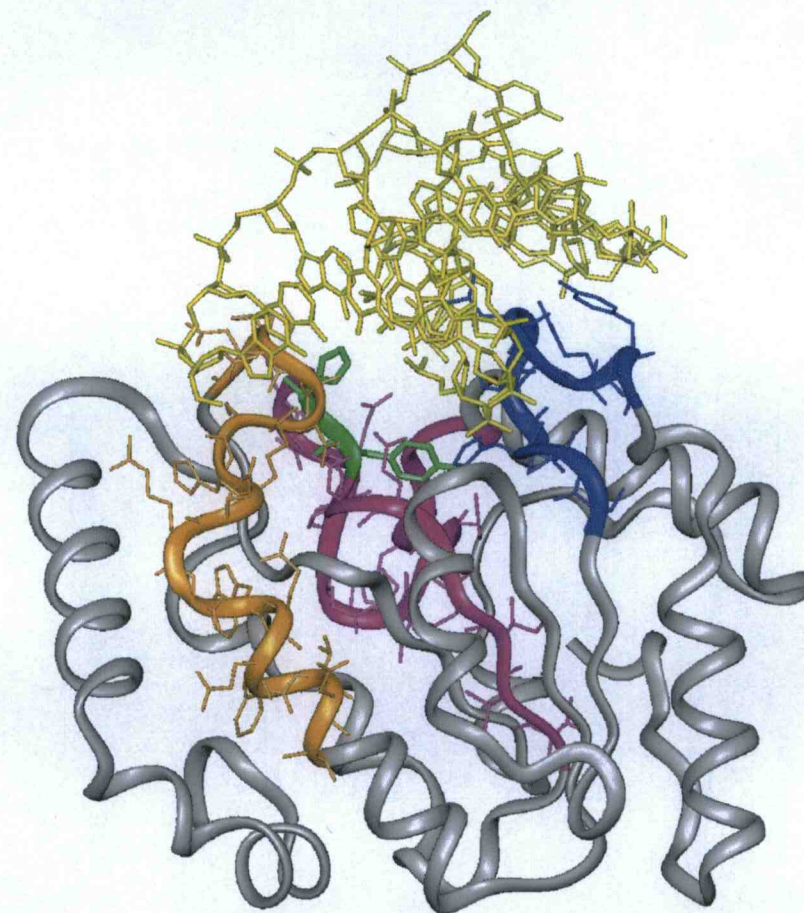
Along with characterizing the fragmentation pattern of nucleopeptides, an equally important objective of this research was to identify regions of Ung that take part in crosslinking and to locate specific amino acids within those regions that are photochemically crosslinked. Three tryptic peptides, T6, T11, and T18, were definitively identified by tandem mass spectrometry to be involved in crosslinking. Within those peptides, T6 amino acid residue His₆₇ was unambiguously identified as taking part in crosslinking (Figure 4.21). Other sites of crosslinking were tentatively identified as Tyr₆₆ within T6 and His₁₃₃ and His₁₃₅ within T11.

Peptides T6 and T18 contain highly conserved amino acid sequences that have been identified as important regions for either DNA binding or catalytic activity of human UDG. The water-activating loop of the protein, containing the DPYH region (an identical region between Ung and UDG), is located in the T6 peptide of Ung. Along with its catalytic importance, this region also contains the specificity-determining tyrosine. The orientation in the catalytic pocket of this amino acid side has been shown to exclude thymine in human UDG (64, 91). One of three Ser-Pro loops that takes part in compressing DNA in the "pinch-push-pull" mechanism of the proteins is located in the T18 peptide of Ung (APHPSPLSAHR). The serines in this region have been proposed to take part in orienting the protein for DNA scanning. Also located within this region is the leucine involved in minor groove reading (96). Based on these crystallographic

Figure 4.21. Correlation of LC/ESI MS/MS nucleopeptide mapping with the crystal structures of uracil-DNA glycosylase. Peptides of *E. coli* uracil-DNA glycosylase identified by LC/ESI MS/MS as taking part in crosslinking are shown in the crystal structure of the enzyme (**A**). Based on the sequence homology between *E. coli* Ung and human Ung, regions of the *E. coli* crystal structure that were identified as being crosslinked are highlighted in the crystal structure of the human enzyme co-crystallized with DNA (96) (**B**). Residues corresponding to T6 are highlighted in purple, T11 residues are in orange, T18 residues are in blue, Try₆₆ and His₆₇ (*E. coli* Ung) and Try₁₄₇ and His₁₄₈ (human Ung) are in green, and DNA is in yellow.



A.



B.

Figure 4.21. Correlation of LC/ESI MS/MS nucleopeptide mapping with the crystal structures of uracil-DNA glycosylase.

findings, it is reasonable that these regions of the protein took part in crosslinking.

The involvement of T11 in the crosslinking was not predictable from crystallographic data. Previous co-crystal structures of human UDG with uracil containing substrates did not identify this portion of the enzyme as being important structurally or catalytically (96, 128). Within the region defined by T11, there are three highly conserved amino acids (Ser₁₃₅, His₁₃₆, and Trp₁₄₂ in *E. coli*) which is normally an indicator of an important role in the enzyme's function. Close examination of the co-crystal structure of UDG with duplex DNA indicates that the region in this protein corresponding to Ung's T11 is too far from the duplex DNA ($>3 \text{ \AA}$) for it to take part in crosslinking. This discrepancy between crosslinking of T11 in Ung and the crystallographically predicted non-binding behavior of T11 in UDG can be explained in two ways. First, structural anomalies in the human Ung/DNA co-crystal structure, such as crystal packing effects, could have resulted in the region of the protein being unnaturally shifted away from the DNA (or the DNA unnaturally shifted away from the protein). Second, since a crystallographic structure represents a static molecule/complex, it is also possible that conformational changes that take place in the protein during its binding to DNA may shift this region of the protein to be in sufficient contact for the formation of a crosslink. This potential conformational change could go undetected/predicted based on the interpretation of the static crystal structure.

5. SULFONIC ACID DERIVATIZATION OF PEPTIDES FOR ENHANCED SEQUENCING BY ELECTROSPRAY IONIZATION MASS SPECTROMETRY

A common method for producing peptide sequence mass spectra is collision-induced dissociation (CID) within an ion-trap or within the collision cell of an electrospray ionization (ESI) mass spectrometer. Sequence spectra of peptides produced from these instruments are often complex due to the nature of the peptide backbone. Three types of bonds are present in the peptide backbone that can undergo cleavage to produce six possible fragments per amino acid residue; three containing the N-terminus (a-, b-, and c-ions) and three containing the C-terminus (x-, y-, and z-ions) (9). There are other complexities associated with sequence spectra. These include internal fragmentation of the peptide, fragmentation along peptide side chains, loss of neutral species during fragmentation such as water and ammonia, and fragment ions having multiple charge states. The complex nature of these spectra often makes interpretation difficult, if not impossible.

Placement of a fixed charge at the N-terminus of a peptide has been shown to direct the fragmentation pattern of post-source decay (PSD) and CID mass spectra simplifying sequence interpretation (111). Recently, Keough and colleagues reported a derivatization method that involved the addition of a sulfonic acid group to the N-terminus of tryptic peptides (65). Addition of this fixed negative charge was able to alter the MALDI PSD fragmentation pathways in favor of abundant y-series fragment ions allowing for simple and unambiguous

de novo peptide sequencing. This peptide derivatization procedure was adapted to simplify the CID spectra produced in ESI mass spectra and its application with protein and peptide models was investigated.

5.1 Results

5.1.1 The Derivatization Reaction

Initial derivatization experiments were conducted using 2-sulfobenzoic acid cyclic anhydride or chlorosulfonylacetyl chloride as derivatizing agents similar to the procedure of Keough and colleagues (65). Chlorosulfonylacetyl chloride was found to be the derivatization reagent of choice due to its ease of handling and higher coupling efficiency. All derivatized peptides reported here were generated using the chlorosulfonylacetyl chloride reagent. Derivatization was found to be fast and simple enough to be used routinely for isolated peptides as well as mixtures of tryptic peptides produced from in-gel digestions. Sample cleanup after derivatization, beyond the capabilities of the LC system coupled in-line to the mass spectrometer, was not required. Based on the chemistry of derivatization, a derivatized peptide increases in mass by 122.0 mass units, the mass of a sulfoacetyl group (Figure 5.1). When compared to a native peptide, a mass increase should be observed for b-ions of derivatized peptides produced from CID experiments, but since the derivatization takes place at the N-terminus of a peptide no mass increase would be expected for y-ions.

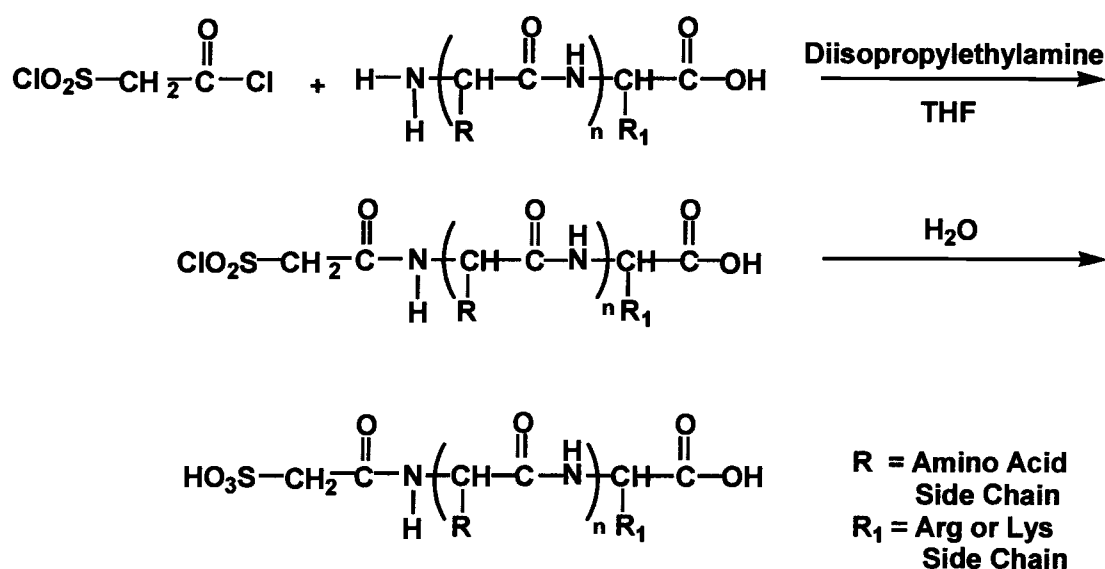


Figure 5.1 N-terminal peptide derivatization scheme. Model peptides or peptides produced from tryptic in-gel digestions of proteins were derivatized with chlorosulfonylacetate, as described under "Experimental Procedures." The derivatized peptides contain a sulfoacetate group at the N-terminus of the peptide that increased the molecular weight of a native peptide by 122.0 mass units.

5.1.2 Tandem Mass Spectra of Singly Charged, Derivatized, Tryptic Peptides

Mixtures of peptides produced from in-gel tryptic digestions of human replication protein A (hRPA) and *E. coli* uracil-DNA glycosylase (Ung) were derivatized and analyzed by LC tandem mass spectrometry. The effects of derivatization on the CID spectra of singly charged peptides were investigated. The singly charged tryptic peptide LFSLELVDESGEIR from hRPA was selected for tandem mass analysis and the resulting fragmentation spectrum produced a molecular ion and twelve fragment ions: six b-ions and six y-ions (Figure 5.2A). After N-terminal sulfonic acid derivatization, the singly charged peptide underwent fragmentation to produce an abundant, consecutive, y-ion series that covered the peptide from y₅ to the N-terminus (Figure 5.2B). The fragmentation peaks were easily identified in the spectrum and the mass difference between consecutive peaks readily yield the amino acid sequence of the peptide; thus, derivatization greatly simplified *de novo* sequencing. A less prominent ion series was also observed corresponding to a water loss of the y-ion series. Sequence information before y₅ was not present due to the low-mass range limitation of the ion-trap instrument (11). In general, native, singly charged tryptic peptides that produced mixtures of incomplete b- and y-ion series underwent extensive fragmentation after derivatization to produce abundant y-ions.

During the ion-trap tandem mass analysis of an in-gel tryptic digestion of Ung, it was found that the singly charged peptide QGVLLNTVLTVR was

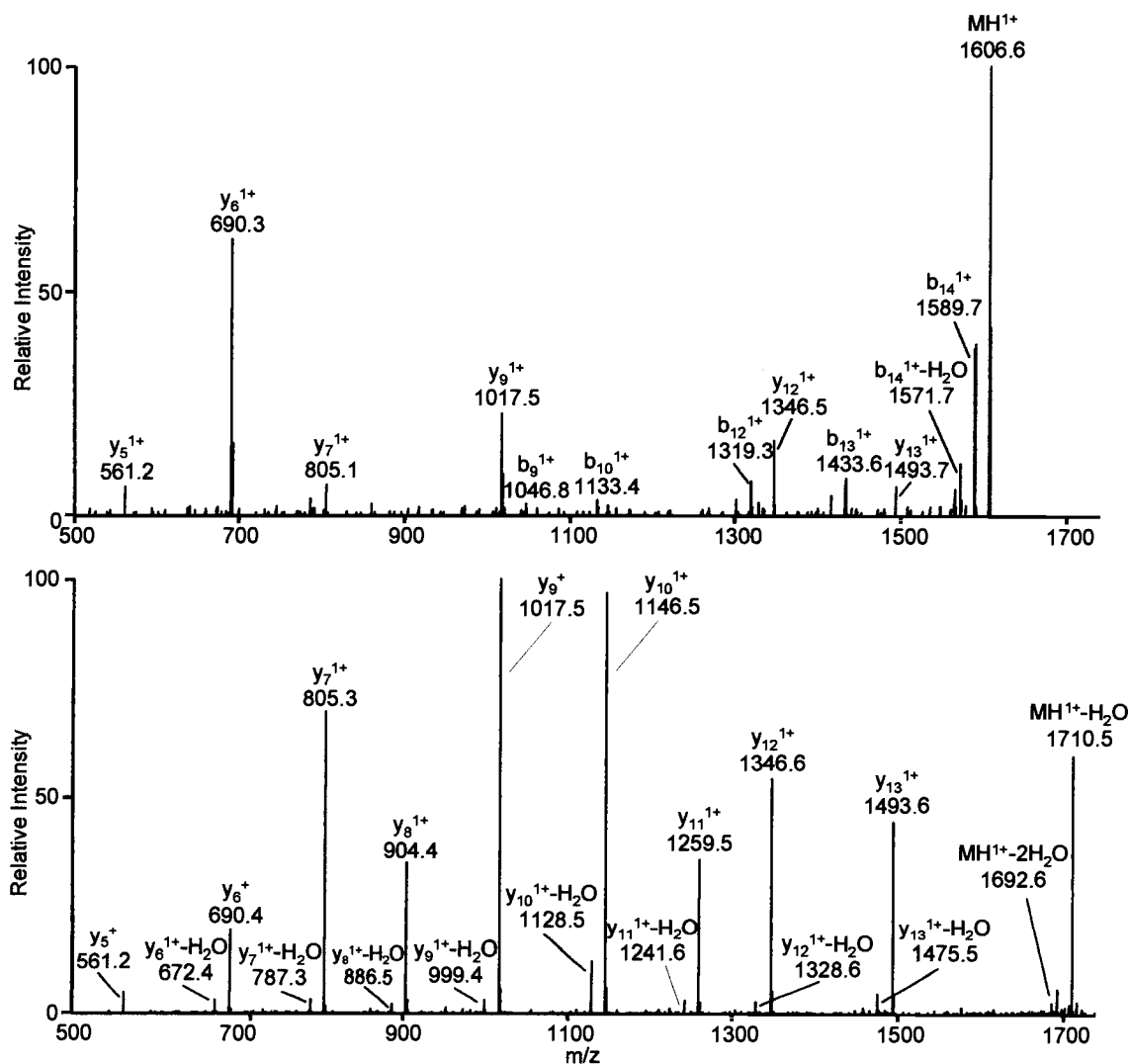


Figure 5.2. Effect of derivatization on the CID spectra of singly charged peptide LFSLELVDESGEIR. An in-gel tryptic digestion mixture of the 70 kDa subunit of hRPA was analyzed by LC ion-trap tandem mass spectrometry as described under "Experimental Procedures." The singly charged tryptic peptide LFSLELVDESGEIR was selected from the mixture for tandem mass analysis both before (A) and after (B) derivatization.

refractory to fragmentation (Figure 5.3A); this phenomenon is not uncommon for singly charged peptides. The fragmentation spectrum of this peptide showed a strong MH^+-NH_3 ion and weak signal strength for other fragmentation species that made sequence interpretation challenging. After derivatization, extensive y-series fragmentation that covered the peptide sequence from y_4 to the N-terminus was observed (Figure 5.3B). Sulfonic acid derivatization resulted in making the peptide more susceptible to fragmentation and produced a continuous y-ion series (within the mass limit of the ion-trap) with high signal to noise that allowed for unambiguous sequence determination.

A concern in derivatizing a peptide's N-terminal α -amine is the possibility of also derivatizing the lysine side chain ϵ -amine. Analysis of LC/ESI spectra of C-terminal lysine containing peptides (produced from proteolytic digestion with trypsin) indicated that the predominant species produced was mono-derivatized peptides with bis-derivatized peptides as the minor products. The retention time of a mono-derivatized peptide was usually different from its bis-derivatized counterpart making, isolation of these species for data-dependent analysis facile. Tandem mass analysis of the mono-derivatized peptide QSGVTIYPPQK showed a mass increase of the molecular ion after derivatization that was consistent with the addition of a sulfoacetyl group to the peptide (compare $MH^{1+}-NH_3$ ion species between Figure 5.4A and Figure 5.4B). The mass of the y-series fragment ions (ions containing the C-terminus) after derivatization revealed that they did not contain the derivatizing reagent, thereby locating the site of derivatization to the

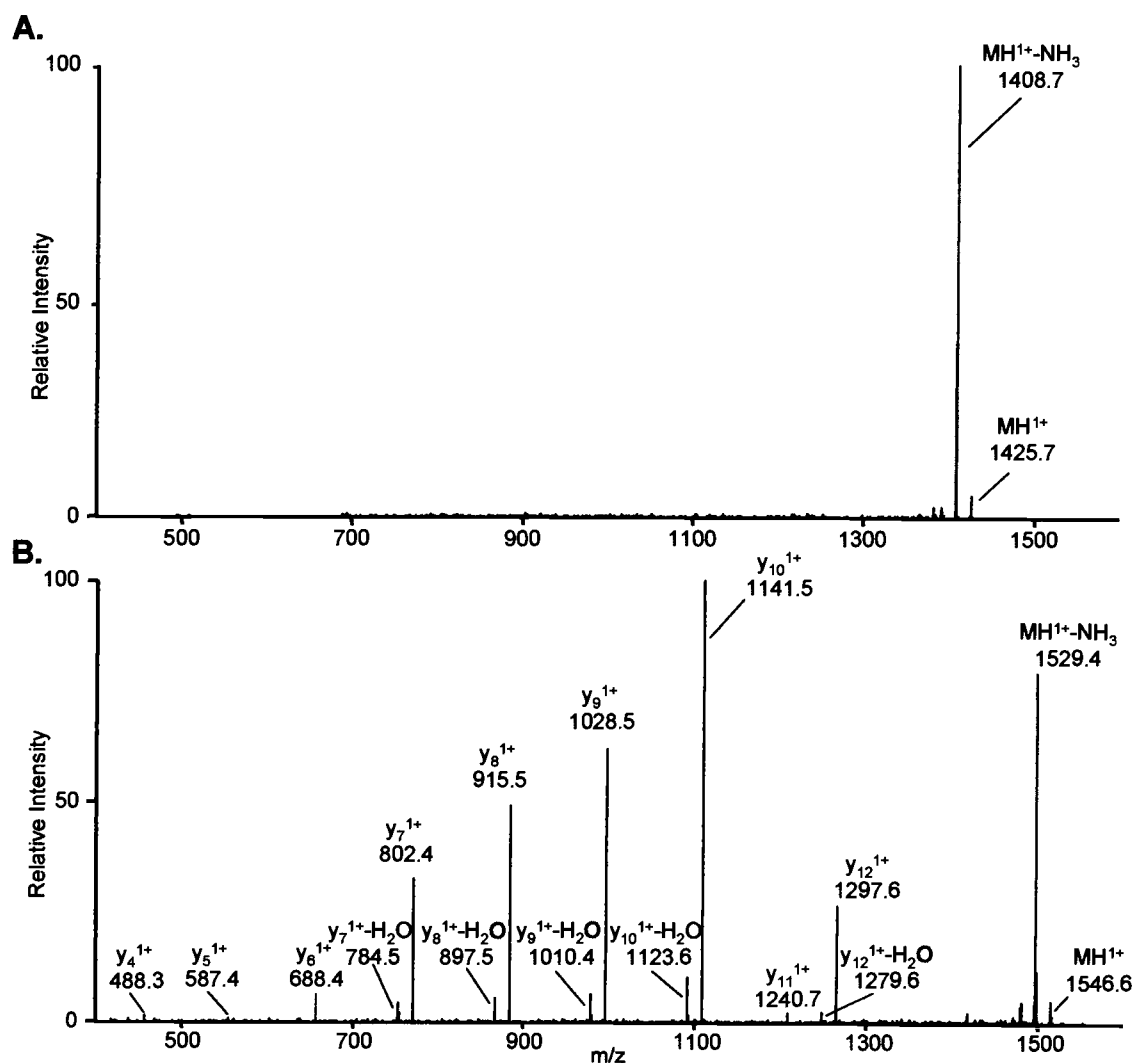


Figure 5.3. Effect of derivatization on the CID spectrum of the singly charged peptide QGVLLNTVLTVR. An in-gel tryptic digestion mixture of *E. coli* Ung was analyzed by LC ion-trap tandem mass spectrometry, as described under "Experimental Procedures." The singly charged tryptic peptide QGVLLNTVLTVR was selected from the mixture for tandem mass analysis before (A) and after (B) derivatization.

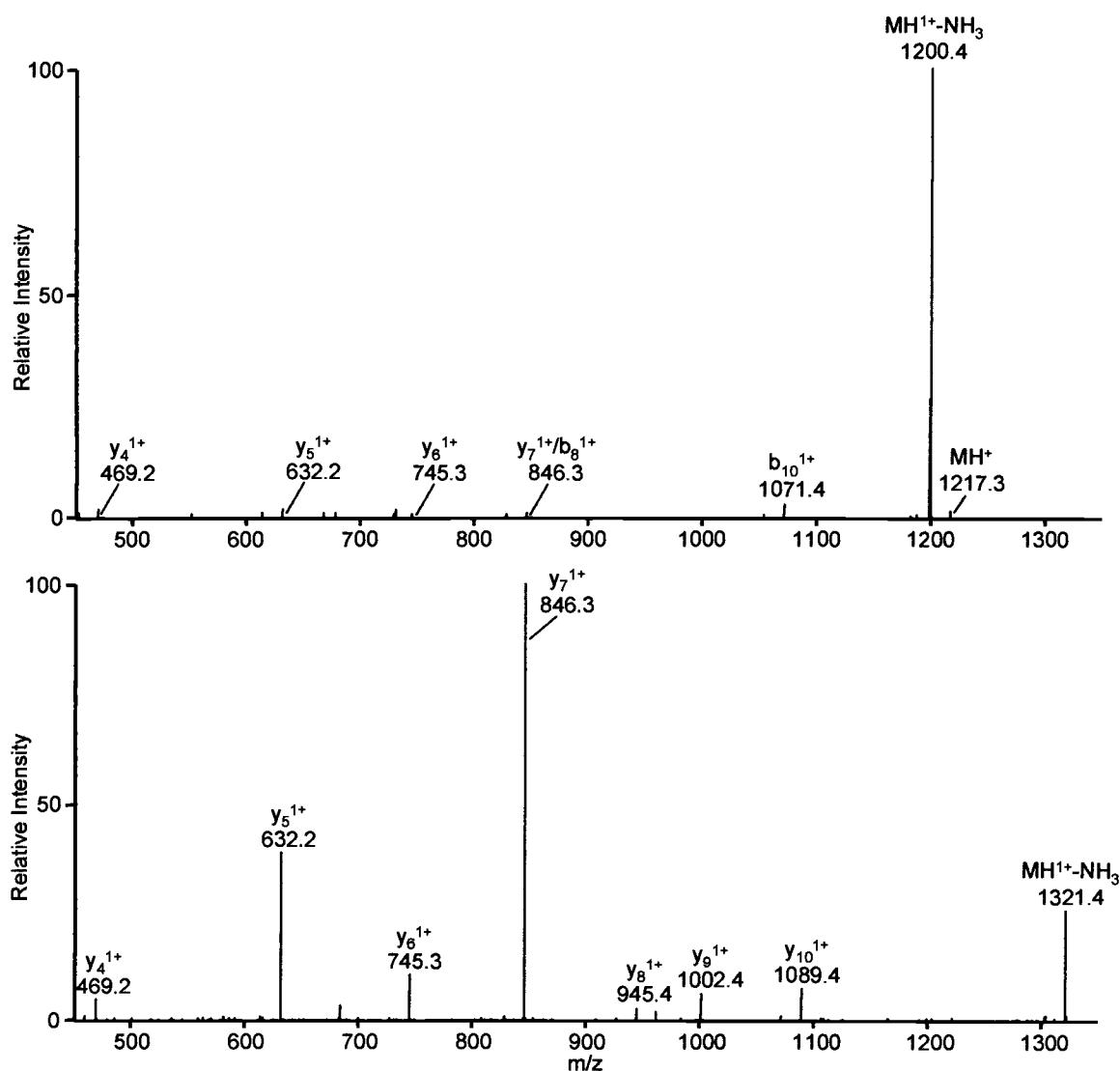


Figure 5.4. Effect of derivatization on tryptic peptides containing C-terminal lysine. An in-gel tryptic digestion mixture of *E. coli* Ung was analyzed by LC ion-trap tandem mass spectrometry, as described under "Experimental Procedures." The peptide QSGVTIYPPQK was selected from the mixture for tandem mass analysis before (A) and after (B) derivatization.

N-terminal amine rather than to lysine (Figure 5.4). The tandem mass spectra of bis-derivatized peptides showed no regular fragmentation pattern and interpretation of these spectra was not possible.

5.1.3 Tandem Mass Spectra of Doubly Charged, Derivatized, Tryptic Peptides

Peptide precursor ions produced by ESI are generally a heterogeneous mixture of multiply charged ions. For tryptic peptides containing basic amino acids the ion formation tendency is usually towards a greater population of ions in higher charge states. The higher the charge state of an ion, the higher the kinetic energy and thus the more energy available during the CID process (106). The effect of the increased charged state on the fragmentation spectra of sulfonic acid-derivatized tryptic peptides was examined. The Ung tryptic peptide ELENTIPGFTRPNHGYLESWAR was selected for analysis because the precursor ion was doubly charged. This peptide was produced from incomplete digestion of the protein, a common occurrence that results from the protease not recognizing a cleavage site. The LC tandem mass spectrum of this peptide showed a strong, doubly charged molecular ion and weak, non-consecutive fragment ions that made it difficult to identify a sequence tag (Figure 5.5A). Tandem mass analysis of the doubly charged precursor after derivatization showed that the molecular ion had undergone much more extensive fragmentation and the y-ion fragments were significantly enhanced (Figure 5.5B).

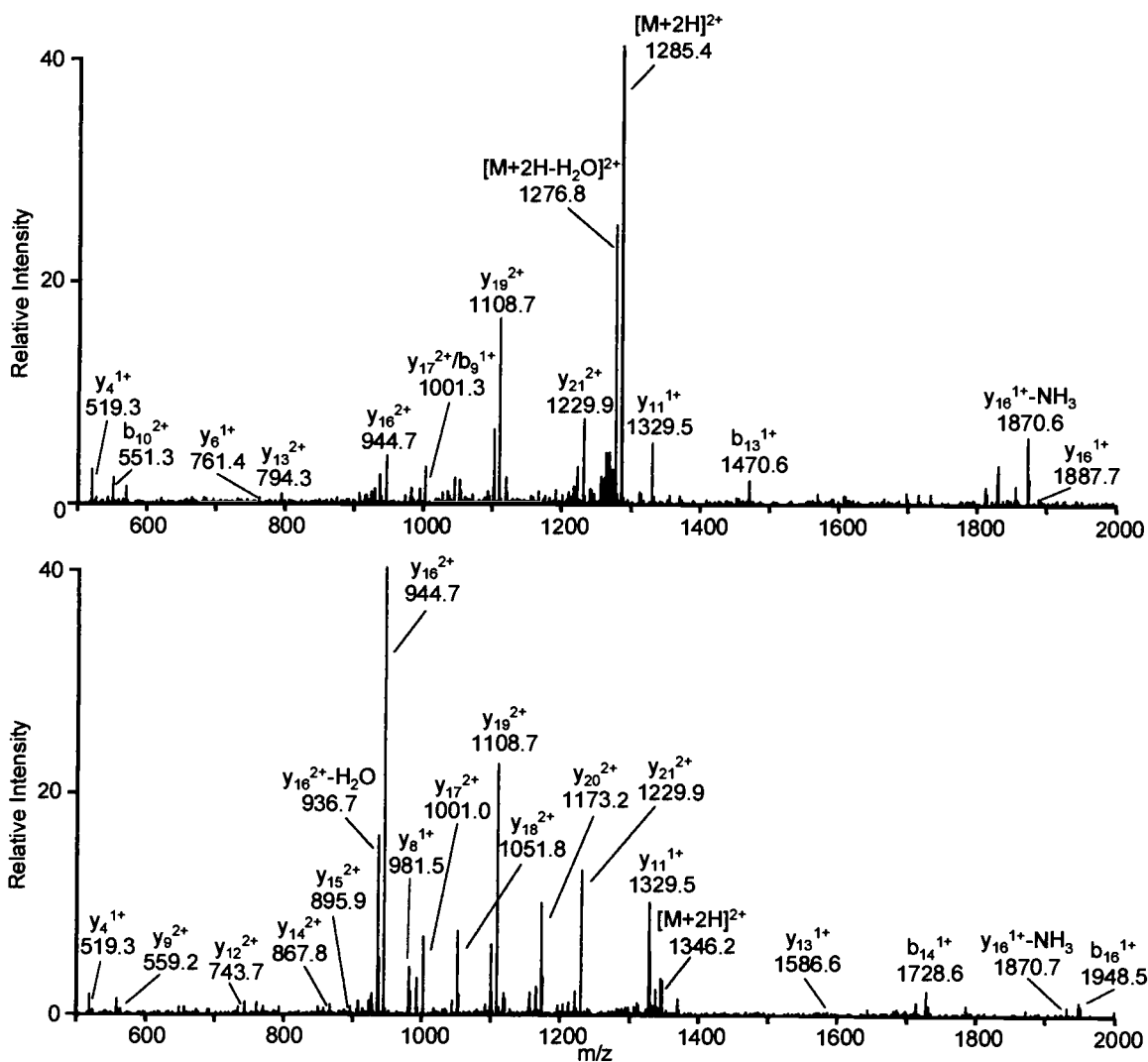


Figure 5.5. Effect of derivatization on the CID spectrum of the doubly charged peptide ELENTIPGFTRPNHGYLESWAR. An in-gel tryptic digestion of *E. coli* Ung was analyzed by LC ion-trap tandem mass spectrometry, as described under "Experimental Procedures." The doubly charged tryptic peptide ELENTIPGFTRPNHGYLESWAR was selected from the mixture for tandem mass analysis both before (A) and after (B) derivatization.

The strong, consecutive y-ion series from y_{16} to y_{21} produced a sequence tag that clearly confirmed the identity of the peptide.

5.1.4 Tandem Mass Spectra of Derivatized, Non-tryptic Peptides

The peptide angiotensin (DRVYIHPFHL) was used as a model to test the effect of sulfonic acid derivatization on the fragmentation spectra of non-tryptic peptides. A doubly charged precursor ion of the native peptide was selected for fragmentation and showed a total of seven fragmentation ions, four b-ions and three y-ions (Figure 5.6A). Derivatized angiotensin was also selected as a doubly charged precursor and the resulting fragmentation spectrum showed a total of sixteen fragment ions: eight b-ions and eight y-ions (Figure 5.6B). For this model non-tryptic peptide, derivatization did not enhance a specific fragmentation group (e.g. solely y-ions) but enhanced fragmentation as a whole (enhanced both b- and y-ions). Even though the derivatized fragmentation spectrum is complicated by the overlap of multiply charged b- and y-ions, computerized database algorithms, such as SEQUEST (33, 155-157), can easily deconvolute the spectrum to aid in identifying the peptide.

5.2 Discussion

Sulfonic acid derivatization of the N-terminus of peptides was used on model peptides and on tryptic peptides produced from in-gel digests of the proteins Ung and hRPA. Similar to the results of Keough and coworkers (65),

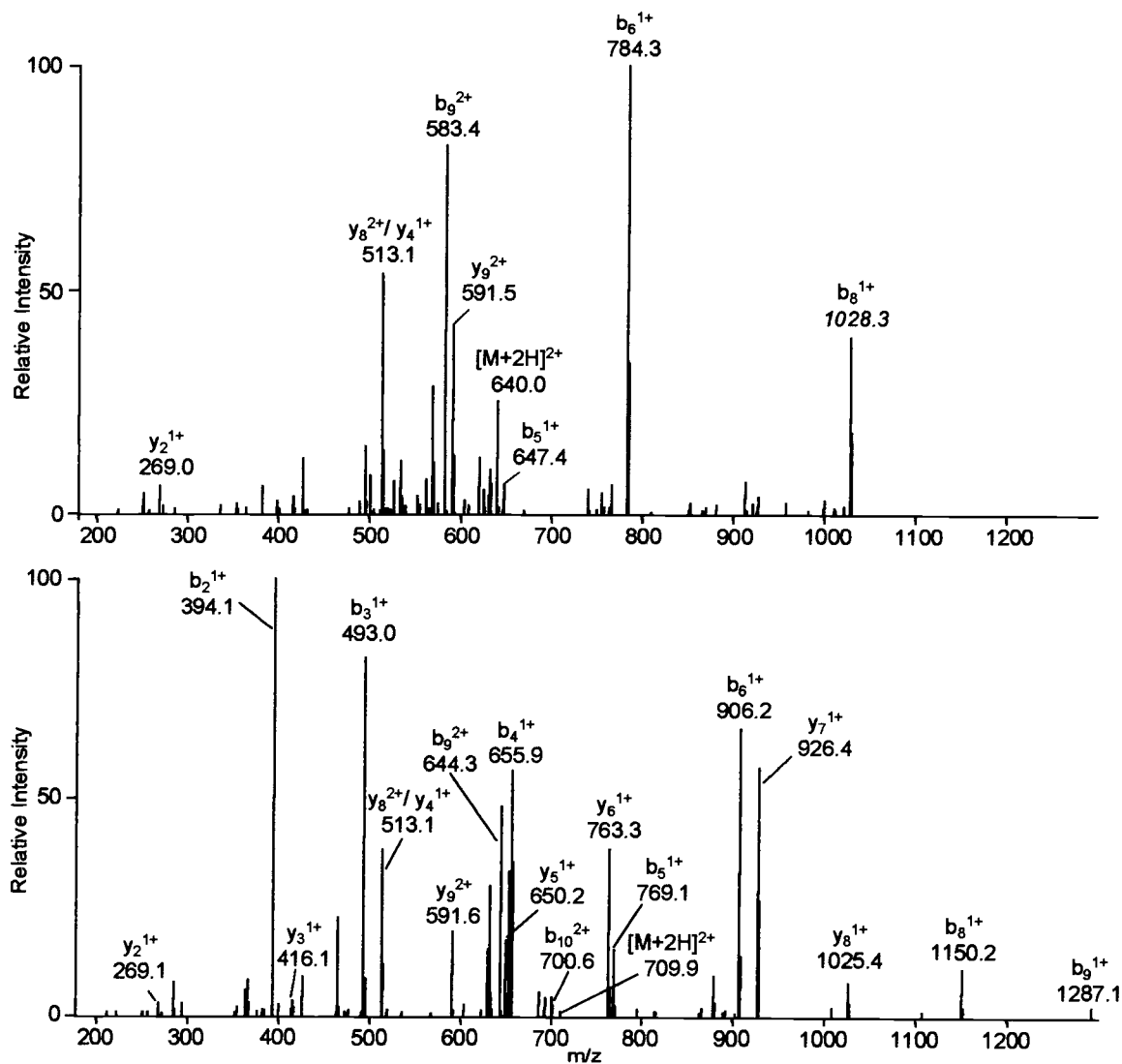


Figure 5.6. Effect of derivatization on the CID spectrum of doubly charged angiotensin. The model peptide angiotensin (DRVYIHPFHL) was analyzed by LC ion-trap tandem mass spectrometry, as described under "Experimental Procedures." The doubly charged precursor ion was selected for tandem mass analysis before (A) and after (B) sulfonic acid derivatization.

electrospray CID spectra of derivatized tryptic peptides showed an enhancement in mainly y-ion signals and a loss of b-ion signals in comparison to CID spectra of nonderivatized tryptic peptides that produced multiple ion types. This finding can be explained based on the proposed fragmentation pathways of tryptic peptides. It is generally accepted that tryptic peptides undergo protonation at the C-terminal basic amino acid (either lysine or arginine). An additional proton yields a heterogeneous population of structures resulting from rapid intramolecular proton transfer across the peptide backbone (21, 133). If Coulombic repulsions are strong enough, the additional proton needed to ionize the peptide backbone may not readily bind, preventing fragmentation of the peptide. It is for this reason that singly charged tryptic peptides often resist low energy CID fragmentation whereas doubly charged tryptic peptides fragment easily. During ionization, the basic C-terminal amino acid of derivatized peptides is protonated first (Figure 5.7). With the sulfoacetyl group present at the N-terminus of the peptide, the deprotonated derivatization group forms a zwitterionic peptide species with a negative charge at the N-terminus and a positive charge at the C-terminus. An additional proton is then needed to ionize the peptide. The ionizing proton can freely migrate across the peptide backbone to ionize amide groups since the most basic residue at the C-terminus is already occupied, promoting thorough fragmentation of the peptide. The resulting b-ion fragments produced will be in a zwitterionic form and the resulting y-ion fragments will contain a fixed charge at the basic C-terminal amino acid. Overall, this will suppress the presence of b-type ions and enhance the presence of y-type ions.

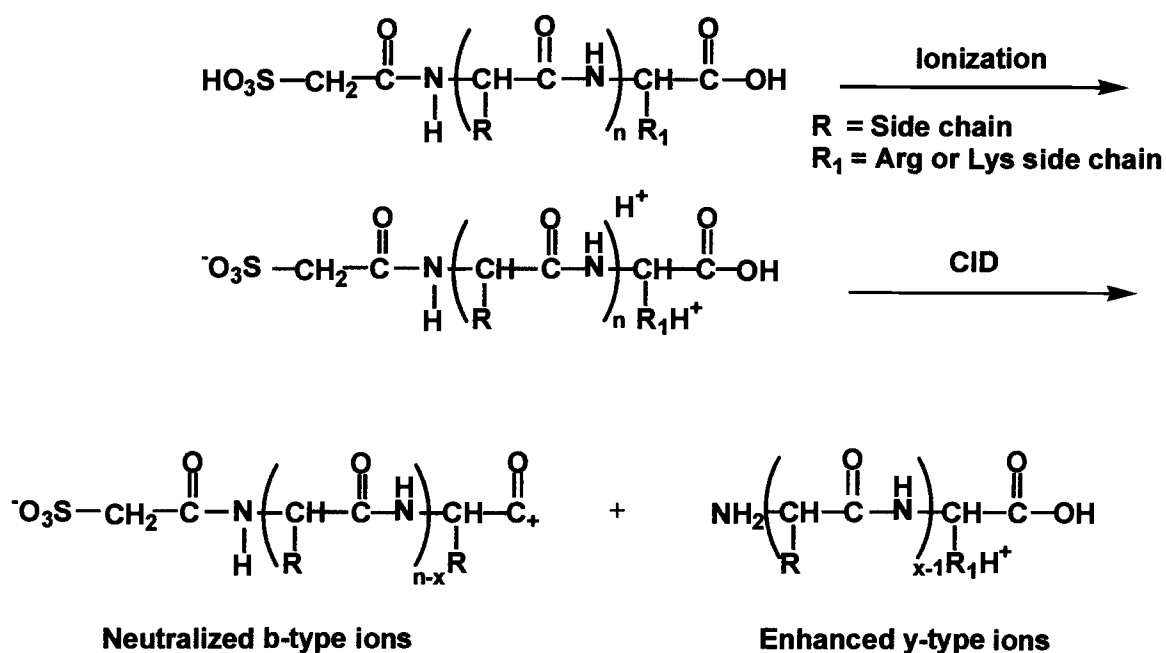


Figure 5.7. Proposed ESI CID fragmentation scheme for tryptic peptides N-terminally derivatized with sulfonic acid.

The simplification of tandem mass spectra is of great utility for the *de novo* sequencing of peptides. Clearly, N-terminal sulfonic acid derivatization could be used as an aid in identifying peptides produced from in-gel digestion of an unknown protein. Unambiguous sequencing of peptides or identifying unambiguous sequence tags within peptides would increase the confidence in determining the identity of an unknown protein. Also, derivatization of peptide mixtures produced from in-gel digestions of known proteins may lead to more peptides being sequenced, increasing the coverage of peptide maps (47, 137).

Along with playing a role in protein identification, the virtues of this derivatization procedure may also have utility in the mass spectrometric analysis of protein-nucleic acid complexes. The complexity of nucleopeptide fragmentation spectra produced from ESI tandem mass spectrometry has been clearly shown in Chapter 4. Any technique that could simplify these spectra would increase the likelihood of identifying sites of crosslinking within nucleoprotein complexes. It may be possible to use sulfonic acid derivatization of nucleopeptides produced from the tryptic digestion of nucleoprotein complexes to increase the fragmentation of the peptide portion of the nucleopeptide and minimize the fragmentation of the DNA portion of the nucleopeptide. The fragmentation spectra of sulfonic acid derivatized nucleopeptides may also show an enhancement in γ -ions and a suppression of b -ions, similar to that of derivatized tryptic peptides, to simplify the interpretation of these compounds. Other derivatizing reagents could be used, such as S-pentafluorophenyl-[tris(2,4,6-trimethoxyphenyl)phosphonium]acetate bromide, to enhance b -ions

and suppress y-ions, to provide fragmentation spectra that corroborate those produced from sulfonic acid derivatization (2, 53, 112). The use of derivatizing agents to enhance and simplify nucleopeptide fragmentation spectra is certainly a novel idea; however, the feasibility of this approach requires further studies.

6. UV-CATALYZED CROSSLINKING OF HUMAN REPLICATION PROTEIN A TO SINGLE-STRANDED DNA SUBSTRATES

There are three interesting questions about the structure/function relationship of hRPA that have yet to be resolved by other studies. Namely, what structural features of the protein modulate the increased association of the protein for increasing lengths of oligonucleotides? What is the importance of specific aromatic amino acids in binding single-stranded DNA? Do the hRPA14 and hRPA32 subunits both take part in binding to DNA? Based on the results obtained from crosslinking studies of *E. coli* Ung, similar studies applied to hRPA have the potential to answer these questions.

To test the feasibility of using UV crosslinking to study this protein, three experiments were identified that, if successful, would indicate the likelihood of success in pursuing future biological characterization. These experiments were to: 1) purify sufficient quantities of active hRPA, 2) determine optimum irradiation times for crosslinking hRPA to various oligonucleotides, and 3) produce a peptide map of the protein and a nucleopeptide map of an hRPA-oligonucleotide complex. This chapter discusses the results from these experiments and lays a foundation for future crosslinking studies of hRPA.

6.1 Results

6.1.1 Overproduction of Recombinant hRPA

The overproduction of recombinant hRPA in *E. coli* BLR21(DE3) cells was tested by transforming these cells with the p11d-tRPA plasmid (46). The plasmid contains a single T7 RNA polymerase promoter followed by the coding sequence of hRPA70, hRPA14, and hRPA32. Transformants were induced with IPTG, lysed, and examined for the expression of the recombinant subunits by denaturing polyacrylamide gel electrophoresis (Figure 6.1). It was found that upon induction, cells containing p11d-tRPA synthesized the 70-kDa and 32-kDa subunits of hRPA. The 14-kDa subunit could not be identified due to the high density of protein at the dye front of the gel.

6.1.2 Purification of hRPA

Subsequent to initial overproduction experiments, a survey of the coding sequences of hRPA70, hRPA32, and hRPA14 revealed the presence of 22 arginine codons (AGA and AGG). These codons are common in eukaryotic genes, but are relatively rare in *E. coli* (158). It was suspected that high-level expression of these genes containing codons rarely used by *E. coli* would result in depletion of the internal aminoacyl tRNA pools. This in turn could lead to the formation of aborted translation products that would be observed as proteolytic fragments or it could lead to misincorporation of lysine at arginine codons during overexpression in *E. coli* (23, 36). To avoid the formation of aborted translation

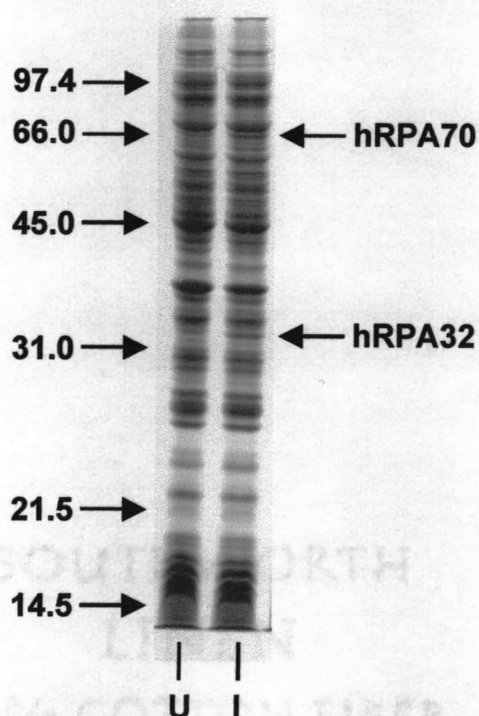


Figure 6.1. Overexpression of hRPA. A mini-culture of *E. coli* BLR21(DE3) cells that had been transformed with p11d-tRPA was split into two fractions which were either induced with IPTG or left uninduced. The cells from both fractions were lysed, the soluble protein was separated on a 12.5% denaturing polyacrylamide gel, and protein bands were visualized by Coomassie staining, as described under "Experimental Procedures." The lanes of uninduced (*U*) and induced (*I*) cultures were compared to identify unique bands. *Arrows* to the left of the gel indicate the positions of molecular weight standards and *arrows* to the right of the gel indicate the overexpressed 70 kDa and 32 kDa subunits of hRPA.

products and the misincorporation of lysine for arginine in hRPA, the plasmid p-RIL was co-transformed with p11d-tRPA into the *E. coli* BLR21(DE3) cell line. The plasmid pRIL (Stratagene) contains the tRNA gene (*argU*), whose product recognizes the two rare arginine codons AGA and AGG, and it was also coded for tRNA genes of isoleucine and leucine.

Initially, the hRPA purification procedure of Wold and colleagues (46) was attempted which called for the use of Affi-gel Blue, hydroxylapatite, and Mono Q columns. Using this scheme, it was difficult to recover the protein from the hydroxylapatite column. This result led to the development of a new purification scheme (Figure 6.2). Using the newly developed scheme, the cell lysate from induced BLR21(DE3) cells transformed with the p11d-tRPA overexpression plasmid and p-RIL was applied to an Affi-Gel Blue column and washed with buffers of increasing chaotropic strength. Individual proteins of approximate molecular weights 70 kDa, 32 kDa, and 14 kDa, as observed by denaturing polyacrylamide gel electrophoresis, were abundant in column washes with 1.5 M sodium thiocyanate (Fig 6.3). The chaotropic strength required to elute proteins of these molecular weights strongly indicated that these fractions contained enriched recombinant heterotrimeric hRPA. Elution fractions containing the proposed subunits of hRPA were pooled and dialyzed against HI-500 mM NaCl buffer, which took the place of the hydroxylapatite column used in the purification procedure by Wold and colleagues (46). It was reasoned that hRPA would have a natural affinity for single-stranded DNA agarose, making it a logical fractionation step to obtain active protein. The dialyzed fraction was further

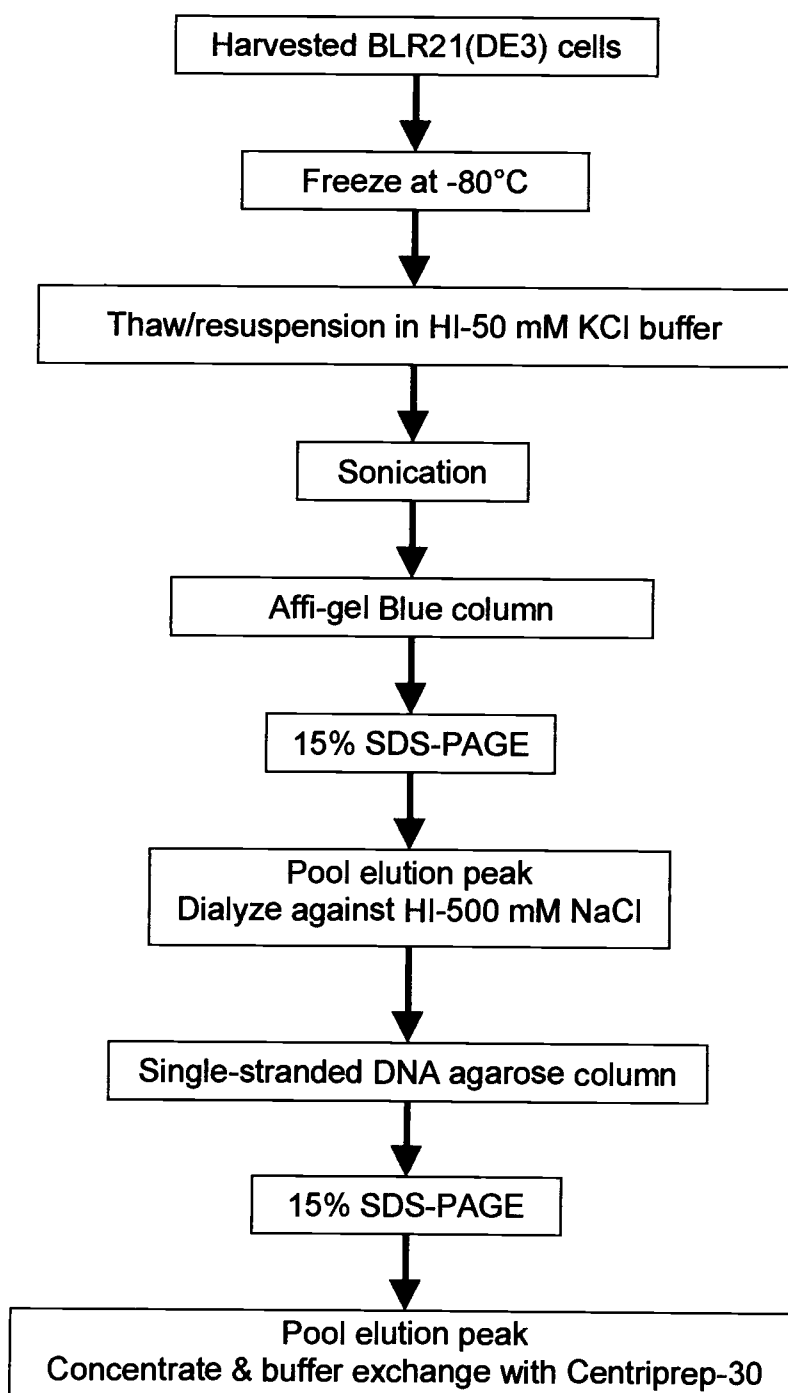


Figure 6.2. Purification scheme for hRPA.

Figure 6.3. Affigel Blue purification of hRPA. A soluble protein lysate of *E. coli* BLR21(DE3) cells transformed with p11d-tRPA and pRIL was loaded onto an Affigel Blue column and eluted fractions were visualized on 15% SDS-polyacrylamide gels by Coomassie stain, as described under "Experimental Procedures." The lysate fraction was visualized on the gel and is designated *L* and fractions 4 and 7 were collected during loading of the column. Fractions 12-17 were collected during washing with HI-50 mM KCl, fractions 20-27 during washing with HI-800 mM KCl, fractions 30-36 during washing with HI-0.5 M NaSCN, fractions 39-49 during washing with HI-1.0 M NaSCN, and fractions 52-91 during elution with HI-1.5 M NaSCN. The positions of molecular weight markers are indicated by *arrows*. From analysis of these gels, fractions 61-81 were pooled for further purification.

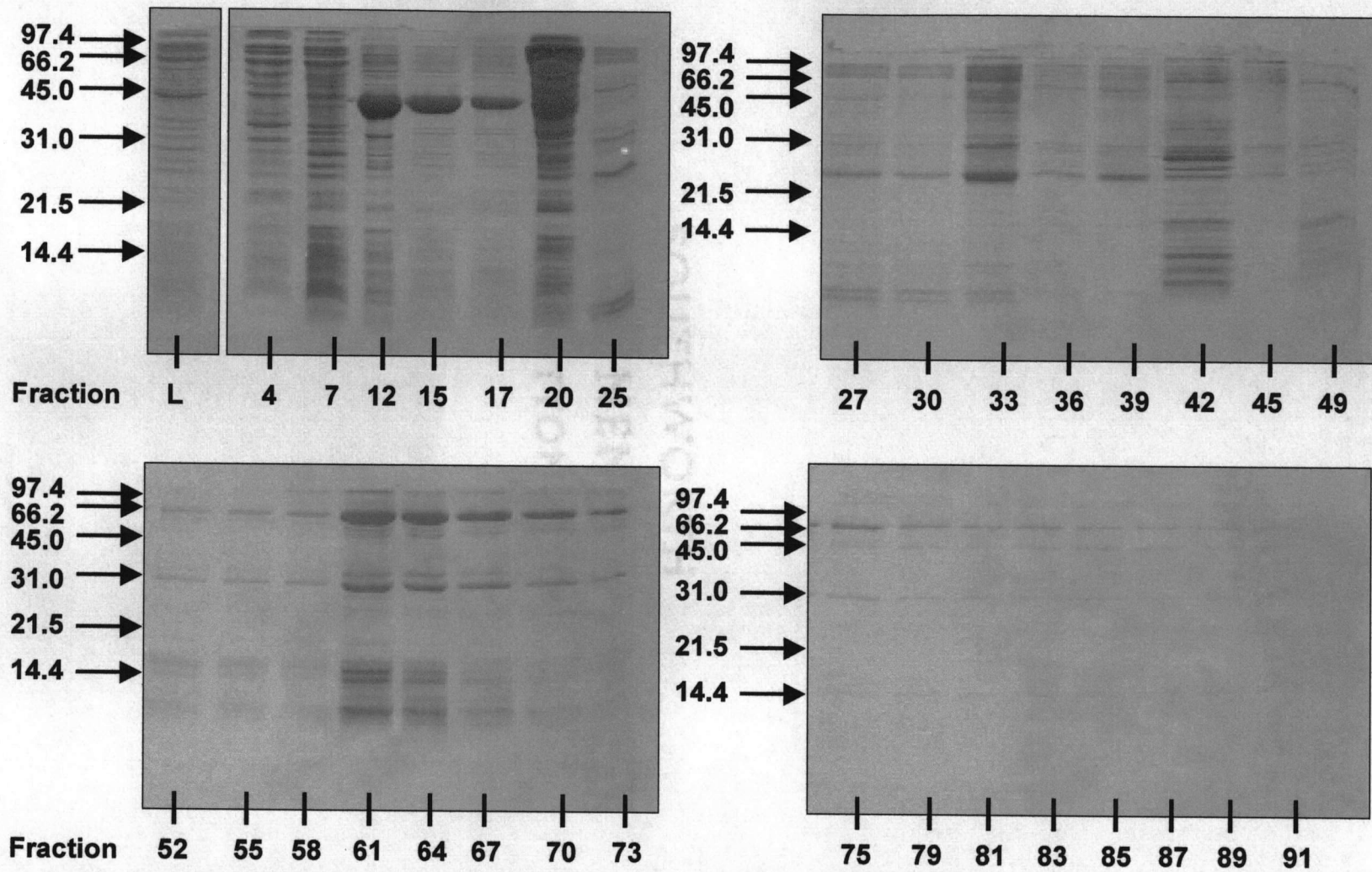


Figure 6.3. Affigel Blue purification of hRPA.

purified by using a single-stranded DNA agarose column. An elution peak containing proteins of approximately 70 kDa, 32 kDa, and 14 kDa began at 700 mM NaCl (Fig 6.4), again indicating that these fractions contained recombinant heterotrimeric hRPA. After pooling the peak of elution and concentration, the purified sample contained the three proposed subunits of hRPA as well as a minor bands at approximately 23 kDa and two bands around 66 kDa (Fig 6.5A). Typically, about 1.0 mg of protein was purified per liter of cultured cells.

Performing proteolytic digestions in-gel and sequencing the resulting peptides identified the purified proteins by mass spectrometry (125). Initially, the 14-kDa band was cut from a SDS-polyacrylamide gel containing the purified proteins (Figure 6.5A). The protein band was digested in-gel with trypsin and then surveyed by MALDI MS to identify tryptic fragments (Figure 6.5B). An ion species of m/z 2039.7 was observed and tentatively identified as the C-terminal tryptic fragment of hRPA14. The in-gel digestion mixture was analyzed by LC/ESI MS and the $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ion species of 2039.7 were present (Figure 6.5C). Tandem mass analysis was conducted on the $[M+2H]^{2+}$ ion species (Figure 6.5D) and the resulting data were used in a peptide sequence database search algorithm. The results of the database search (Figure 6.6) revealed that this peptide was in fact the C-terminal tryptic peptide of hRPA14, thus identifying this protein. A similar protein identification scheme was used to identify the five other protein bands (Figure 6.5A). This analysis revealed that the 70-kDa and the 32-kDa protein bands corresponded to hRPA70 and hRPA32. The two protein bands of approximately 66 kDa and 57 kDa were both

Figure 6.4. Single-stranded DNA agarose column purification of hRPA.

Elution fractions 61-81 from an Affigel Blue column (Figure 6.3) containing hRPA were pooled, dialyzed against 500 mM NaCl, loaded onto a single-stranded DNA agarose column, and collected fractions were visualized on 15% SDS-polyacrylamide gels by Coomassie stain, as described under "Experimental Procedures." A fraction of the sample loaded onto the column was visualized on the gel and is designated *L*. Fractions 5-11 were collected during loading of the column, fractions 13-22 during washing with HI-500 mM NaCl, and fractions 25-69 during elution with a linear gradient of 500 mM NaCl to 3 M NaCl in HI buffer. The concentration of NaCl in collected fractions was determined from conductivity measurements and indicated by *M NaCl*. The positions of molecular weight markers are indicated by *arrows*. Fractions 31-60 were pooled, buffer exchanged to HI-200 mM KCl, and concentrated.

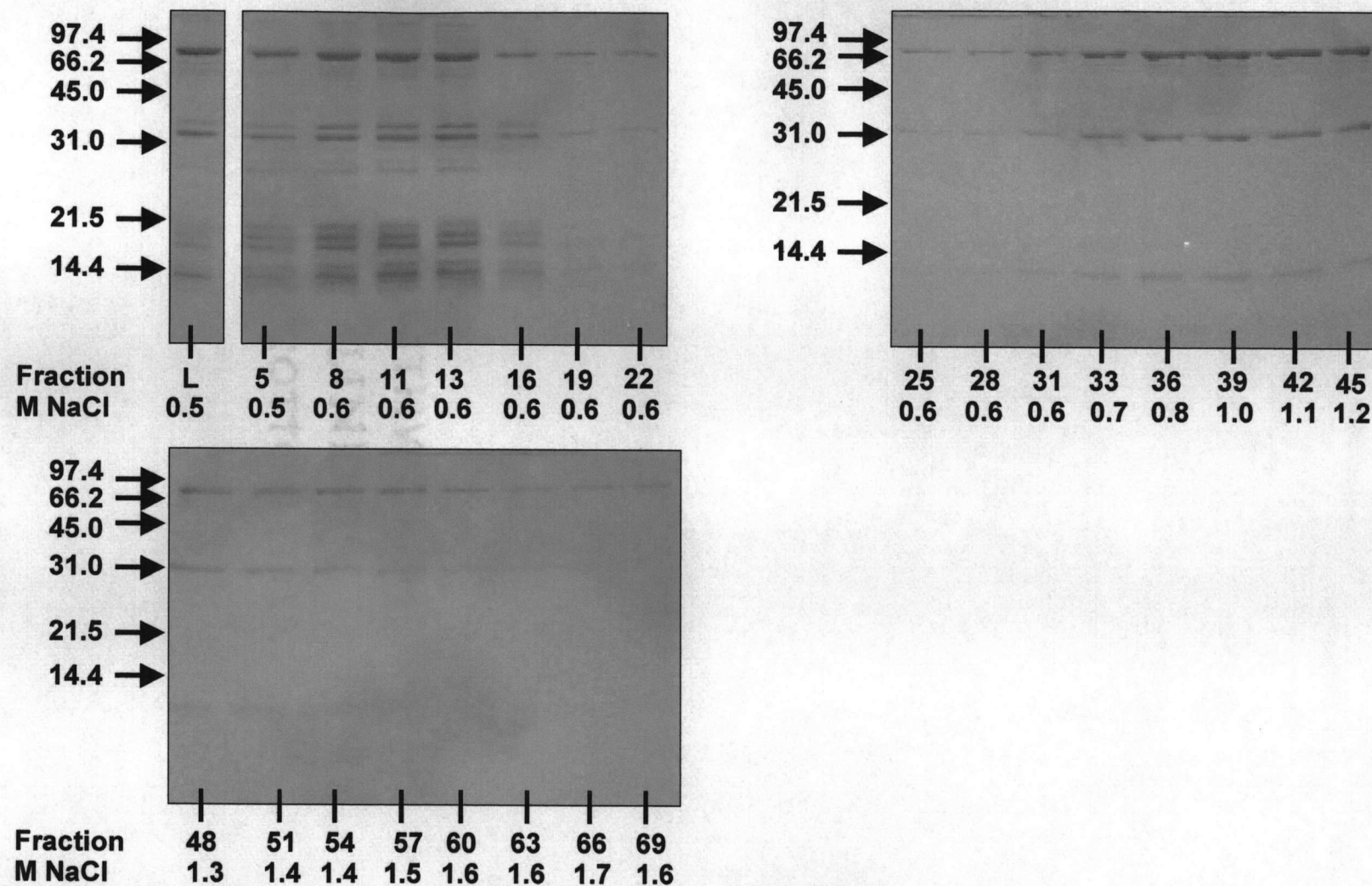


Figure 6.4. Single-stranded DNA agarose column purification of hRPA.

Figure 6.5. Mass spectrometric analysis of purified hRPA. *A*, fractions 31-60 from a single-stranded DNA agarose column (Figure 6.4) containing highly purified proteins of ~ 70 kDa, ~ 32 kDa, and ~ 14 kDa (thought to be subunits of hRPA) were pooled, exchanged with HI-200 mM KCl buffer, and concentrated to 300 µg/ml. The concentrated sample was analyzed by loading 2.5 µg (20 pmol) of protein onto a 15% SDS-polyacrylamide gel and visualizing with silver stain. *Arrows* indicate the positions of molecular weight markers. *B*, the visualized protein bands were excised from the gel, digested with trypsin, and the resulting tryptic peptides were analyzed by MALDI-MS. The tryptic in-gel digestion of the ~ 14 kDa protein is shown. *C*, electrospray ionization LC-MS¹ (ion-trap) of the ~ 14 kDa tryptic digestion was conducted and peaks corresponding to the $[M+2H]^{2+}$ (m/z 1020.3) and $[M+3H]^{3+}$ (m/z 680.6) ion species of mass 2038.6 were identified. *D*, LC-MS² analysis was then conducted on the m/z 1020.3 ion species, as observed in the MS¹ analysis (*panel C*), in order to produce peptide sequence information for subsequent use in the peptide database search algorithm Protein Prospector.

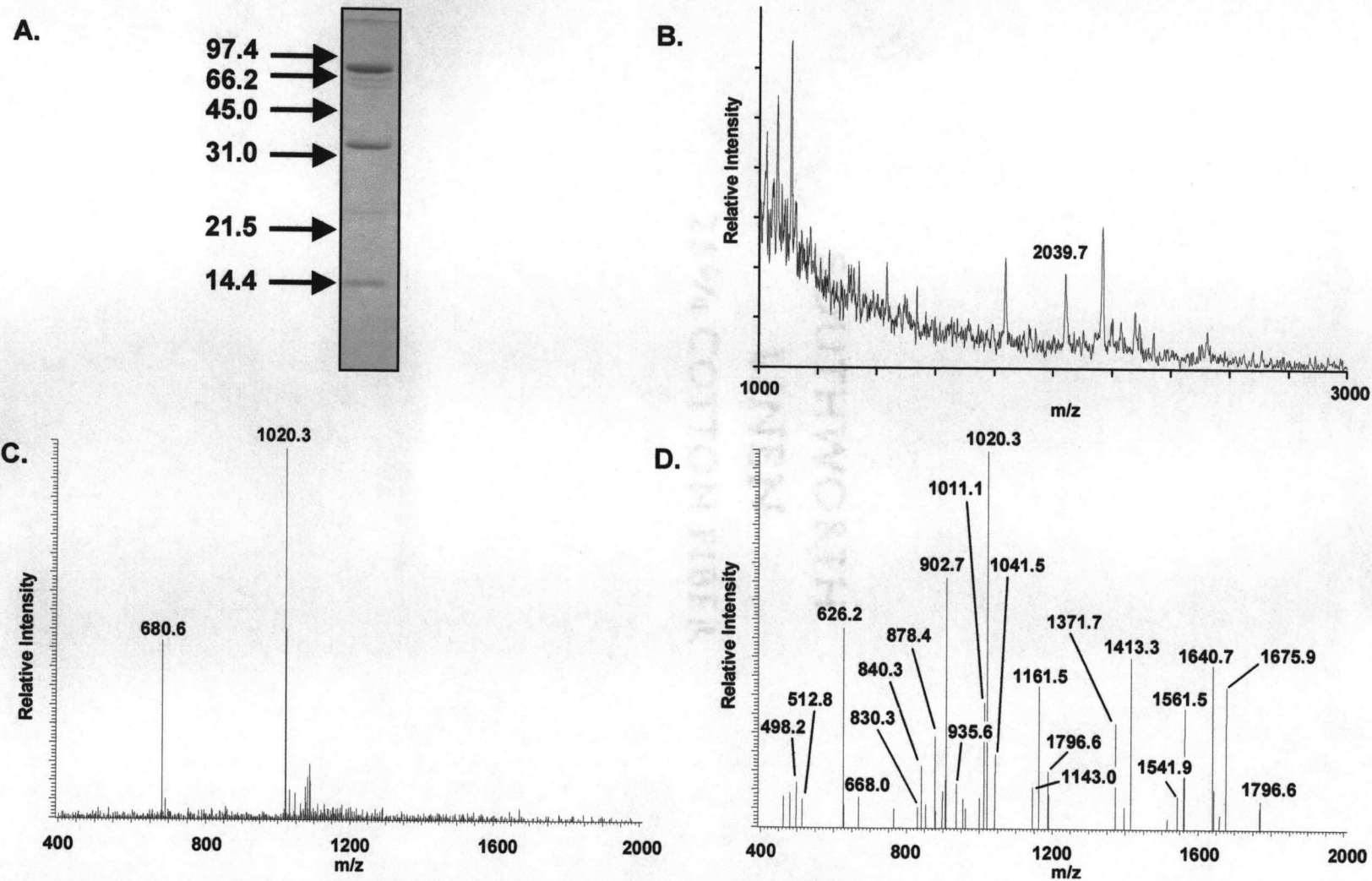


Figure 6.5. Mass spectrometric analysis of purified hRPA.

MS-Tag Search Results

Sample ID (comment): **14 kDa gel band**

Database searched: **SwissProt.10.4.99**

Full Molecular Weight range: **77996** entries.

Full pI range: **77996** entries.

Species search (**HOMO SAPIENS**) selects **5128** entries.

Number of sequences passing through parent mass filter: **175**

MS-Tag search selects **1** entry.

Parent mass: **1020.2900⁺²** (+/- 1.0000 Da)

M+H equivalent: **2039.5720** (+/- 1.0000 Da)

Fragment Ions used in search: **164.70, 498.20, 512.80, 626.20, 668.00, 764.50, 830.30, 840.30, 878.40, 907.20, 935.00, 1011.10, 1020.30, 1041.50, 1143.00, 1161.50, 1188.30, 1371.70, 1413.30, 1514.70, 1541.90, 1561.50, 1675.90, 1796.60** (+/- 1.00 Da)

Ion Types Considered: **a b b-NH3 b-H2O b+H2O y y-NH3 y-H2O**

Search Mode identity	Max. # Unmatched Ions 10	Peptide Masses are average	Digest Used Trypsin	Max. # Missed Cleavages 1	Cysteines Modified by acrylamide	Peptide N terminus Hydrogen (H)	Peptide C terminus Free Acid (O H)
----------------------------	-----------------------------------	----------------------------------	---------------------------	---------------------------------	--	---------------------------------------	--

Result Summary

Rank	# Unmatched Ions	Sequence	MH ⁺ Calculated (Da)	MH ⁺ Error (Da)	Protein MW (Da)/pI	Species	SwissProt.10.4.99 Accession #	Protein Name
1	10/24	(K) IIHDFPQFYPLGIVQHD	2040.3369	-0.7649	13568.8 / 4.97	HUMAN	P35244	REPLICATION PROTEIN A 14 KD SUBUNIT (RP-A) (RF-A) (REPLICATION FACTOR-A PROTEIN 3)

MS-Tag 4.2.1, ProteinProspector 3.2.1

© Copyright (1995-1999) The Regents of the University of California.

Figure 6.6. Database search results for identification of the ~14 kDa purified protein. A list of the electrospray ionization LC-MS/MS m/z data from Figure 6.5, panel D was entered into the Protein Prospector protein database search engine and the results of that search against the SwissProt protein data bank are shown.

identified as being related to hRPA70, most likely proteolytic fragments of hRPA70. The 23-kDa protein band was identified as the *E. coli* single-stranded DNA-binding protein (Ecssb). Densitometry experiments (data not shown) estimated that the hRPA70, hRPA32, and hRPA14 made up greater than 90% of the purified protein.

6.1.3 Single-Stranded DNA Binding Capability of hRPA

An electrophoretic mobility shift analysis was used to characterize the DNA binding capability of purified recombinant hRPA. A constant quantity of oligonucleotide [^{32}P]dT₃₀ was incubated with increasing amounts of hRPA, and mobility-shifted bands on a polyacrylamide gel corresponding to complexes between hRPA and [^{32}P]dT₃₀ were observed (Figure 6.7). At hRPA:dT₃₀ mixture ratios less than one, a distinct mobility-shifted band was observed. Starting at protein to oligonucleotide mixture ratios of one, additional shifted bands of altered mobility were observed and, close to 100% of the oligonucleotide had been bound by the protein. At a hRPA to dT₃₀ mixture ratio of 5, 100% of the mobility-shifted complex was of altered mobility. This result shows that essentially the entire production of the purified recombinant protein was capable of binding DNA.

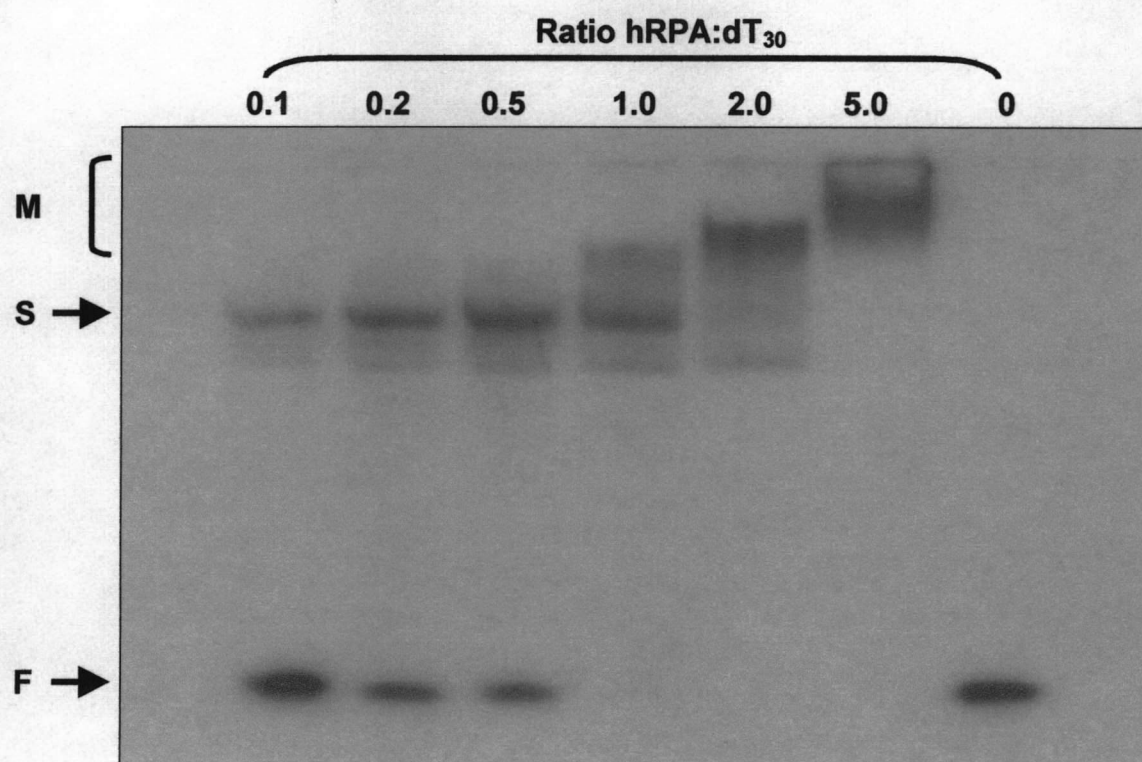


Figure 6.7. Electrophoretic mobility shift analysis to measure the capability of hRPA to bind DNA. Varying amounts (0 to 50 pmol) of purified hRPA were incubated with 10 pmol [³²P]dT₃₀ for 30 min at room temperature. The reactions were separated on a 5% polyacrylamide gel and the gel was dried and subjected to autoradiography, as described under "Experimental Procedures." Location of free [³²P]dT₃₀, [³²P]dT₃₀ bound to a single hRPA molecule, and [³²P]dT₃₀ bound by multiple hRPA molecules are indicated by (F), (S), and (M), respectively.

6.1.4 UV-Catalyzed Crosslinking of hRPA to dT₃₀

The ability to form UV-catalyzed crosslinks between hRPA and dT₃₀ was tested by irradiating a mixture of hRPA and [³²P]dT₃₀ and analyzing the products by SDS polyacrylamide gel electrophoresis (Figure 6.8). Silver staining of the gel showed the presence of a new mobility shifted protein band that was dependent on both DNA and UV irradiation (Figure 6.8A, lane 6). The position of the mobility shifted band suggested that it corresponded to the 70-kDa subunit of hRPA crosslinked to [³²P]dT₃₀. The gel was also analyzed by autoradiography, which showed the presence of four components in the gel that contained [³²P]dT₃₀. One of the components was free [³²P]dT₃₀ (Figure 6.8B, lane 4) while the other three components were formed only after UV-irradiation (Figure 6.8, lane 6), suggesting these components were crosslinked nucleoprotein complexes. The first nucleoprotein complex, labeled XL-1, was shifted in mobility to the same extent as the crosslinking product observed after silver staining (compare Figure 6.8A, lane 6 and Figure 6.8B, lane 6) also suggesting that this crosslinked species was hRPA70 covalently attached to [³²P]dT₃₀. The second crosslinking product, XL-2, was visible only by autoradiography and was shifted in mobility to a lesser extent than XL-1. It was possible that this second crosslinking product resulted from crosslinking of the proteolytic fragments of hRPA70 (Figure 6.8a, lanes 1, 2, and 3; apparent molecular weights less than 66.2 kDa) to [³²P]dT₃₀. The third crosslinking product, XL-3, migrated in the gel with an apparent molecular weight of 31 kDa. Since hRPA32 was retarded in the gel slightly more than XL-3, it was unlikely that this crosslinking product is

Figure 6.8. Dependence of UV light for the formation of UV-catalyzed crosslinks between hRPA and dT₃₀. Seven samples (20 μ L) were prepared containing either 20 pmol hRPA, 20 pmol [³²P]dT₃₀, 20 pmol [³²P]dT₃₀ and 20 pmol hRPA, or 20 pmol [³²P]dT₃₀ and 20 pmol heat-denatured hRPA (denoted by *). The samples were either unirradiated or irradiated for 1 min with 254-nm light as indicated. Each sample was loaded onto a 12.5% SDS polyacrylamide mini-gel and electrophoresis was carried out. The gel was stained with silver (*panel A*), then dried, and subjected to autoradiography (*panel B*). The positions of hRPA70, hRPA32, hRPA14, [³²P]dT₃₀, and hRPA $\times$ [³²P]dT₃₀ crosslinking products (XL) are indicated by *arrows* to the right of the panels. The position of molecular weight markers are indicated by *arrows* to the left of the panels.

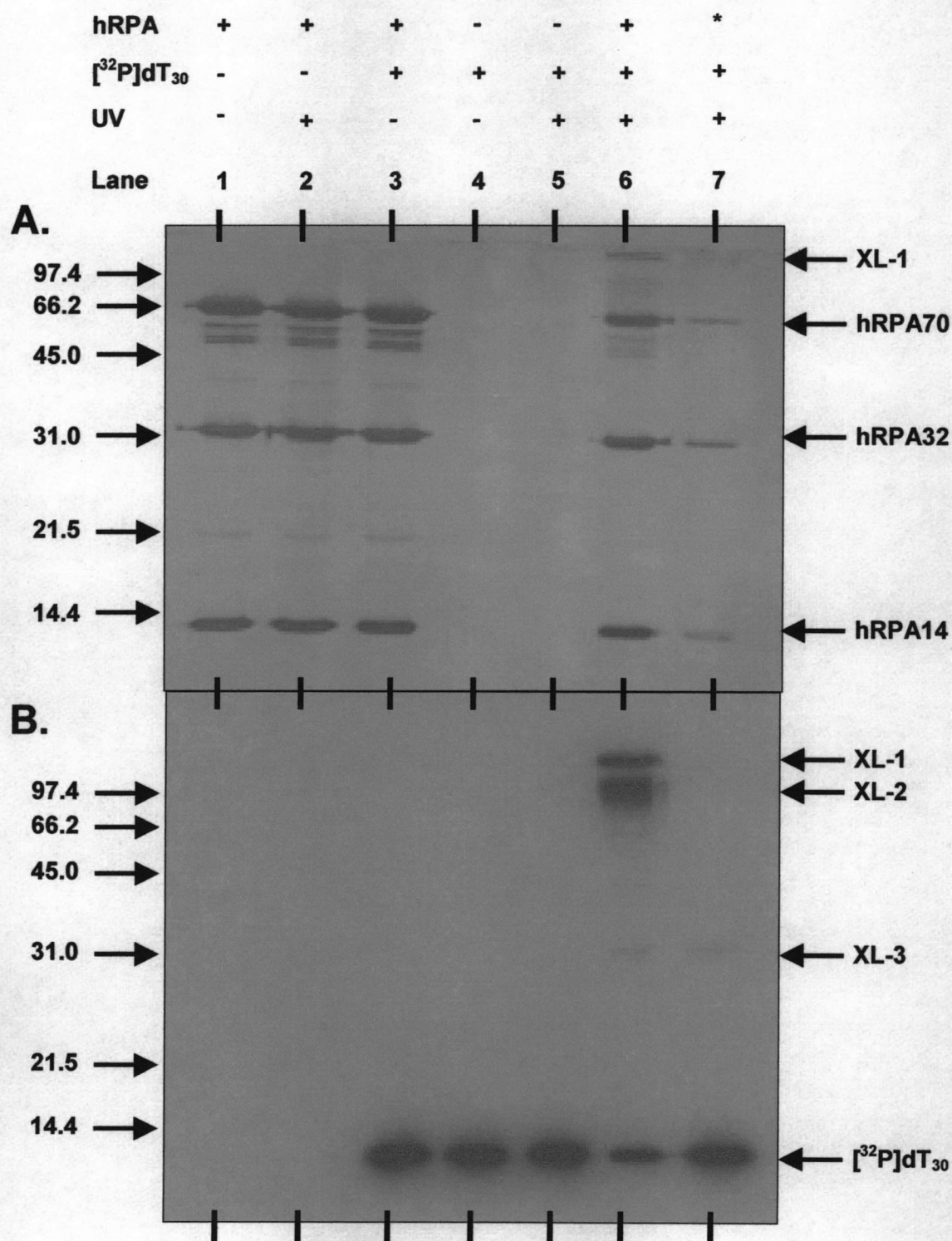


Figure 6.8. Dependence of UV light for the formation of UV-catalyzed crosslinks between hRPA and dT₃₀.

associated with hRPA32. A faint protein band was observed in the silver stained gel with an apparent molecular weight of 23 kDa (Figure 6.8A, lanes 1, 2, and 3) and was identified as Ecssb (Section 6.1.2). The covalent attachment of [^{32}P]dT₃₀ to Ecssb as a result of crosslinking could give a predicted molecular weight close to 31 kDa; thus, the faint crosslinking product XL-3 was possibly Ecssb crosslinked to [^{32}P]dT₃₀. The specificity of crosslinking was also investigated by heat denaturing purified hRPA, incubating the denatured protein with [^{32}P]dT₃₀, irradiating with UV-light, and then analyzing the photoproducts by SDS-polyacrylamide gel electrophoresis. Silver staining and autoradiography of the gel did not indicate the formation of a crosslinking product associated with hRPA70 (Figure 6.8A, lane 7 and Figure 6.8B, lane 7). A crosslinking product, however, was observed in a region of the gel suggesting Ecssb crosslinked to dT₃₀ (Figure 6.8B, lane 7). When taken together, the results of the silver stained gel and the autoradiogram establish that hRPA70, and the proteolytic fragments of hRPA70, is the only subunit of the hRPA heterotrimer that takes part in crosslinking to dT₃₀ under the conditions tested.

6.1.5 UV-Dose Response of Crosslinking hRPA to dT₁₀, dT₃₀, and dT₇₀

In order to determine the optimum crosslinking time for hRPA to oligonucleotides dT₁₀, dT₃₀, and dT₇₀, these oligonucleotides were mixed individually with hRPA and crosslinked for increasing doses of UV light. The photoproducts were analyzed on dried SDS polyacrylamide gels. For oligonucleotide dT₁₀ (Figure 6.9A), only a single crosslinking product was present

Figure 6.9. Dose response of Crosslinking hRPA to dT₁₀, dT₃₀, dT₇₀.

Crosslinking reaction mixtures (200 μ L) containing 50 pmol of hRPA, 50 pmol of a [³²P] oligonucleotide, and buffer HI-30 mM KCl were irradiated with UV light for 0, 1, 2, 5, 10, and 15 min (lanes 2 through 7, respectively), as described under "Experimental Procedures." At each time point, 5- μ L samples were removed and subsequently analyzed by SDS-polyacrylamide gel electrophoresis. Gels were dried and crosslinking products were visualized by autoradiography. **A**, the oligonucleotide [³²P]dT₁₀ was used in the crosslinking reaction and the photoproducts were analyzed by 15% SDS polyacrylamide gel electrophoresis. Lane 1 contains 1.4 pmol of unirradiated [³²P]dT₁₀. **B**, the oligonucleotide [³²P]dT₃₀ was used in the crosslinking reaction and the photoproducts were analyzed by 12% SDS polyacrylamide gel electrophoresis. Lane 1 contains 1.4 pmol of unirradiated [³²P]dT₃₀. **C**, the oligonucleotide [³²P]dT₇₀ was used in a crosslinking reaction and the photoproducts were analyzed by 8% SDS polyacrylamide gel electrophoresis. Lane 1 contains 1.4 pmol of unirradiated [³²P]dT₇₀. Arrows in each figure panel indicate the photoproducts used for subsequent quantitation (Figure 6.10). Asterisks indicate the position of secondary crosslinking products and *P* indicates the positions of oligonucleotide probes.

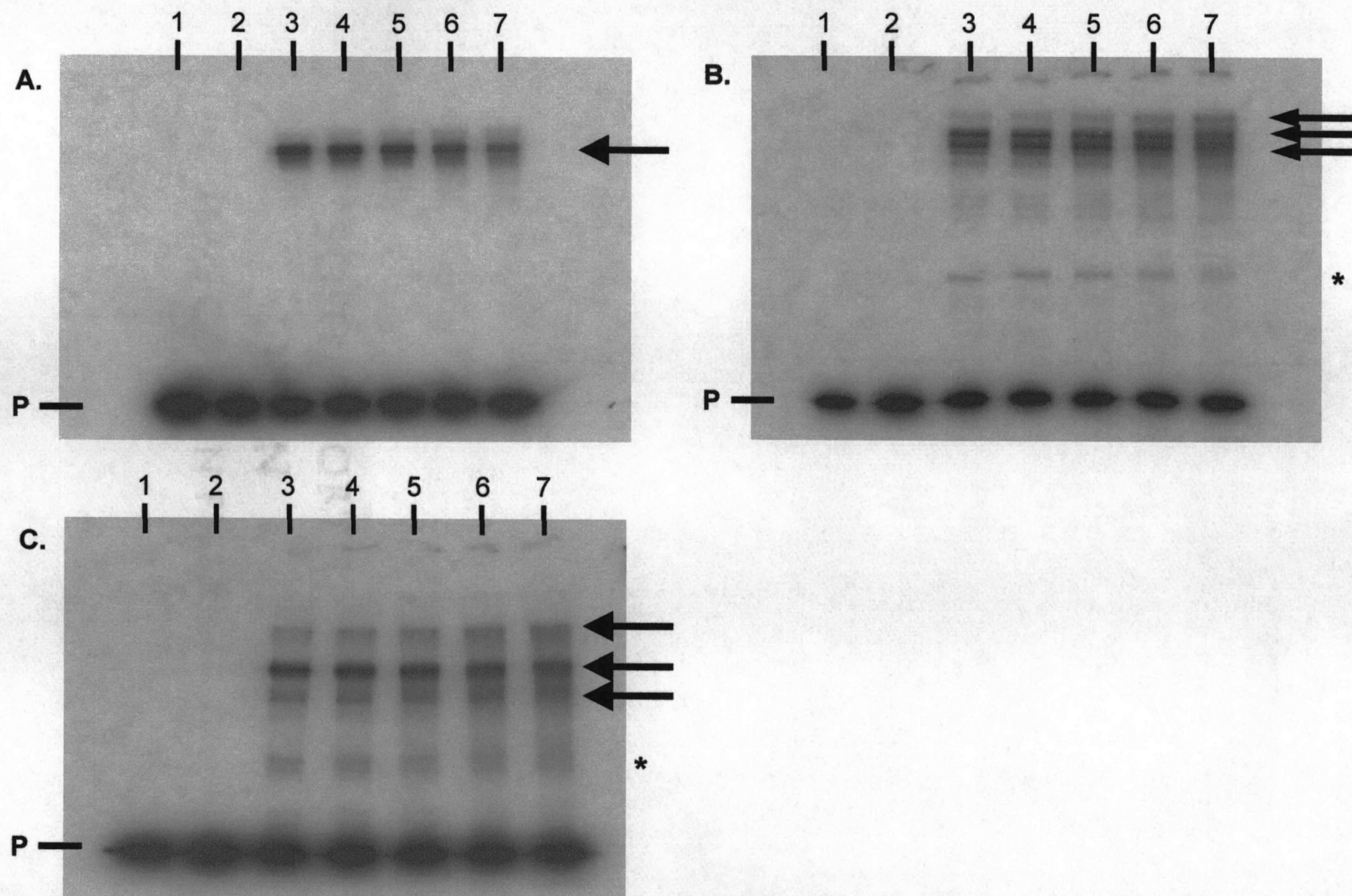


Figure 6.9. Dose response of crosslinking hRPA to dT₁₀, dT₃₀, and dT₇₀.

whereas for oligonucleotides dT₃₀ and dT₇₀ multiple crosslinking products were present (Figure 6.9B and C). The intensity of the mobility-shifted bands (as indicated by arrows in Figure 6.9) were quantified from PhosphorImages and plotted versus the irradiation time (Figure 6.10). In the case of all three oligonucleotides, the product yield was at a maximum after 1 minute of UV exposure. From this result, an optimum crosslinking time of 1 min for preparative reactions between hRPA and dT₁₀, dT₃₀, and dT₇₀ was established.

6.1.6 Peptide Mapping of hRPA70

Crosslinking analysis identified hRPA70 as the major subunit involved in crosslinking to dT₃₀ (Figure 6.8). Therefore, the 70-kDa subunit of hRPA was digested in-gel with trypsin and the products were analyzed by LC/ESI ion trap mass spectrometry to produce a map of the subunit. The masses of the tryptic fragments were compared to a predicted tryptic digestion of the 70-kDa subunit to identify regions of the subunit covered in the map. Mass spectrometry detected 31 tryptic peptides of which 26 were sequenced by tandem mass analysis (Table 6.1). The detected peptides covered 63% of the amino acid sequence of hRPA70.

6.1.7 MALDI MS Analysis of hRPA70 × dT₃₀ Nucleopeptides

A preparative crosslinking reaction of hRPA with dT₃₀ was conducted and the products of the photoreaction were analyzed by a 10% SDS polyacrylamide

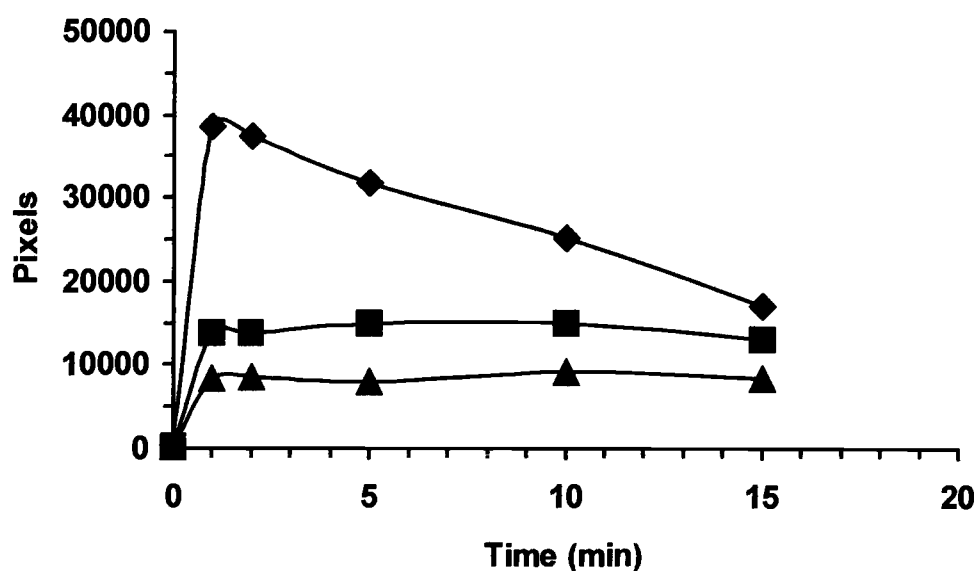


Figure 6.10. Quantitation of UV dose response during crosslinking of hRPA to dT₁₀, dT₃₀, and dT₇₀. Individual crosslinking reaction mixtures between hRPA and ³²P labeled dT₁₀, dT₃₀, and dT₇₀ were analyzed on dried SDS polyacrylamide gels (Figure 6.9) by using PhosphorImager analysis. The photoproducts (as indicated by *arrows* in Figure 6.9) were quantified and plotted. (◆), (■), and (▲) indicate quantified photoproducts at various time points of dT₁₀, dT₃₀, and dT₇₀, respectively.

Table 6.1. Tryptic peptide mapping of carboxyamidomethylated human RPA70 by LC-ESI/IT-MS.

No.	Peptide Position	Predicted Mass (M+H ⁺) ^a	ESI/IT-MS (M+H ⁺)	MS/MS
1	1 - 16	1646.9	1646.6	
2,3,4	17 - 41	1358.8 ^b	1358.1 ^b	✓
5	42 - 43	338.2	N.D. ^c	
6	44 - 81	4348.1 (2) ^d	N.D.	
7	82 - 88	834.5	834.4	✓
8	89 - 91	347.2	N.D.	
9	92 - 92	175.1	N.D.	
10	93 - 103	1285.8	N.D.	
11	104 - 111	760.4	760.4	✓
12,13	112 - 157	1486.4 ^e	1486.8 ^e	✓
14	158 - 163	596.3	596.3	✓
15	167 - 167	452.3	N.D.	
16	168 - 183	1485.7	1485.5	✓
17	184 - 196	1402.8	1402.5	
18	197 - 202	806.4 (1)	806.4 (1)	
19	203 - 206	461.3	N.D.	
20	207 - 210	503.3	503.3	✓
21,22	211 - 216	1121.6	1121.4	✓
23	221 - 234	1606.8	1606.7	
24,25	235 - 253	1105.6	1106.6	✓
26	254 - 259	806.4	806.3	✓
27	260 - 263	418.3	N.D.	
28	264 - 267	445.3	N.D.	
29	268 - 273	693.3	693.4	
30	274- 311	4517.9 (1)	N.D.	
31,32	312 - 324	1448.8 (1)	1448.6 (1)	✓
33	325 - 331	813.4	813.3	✓
34	332 - 335	488.3	488.3	✓
35	336 - 339	490.2	N.D.	
36	340 - 343	446.3	N.D.	
37	344 - 344	175.1	N.D.	
38	345 - 354	1141.6	1141.5	✓
39,40	355 - 372	1968.0	1967.8	✓
39,40,41	355 - 379	1359.8 ^b	1359.8 ^b	✓
42	380- 382	303.2	N.D.	
43	383 - 389	737.4	737.4	✓
44	390 - 410	2205.2	N.D.	
45	411 - 412	288.2	N.D.	
46	413 - 431	2008.0	N.D.	
47	432 - 443	1163.5	1163.4	✓

Table 6.1 (continued)

No.	Peptide Position	Predicted Mass (M+H ⁺) ^a	ESI/IT-MS (M+H ⁺)	MS/MS
48	444 - 449	752.4	752.4	✓
49,50	450 - 458	1273.6 ^b	1273.4 ^b	✓
51	473 - 473	147.1	N.D.	
52	474 - 488	1918.7 (3)	1918.6 (3)	
53	489 - 489	147.1	N.D.	
54	490 - 499	1205.6	1205.5	✓
55	500 - 502	436.2 (1)	N.D.	
56	503 - 511	1157.5 (1)	1157.4	✓
57	512 - 513	338.2	N.D.	✓
58	514 - 551	4301.1 (1)	N.D.	
59,60	552 - 568	1044.0 ^b	1043.9 ^b	✓
61	569 - 573	669.4	669.4	✓
62	574 - 575	274.2	N.D.	
63,64	576 - 586	1340.7	1339.6	✓
65	587 - 588	260.2	N.D.	
66,67	589 - 600	1393.6	1393.6	✓
68	601 - 604	524.3	N.D.	
69	605 - 605	175.1	N.D.	
70	606 - 611	718.4	N.D.	
71	612 - 612	175.1	N.D.	
72	613 - 616	421.2	N.D.	

^a Predicted peptide masses (monoisotopic) calculated using the computer program GPMW, version 3.1, Lighthouse Data.

^b Predicted Mass for [M+2H]²⁺ ion species.

^c N.D. refers to not detected.

^d Values in parentheses indicate the number of cysteine residues within the peptide modified by carboxyamidomethylation.

^e Predicted Mass for [M+3H]³⁺ ion species.

gel. Two mobility-shifted bands retarded in the gel above hRAP70 were observed after silver staining. The reaction products were also analyzed on a 7.5% SDS polyacrylamide gel (Figure 6.11). Correlating the mobility of these bands to the position of molecular weight markers suggested that these bands were not multiple hRAP70 molecules bound to a single oligonucleotide. If this were the case, a mobility-shifted band would be expected around 140 kDa (missing in Figure 6.11, lane 2). Instead, these bands might be either 1) a single hRAP70 molecule crosslinking to multiple oligonucleotides or 2) crosslinking products of hRAP70 and a proteolytic fragment of hRAP70 (M.W. ~66 kDa) containing a single oligonucleotide.

The two bands from the 10% SDS polyacrylamide gel were separately excised and proteolytically digested in gel with trypsin. Using a nucleopeptide isolation and purification procedure identical to that of the Ung nucleoprotein complex (Figure 3.9), resulting hRAP70 nucleopeptide species from each mobility-shifted band were analyzed by MALDI MS (Figure 6.12). The molecular weight (average) of dT₃₀ is 9064.0; therefore, the ion species observed above m/z 9065.0 were presumed to be nucleopeptides of hRAP70. Comparing the nucleopeptide mass spectra produced from the two mobility-shifted bands showed ion species common to both spectra (compare ion species 9107.9, 9820.2, and 10215 and 11329 in Figure 6.12A with ion species 9113.9, 9816.9, 10200, and 11342, respectively, in Figure 6.12B). An ion signal corresponding to chemically adducted, free dT₃₀ is present in both spectra (m/z 9107.9 for Figure 6.12A and m/z 9113.9 for Figure 6.12B). There are ion species that are unique

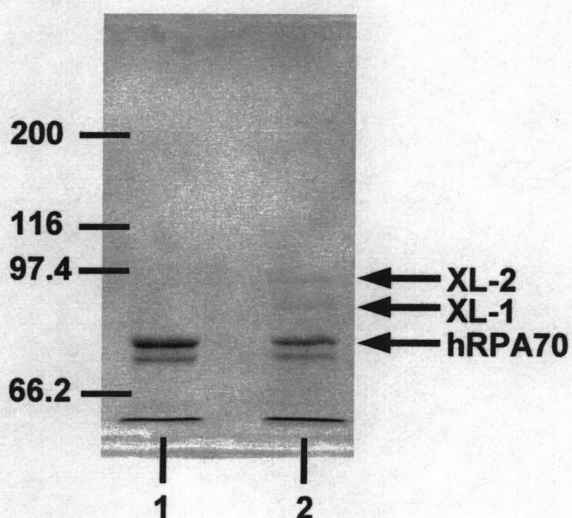


Figure 6.11. Analysis of a preparative scale crosslinking reaction of hRPA to dT₃₀. Ten mixtures (200 μ L each) containing 800 pmol of hRPA and 800 pmol of dT₃₀ were irradiated at 254 nm for 1 min and separated by denaturing polyacrylamide gel electrophoresis for subsequent nucleopeptide isolation, as described under "Methods and Materials." Aliquots (10 μ L) were removed from a crosslinking mixture both before (*Lane 1*) and after irradiation (*Lane 2*), analyzed on a 8% denaturing polyacrylamide gel, and protein bands were visualized by Coomassie stain. The positions of molecular weight markers indicated by arrows to the left of the gel. The position of hRPA70, a lower mobility shifted band (XL-1), and an upper mobility shifted band (XL-2) are indicated by arrows to the right of the gel.

Figure 6.12. MALDI mass spectrum of purified hRPA70 × dT₃₀ tryptic nucleopeptides. Crosslinking reaction mixtures of hRPA (8nmol total) and dT₃₀ (8nmol total) were separated on a 10% SDS polyacrylamide gel and protein bands were visualized by silver staining. Two mobility shifted bands relative to the 70-kDa subunit of hRPA were cut from the gel, reduced, carboxyamidomethylated, and digested with trypsin (Figure 6.11). Extracted nucleopeptides were eluted from a NAC-52 anion exchange cartridge, concentrated and desalted using a Centricon-3, further concentrated by vacuum centrifugation to approximately 10 μ L, and analyzed by MALDI MS, as described under "Experimental Procedures." (A) MALDI mass spectrum of nucleopeptides purified from a crosslinked gel band of lowest gel mobility (XL-1 from Figure 6.11). (B) MALDI mass spectrum of nucleopeptides purified from a crosslinked band of second lowest gel mobility (XL-2 from Figure 6.12).

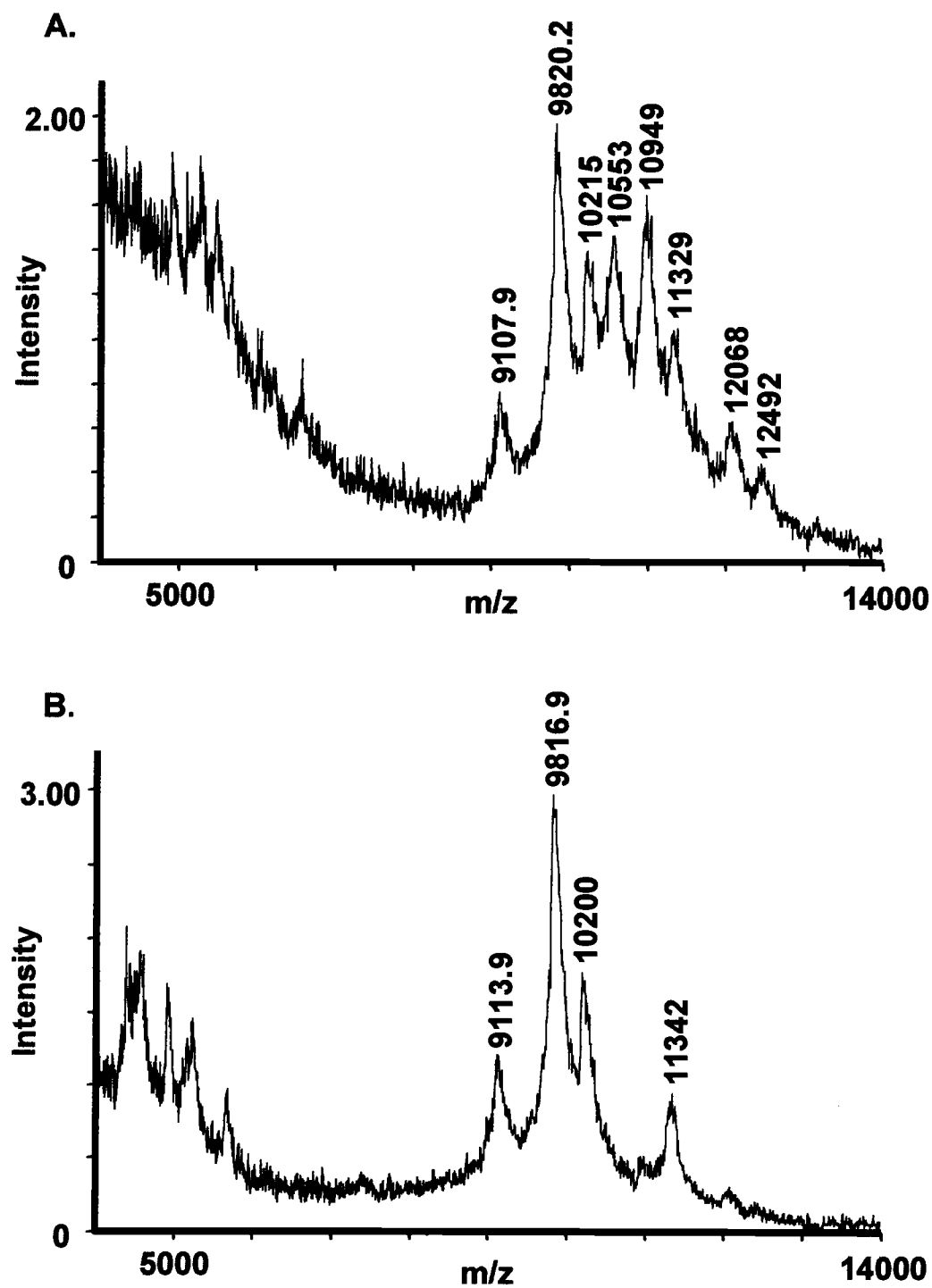


Figure 6.12. MALDI mass spectrum of purified hRPA70 $\times$ dT₃₀ tryptic nucleopeptides.

to the gel band XL-1 (ion species 10553, 10949, 12068, and 12492 in Figure 6.12 A).

Tentative identification of the nucleopeptides was attempted by subtracting the molecular weight of dT₃₀ from the measured ion species and then comparing the resulting molecular weights to a tryptic map of hRPA70. Closer analysis of the nucleopeptide spectrum showed that the mass assignments were made to very broad peaks (~150 m.u. at FWHM) which drastically reduces the accuracy of the assignments. This lack of mass accuracy coupled with the potential for truncation of the peptide portion of the nucleopeptide (as observed for the T6 nucleopeptide of Ung; Table 4.1 and Table 4.2) prevented the tentative identification of the nucleopeptides with any level of certainty. More thorough purification of the hRPA70 × dT₃₀ nucleopeptides will be required in order to use MALDI MS in the nucleopeptide identification.

6.1.8 Tandem Mass Spectrometry of Nuclease P1-digested hRPA70 × dT₃₀ Nucleopeptides

Even though tentative identification of the hRPA70 × dT₃₀ nucleopeptides by MALDI MS could not be made due to issues of sample purity, LC/ESI ion-trap mass spectrometry experiments were attempted. The hRPA70 × dT₃₀ nucleopeptides were digested with nuclease P1 and then analyzed. The LC/ESI MS spectrum of this sample did not provide any mass spectra that could be

identified as potential nuclease digested species. This lack of information could be a result of low sample quantity or a result of labile crosslinks breaking during the nuclease P1 digestion.

6.2 Discussion

The work here begins UV catalyzed crosslinking experiments to structurally characterize the DNA binding domain of hRPA. This characterization was conducted in the same manner as that of Ung; covalent protein-DNA complexes were formed by UV-crosslinking, the complexes were proteolytically and nucleolytically digested, and peptide-DNA fragments were isolated and analyzed by mass spectrometry. The ultimate goal was to use collected mass spectrometry data to identify peptide regions and specific amino acids within hRPA that take part in crosslinking.

As part of pursuing this goal, experiments to purify the protein, determine an optimum crosslinking time, and generate a nucleopeptide map of a hRPA-oligonucleotide complex were performed to gauge the feasibility of this project. From these experiments it was found that milligram quantities of hRPA could be readily purified, an optimum irradiation time of one minute for crosslinking hRPA to dT₁₀, dT₃₀, and dT₇₀ was established, and both peptide and nucleopeptide maps of the protein were generated. The combination of these results indicates a strong likelihood of success to using UV-catalyzed crosslinking to map the DNA-binding domain of hRPA.

The presence of multimeric protein-DNA complexes forming between hRPA and single-stranded oligonucleotides was not expected. An electrophoretic mobility shift analysis used to measure the DNA capability of purified hRPA revealed three distinct mobility shifted bands (Figure 6.7). The formation of super-mobility shifted complexes increased with saturating amounts of hRPA, as compared to oligonucleotide, suggesting these complexes were multiple hRPA molecules bound to single dT₃₀ molecules. Other research groups have reported the formation of multimeric hRPA complexes with oligonucleotides. EMSA and gel permeation chromatography studies have shown the presence of two hRPA molecules bound to oligonucleotides of 70 nucleotides in length, but they have not shown the presence of multimeric complexes with oligonucleotides 30 nucleotides in length (46, 66). From glutaraldehyde crosslinking studies, however, up to three hRPA molecules crosslinked a 30-mer oligonucleotide (12). This study estimated the minimum DNA binding size of a single hRPA molecule to be 8-10 nucleotides in length and it also detected an hRPA complex that bound single-stranded DNA in 30-nucleotide steps. The minimum DNA binding site size suggests the possibility of up to three hRPA molecules binding to one 30 mer oligonucleotide. This is one possible explanation of the electrophoretic mobility analysis results observed here (Figure 6.7). A major deficiency in the electrophoretic mobility, gel permeation, and glutaraldehyde crosslinking analysis is that these techniques do not reveal the specific nature of the multimeric hRPA-dT₃₀ complex. These analyses can not fully distinguish between the possibility of three hRPA molecules each directly interacting with a single oligonucleotide

molecule or the possibility of three hRPA molecules interacting with one another (protein-protein interactions) and only one of the hRPA molecules directly interacting with an oligonucleotide molecule. The later possibility is a second explanation of the electrophoretic mobility results observed here.

Crosslinking experiments of hRPA with oligonucleotides of increasing size also showed the presence of multiple protein-DNA complexes. Crosslinking complexes were observed on denaturing polyacrylamide gels; thus, these complexes were separated into subunits of hRPA that contained covalently attached oligonucleotides. As the size of the oligonucleotide increased, additional crosslinking products were observed. For dT₁₀, a single crosslinking product was observed (Figure 6.9A). Compared to other crosslinking experiments analyzed on gels of this acrylamide percentage, this product was presumed to be a hRPA70 × dT₁₀ complex. As the size of the oligonucleotide increased to dT₃₀ or dT₇₀, three distinct crosslinking products were observed (Figure 6.9B and Figure 6.9C). Since these gels were not stained to observe the protein bands, molecular weight standards were not available to estimate the molecular weight of the crosslinking products. Therefore, it is difficult to ascertain the identity of these products, but it is obvious that multimeric complexes form as the length of the oligonucleotide increases. Similar to the electrophoretic shift analysis, these complexes could be explained as either multiple oligonucleotides attached to a subunit of hRPA or multiple hRPA subunits attached to a single oligonucleotide.

Currently, structural information at the amino acid level for hRPA is limited to a x-ray crystal structure of a truncated form of hRPA70 (14), a x-ray crystal structure of the hRPA32-hRPA14 complex (13), and characterization of deletion/site-directed mutants of hRPA. Even though these studies have provided valuable insight into the structural elements of hRPA that bind to oligonucleotides, very few structural experiments have been performed on heterotrimeric hRPA. Future crosslinking studies have the potential to uncover the structural features of the heterotrimeric protein that facilitate the increased binding affinity of the protein to oligonucleotides of increasing size, reveal the role that aromatic amino acids within hRPA70 play in the amino acid-nucleic acid interactions of this protein, as well as give insight into the roles each subunit plays in binding to DNA. A better understanding of how hRPA binds to DNA will elucidate the function of this protein in DNA metabolism.

BIBLIOGRAPHY

1. Adachi, Y. and Laemmli, U.K. (1992) Identification of nuclear pre-replication centers poised for DNA synthesis in *Xenopus* egg extracts: immunolocalization study of replication protein A. *J. Cell. Biol.* **119**, 1-15.
2. Adamczyk, M., Gebler, J.C., and Wu, J. (1999) Charge derivatization of peptides to simplify their sequencing with an ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* **13**, 1413-22.
3. Alexander, P. and Moroson, H. (1962) Cross-linking of deoxyribonucleic acid to protein following ultraviolet irradiation of different cells. *Nature* **194**, 882-3.
4. Allen, T.D., Wick, K.L., and Matthews, K.S. (1991) Identification of amino acids in lac repressor protein cross-linked to operator DNA specifically substituted with bromodeoxyuridine. *J. Biol. Chem.* **266**, 6113-9.
5. Barrett, T.E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L.H. (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **92**, 117-29.
6. 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**, 21870-9.
7. Bennett, S.E. and Mosbaugh, D.W. (1992) Characterization of the *Escherichia coli* uracil-DNA glycosylase-inhibitor protein complex. *J. Biol. Chem.* **267**, 22512-21.
8. Bennett, S.E., Sanderson, R.J., and Mosbaugh, D.W. (1995) Processivity of *Escherichia coli* and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration. *Biochemistry* **34**, 6109-19.
9. Bennett, S.E., Schimerlik, M.I., and Mosbaugh, D.W. (1993) Kinetics of the uracil-DNA glycosylase/inhibitor protein association. Ung interaction with Ugi, nucleic acids, and uracil compounds. *J. Biol. Chem.* **268**, 26879-85.
10. Biemann, K. (1990) Nomenclature for peptide fragment ions. *Methods Enzymol.* **193**, 886-7.

11. Bier, M.E. and Schwartz, J.C., *Electrospray-ionization quadrupole ion-trap mass spectrometry*, in *Electrospray Ionization Mass Spectrometry*, R.B. Cole, Editor. 1997, John Wiley & Sons: New York. p. 235-89.
12. Blackwell, L.J. and Borowiec, J.A. (1994) Human replication protein A binds single-stranded DNA in two distinct complexes. *Mol. Cell. Biol.* **14**, 3993-4001.
13. Bochkarev, A., Bochkareva, E., Frappier, L., and Edwards, A.M. (1999) The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *Embo J.* **18**, 4498-504.
14. 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**, 176-81.
15. Bochkareva, E., Frappier, L., Edwards, A.M., and Bochkarev, A. (1998) The RPA32 subunit of human replication protein A contains a single-stranded DNA-binding domain. *J. Biol. Chem.* **273**, 3932-6.
16. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-54.
17. Brill, S.J. and Stillman, B. (1989) Yeast replication factor-A functions in the unwinding of the SV40 origin of DNA replication. *Nature* **342**, 92-5.
18. Brown, G.W., Melendy, T.E., and Ray, D.S. (1992) Conservation of structure and function of DNA replication protein A in the trypanosomatid *Crithidia fasciculata*. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10227-31.
19. Budowsky, E.I. and Abdurashidova, G.G. (1989) Polynucleotide-protein cross-links induced by ultraviolet light and their use for structural investigation of nucleoproteins. *Prog. Nucleic Acid Res. Mol. Biol.* **37**, 1-65.
20. Budowsky, E.I., Axentyeva, M.S., Abdurashidova, G.G., Simukova, N.A., and Rubin, L.B. (1986) Induction of polynucleotide-protein cross-linkages by ultraviolet irradiation. Peculiarities of the high-intensity laser pulse irradiation. *Eur. J. Biochem.* **159**, 95-101.
21. Burlet, O., Orkiszewski, R.S., Ballard, K.D., and Gaskell, S.J. (1992) Charge promotion of low-energy fragmentations of peptide ions. *Rapid Commun. Mass Spectrom.* **6**, 658-62.

22. Burns, J.L., Guzder, S.N., Sung, P., Prakash, S., and Prakash, L. (1996) An affinity of human replication protein A for ultraviolet-damaged DNA. *J. Biol. Chem.* **271**, 11607-10.
23. Calderone, T.L., Stevens, R.D., and Oas, T.G. (1996) High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. *J. Mol. Biol.* **262**, 407-12.
24. Chen, S., Lee, T.D., Legesse, K., and Shively, J.E. (1986) Identification of the arylazido-beta-alanyl-NAD⁺-modified site in rabbit muscle glyceraldehyde-3-phosphate dehydrogenase by microsequencing and fast atom bombardment mass spectrometry. *Biochemistry* **25**, 5391-5.
25. Cheng, X., Harms, A.C., Goudreau, P.N., Terwilliger, T.C., and Smith, R.D. (1996) Direct measurement of oligonucleotide binding stoichiometry of gene V protein by mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7022-7.
26. Cheng, X., Morin, P.E., Harms, A.C., Bruce, J.E., Ben-David, Y., and Smith, R.D. (1996) Mass spectrometric characterization of sequence-specific complexes of DNA and transcription factor PU.1 DNA binding domain. *Anal. Biochem.* **239**, 35-40.
27. Connor, D.A., Falick, A.M., and Shetlar, M.D. (1998) UV light-induced cross-linking of nucleosides, nucleotides and a dinucleotide to the carboxy-terminal heptad repeat peptide of RNA polymerase II as studied by mass spectrometry. *Photochem. Photobiol.* **68**, 1-8.
28. 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.
29. Cotter, R.J. (1992) Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal. Chem.* **64**, 1027A-39A.
30. Cox, K.A., Morris, M., Whiting, A., and Gaskell, S.J. (1996) Role of the Site Protonation in the Low-Energy Decomposition of Gas-Phase Peptide Ions. *J. Am. Soc. Mass Spectrom.* **7**, 522-31.
31. Dongre, A.R., Jones, J.L., Somogyi, A., and Wysocki, V.H. (1996) Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J Am. Chem. Soc.* **118**, 8365-74.

32. Duncan, B.K. and Chambers, J.A. (1984) The cloning and overproduction of *Escherichia coli* uracil-DNA glycosylase. *Gene* **28**, 211-9.
33. Eng, J.K., McCormack, A.L., and Yates, J.R., 3rd. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-89.
34. Fenselau, C. (1997) MALDI MS and strategies for protein analysis. *Anal. Chem.* **69**, 661A-5A.
35. Fitzgerald, M.C. and Smith, L.M. (1995) Mass spectrometry of nucleic acids: the promise of matrix-assisted laser desorption-ionization (MALDI) mass spectrometry. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 117-40.
36. Forman, M.D., Stack, R.F., Masters, P.S., Hauer, C.R., and Baxter, S.M. (1998) High level, context dependent misincorporation of lysine for arginine in *Saccharomyces cerevisiae* a1 homeodomain expressed in *Escherichia coli*. *Protein Sci.* **7**, 500-3.
37. Friedberg, E.C., Walker, G.C., and Siede, W., *DNA Repair and Mutagenesis*. 1995, Washington, D.C.: ASM Press.
38. 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 Sci.* **8**, 2806-12.
39. Gomes, X.V., Henricksen, L.A., and Wold, M.S. (1996) Proteolytic mapping of human replication protein A: evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA. *Biochemistry* **35**, 5586-95.
40. Gomes, X.V. and Wold, M.S. (1995) Structural analysis of human replication protein A. Mapping functional domains of the 70-kDa subunit. *J. Biol. Chem.* **270**, 4534-43.
41. Gomes, X.V. and Wold, M.S. (1996) Functional domains of the 70-kilodalton subunit of human replication protein A. *Biochemistry* **35**, 10558-68.
42. Gorman, M.A., Morera, S., Rothwell, D.G., de La Fortelle, E., Mol, C.D., Tainer, J.A., Hickson, I.D., and Freemont, P.S. (1997) The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. *Embo J.* **16**, 6548-58.

43. Gross, J., Leisner, A., Hillenkamp, F., Hahner, S., Karas, M., Schafer, J., Lutzenkirchen, F., and Nordhoff, E. (1998) Investigations of the metastable decay of DNA under ultraviolet matrix-assisted laser desorption/ionization conditions with post-source-decay analysis and hydrogen/deuterium exchange. *J. Am. Soc. Mass Spectrom.* **9**, 866-78.
44. Guo, B. (1999) Mass Spectrometry in DNA analysis. *Anal. Chem.* **71**, 333R-7R.
45. He, Z., Henricksen, L.A., Wold, M.S., and Ingles, C.J. (1995) RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* **374**, 566-9.
46. Henricksen, L.A., Umbricht, C.B., and Wold, M.S. (1994) Recombinant replication protein A: expression, complex formation, and functional characterization [published erratum appears in J. Biol. Chem. 1994 Jun 10;269(23):16519]. *J. Biol. Chem.* **269**, 11121-32.
47. Hess, D., Covey, T.C., Winz, R., Brownsey, R.W., and Aebersold, R. (1993) Analytical and micropreparative peptide mapping by high performance liquid chromatography/electrospray mass spectrometry of proteins purified by gel electrophoresis. *Protein Sci.* **2**, 1342-51.
48. Higley, M. and Lloyd, R.S. (1993) Processivity of uracil DNA glycosylase. *Mutat. Res.* **294**, 109-16.
49. Hillenkamp, F. and Karas, M. (1990) Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol.* **193**, 280-95.
50. Hillenkamp, F., Karas, M., Beavis, R.C., and Chait, B.T. (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* **63**, 1193A-203A.
51. Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R., Evertsz, E.M., and von Hippel, P.H. (1991) Laser Cross-linking of protein-nucleic acid complexes. *Methods Enzymol.* **208**, 211-58.
52. Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R., and von Hippel, P.H. (1993) Laser cross-linking of proteins to nucleic acids. I. Examining physical parameters of protein-nucleic acid complexes. *J. Biol. Chem.* **268**, 15712-20.
53. Huang, Z.H., Wu, J., Roth, K.D., Yang, Y., Gage, D.A., and Watson, J.T. (1997) A picomole-scale method for charge derivatization of peptides for sequence analysis by mass spectrometry. *Anal. Chem.* **69**, 137-44.

54. Hunt, D.F., Yates, J.R.d., Shabanowitz, J., Winston, S., and Hauer, C.R. (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6233-7.
55. Iftode, C., Daniely, Y., and Borowiec, J.A. (1999) Replication protein A (RPA): the eukaryotic SSB. *Crit. Rev. Biochem. Mol. Biol.* **34**, 141-80.
56. Jellinek, T. and Johns, R.B. (1970) The mechanism of photochemical addition of cysteine to uracil and formation of hihydrouracil. *Photochem. Photobiol.* **11**, 349-59.
57. Jensen, O.N., Barofsky, D.F., Young, M.C., von Hippel, P.H., Swenson, S., and Seifried, S.E. (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.
58. Jensen, O.N., Barofsky, D.F., Young, M.C., von Hippel, P.H., Swenson, S., and Seifried, S.E., *Mass spectrometric protocol for the analysis of UV-crosslinked protein-nucleic acid complexes*, in *Techniques in Protein Chemistry V*, J.W. Crabb, Editor. 1994, Academic Press: San Diego. p. 27-37.
59. Jensen, O.N., Kulkarni, S., Aldrich, J.V., and Barofsky, D.F. (1996) Characterization of peptide-oligonucleotide heteroconjugates by mass spectrometry. *Nucleic Acids Res.* **24**, 3866-72.
60. 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.
61. Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**, 2299-301.
62. Katz, D.L. and Barofsky, D.F. *A new design for a MALDI tandem time-of-flight mass spectrometer* in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*. 1999. Dallas, TX.
63. Kaufmann, R. (1995) Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology. *J. Biotechnol.* **41**, 155-75.
64. Kavli, B., Slupphaug, G., Mol, C.D., Arvai, A.S., Peterson, S.B., Tainer, J.A., and Krokan, H.E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil- DNA glycosylase. *Embo J.* **15**, 3442-7.

65. Keough, T., Youngquist, R.S., and Lacey, M.P. (1999) A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7131-6.
66. Kim, C., Paulus, B.F., and Wold, M.S. (1994) Interactions of human replication protein A with oligonucleotides. *Biochemistry* **33**, 14197-206.
67. 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**, 3050-9.
68. Kim, C. and Wold, M.S. (1995) Recombinant human replication protein A binds to polynucleotides with low cooperativity. *Biochemistry* **34**, 2058-64.
69. Kim, D.K., Stigger, E., and Lee, S.H. (1996) Role of the 70-kDa subunit of human replication protein A (I). Single-stranded dna binding activity, but not polymerase stimulatory activity, is required for DNA replication. *J. Biol. Chem.* **271**, 15124-9.
70. Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-69.
71. Kneale, G.G., ed. *DNA-Protein Interactions*. Methods in Molecular Biology, ed. J.M. Walker. Vol. 30. 1994, Humana Press: Totowa. 427.
72. Krokan, H.E., Standal, R., and Slupphaug, G. (1997) DNA glycosylases in the base excision repair of DNA. *Biochem. J.* **325**, 1-16.
73. Kubasek, W.L., Spann, D., and Hockensmith, J.W. (1993) Laser cross-linking of proteins to nucleic acids: photodegradation and alternative photoproducts of the bacteriophage T4 gene 32 protein. *Photochem. Photobiol.* **58**, 1-10.
74. Lau, A.Y., Scharer, O.D., Samson, L., Verdine, G.L., and Ellenberger, T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249-58.
75. Lavrik, O.I., Nasheuer, H.P., Weisschart, K., Wold, M.S., Prasad, R., Beard, W.A., Wilson, S.H., and Favre, A. (1998) Subunits of human replication protein A are crosslinked by photoreactive primers synthesized by DNA polymerases. *Nucleic Acids Res.* **26**, 602-7.

76. Lin, Y.L., Chen, C., Keshav, K.F., Winchester, E., and Dutta, A. (1996) Dissection of functional domains of the human DNA replication protein complex replication protein A. *J. Biol. Chem.* **271**, 17190-8.
77. Lindahl, T. (1980) Uracil-DNA glycosylase from *Escherichia coli*. *Methods Enzymol.* **65**, 284-90.
78. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-15.
79. Lindahl, T., Karran, P., and Wood, R.D. (1997) DNA excision repair pathways. *Curr. Opin. Genet. Dev.* **7**, 158-69.
80. Lindahl, T., Ljungquist, S., Siebert, W., Nyberg, B., and Sperens, B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. *J. Biol. Chem.* **252**, 3286-94.
81. Liu, C., Tolic, L.P., Hofstadler, S.A., Harms, A.C., Smith, R.D., Kang, C., and Sinha, N. (1998) Probing RegA/RNA interactions using electrospray ionization-fourier transform ion cyclotron resonance-mass spectrometry. *Anal. Biochem.* **262**, 67-76.
82. Liu, D., Prasad, R., Wilson, S.H., DeRose, E.F., and Mullen, G.P. (1996) Three-dimensional solution structure of the N-terminal domain of DNA polymerase beta and mapping of the ssDNA interaction interface. *Biochemistry* **35**, 6188-200.
83. McLuckey, S.A., Van Berkel, G.J., and Glish, G.L. (1992) Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **3**, 60-70.
84. 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**, 552-8.
85. Meisenheimer, K.M. and Koch, T.H. (1997) Photocross-linking of nucleic acids to associated proteins. *Crit. Rev. Biochem. Mol. Biol.* **32**, 101-40.
86. Merril, C.R., Goldman, D., Sedman, S.A., and Ebert, M.H. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211**, 1437-8.
87. Miller, P.E. and Denton, M.B. (1986) The quadrupole mass filter: basic operation concepts. *J. Chem. Ed.* **63**, 617-22.

88. Mirzabekov, A.D., Bavykin, S.G., Belyavsky, A.V., Karpov, V.L., Preobrazhenskaya, O.V., Shick, V.V., and Ebralidse, K.K. (1989) Mapping DNA-protein interactions by cross-linking. *Methods Enzymol.* **170**, 386-408.
89. Mitsis, P.G., Kowalczykowski, S.C., and Lehman, I.R. (1993) A single-stranded DNA binding protein from *Drosophila melanogaster*: characterization of the heterotrimeric protein and its interaction with single-stranded DNA. *Biochemistry* **32**, 5257-66.
90. Mol, C.D., Arvai, A.S., Sanderson, R.J., Slupphaug, G., Kavli, B., Krokan, H.E., Mosbaugh, D.W., and Tainer, J.A. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701-8.
91. Mol, C.D., Arvai, A.S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H.E., and Tainer, J.A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80**, 869-78.
92. Mol, C.D., Kuo, C.F., Thayer, M.M., Cunningham, R.P., and Tainer, J.A. (1995) Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature* **374**, 381-6.
93. Mosbaugh, D.W. and Bennett, S.E. (1994) Uracil-excision DNA repair. *Prog. Nucleic Acid Res. Mol. Biol.* **48**, 315-70.
94. Murzin, A.G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *Embo J.* **12**, 861-7.
95. Oakley, B.R., Kirsch, D.R., and Morris, N.R. (1980) A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Anal. Biochem.* **105**, 361-3.
96. 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**, 5214-26.
97. Park, H.W., Kim, S.T., Sancar, A., and Deisenhofer, J. (1995) Crystal structure of DNA photolyase from *Escherichia coli*. *Science* **268**, 1866-72.
98. Parker, A.E., Clyne, R.K., Carr, A.M., and Kelly, T.J. (1997) The *Schizosaccharomyces pombe* rad11+ gene encodes the large subunit of replication protein A. *Mol. Cell. Biol.* **17**, 2381-90.

99. Patrick, S.M. and Turchi, J.J. (1998) Human replication protein A preferentially binds cisplatin-damaged duplex DNA in vitro. *Biochemistry* **37**, 8808-15.
100. Patrick, S.M. and Turchi, J.J. (1999) Replication protein A (RPA) binding to duplex cisplatin-damaged DNA is mediated through the generation of single-stranded DNA. *J. Biol. Chem.* **274**, 14972-8.
101. 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**, 430-4.
102. Piyadasa, C.K., 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**, 1655-64.
103. Prasad, R., Kumar, A., Widen, S.G., Casas-Finet, J.R., and Wilson, S.H. (1993) Identification of residues in the single-stranded DNA-binding site of the 8-kDa domain of rat DNA polymerase beta by UV cross-linking. *J. Biol. Chem.* **268**, 22746-55.
104. Purmal, A.A., Wallace, S.S., and Kow, Y.W. (1996) The phosphodiester bond 3' to a deoxyuridine residue is crucial for substrate binding for uracil DNA N-glycosylase. *Biochemistry* **35**, 16630-7.
105. 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**, 331-46.
106. Qian, M.G., Zhang, Y., and Lubman, D.M. (1995) Collision-induced dissociation of multiply charged peptides in an ion-trap storage/reflectron time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **9**, 1275-82.
107. Qin, J. and Chait, B.T. (1997) Identification and characterization of posttranslational modifications of proteins by MALDI ion trap mass spectrometry. *Anal. Chem.* **69**, 4002-9.
108. 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**, 4880-7.

109. Reinisch, K.M., Chen, L., Verdine, G.L., and Lipscomb, W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143-53.
110. 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**, 505-15.
111. Roth, K.D., Huang, Z.H., Sadagopan, N., and Watson, J.T. (1998) Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrom. Rev.* **17**, 255-74.
112. Sadagopan, N. and Watson, J.T. (2000) Investigation of the tris(trimethoxyphenyl)phosphonium acetyl charged derivatives of peptides by electrospray ionization mass spectrometry and tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **11**, 107-19.
113. Saito, I., Sugiyama, H., and Matsuura, T. (1983) Isolation and characterization of the thymine-lysine adduct in UV-irradiated nuclei. The role of thymine-lysine photoaddition in photocross-linking of proteins to DNA. *J. Am. Chem. Soc.* **105**, 6989-91.
114. Sannes-Lowery, K.A., Hu, P., Mack, D.P., Mei, H.Y., and Loo, J.A. (1997) HIV-1 Tat peptide binding to TAR RNA by electrospray ionization mass spectrometry. *Anal. Chem.* **69**, 5130-5.
115. Sauer, R.T., ed. *Protein-DNA Interactions*. Methods in Enzymology, ed. J.N. Abelson and M.I. Simon. Vol. 208. 1991, Academic Press: San Diego. 700.
116. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487-93.
117. Savva, R. and Pearl, L.H. (1993) Crystallization and preliminary X-ray analysis of the uracil-DNA glycosylase DNA repair enzyme from herpes simplex virus type 1. *J. Mol. Biol.* **234**, 910-2.
118. Savva, R. and Pearl, L.H. (1995) Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. *Nat. Struct. Biol.* **2**, 752-7.

119. Seroussi, E. and Lavi, S. (1993) Replication protein A is the major single-stranded DNA binding protein detected in mammalian cell extracts by gel retardation assays and UV cross-linking of long and short single-stranded DNA molecules. *J. Biol. Chem.* **268**, 7147-54.
120. Shaw, A.A., Falick, A.M., and Shetlar, M.D. (1992) Photoreactions of thymine and thymidine with N-acetyltyrosine. *Biochemistry* **31**, 10976-83.
121. Shaw, A.A. and Shetlar, M.D. (1989) Ring-opening photoreactions of cytosine and 5-methylcytosine with aliphatic alcohols. *Photochem. Photobiol.* **49**, 267-71.
122. Shetlar, M.D., Carbone, J., Steady, E., and Hom, K. (1984) Photochemical addition of amino acids and peptides to polyuridylic acid. *Photochem. Photobiol.* **39**, 141-4.
123. Shetlar, M.D., Christensen, J., and Hom, K. (1984) Photochemical addition of amino acids and peptides to DNA. *Photochem. Photobiol.* **39**, 125-33.
124. Shetlar, M.D., Home, K., Carbone, J., Moy, D., Steady, E., and Watanabe, M. (1984) Photochemical addition of amino acids and peptides to homopolyribonucleotides of the major DNA bases. *Photochem. Photobiol.* **39**, 135-40.
125. Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H., and Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14440-5.
126. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850-8.
127. Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N.M., Haug, T., Levine, D.W., and Krokan, H.E. (1995) Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry* **34**, 128-38.
128. 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**, 87-92.
129. Smith, K.C. (1962) Dose dependent decrease in extractability of DNA from bacteria following irradiation with ultraviolet light or with visible light plus dyes. *Biochem. Biophys. Res. Commun.* **8**, 157-63.

130. Stults, J.T. (1995) Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). *Curr. Opin. Struct. Biol.* **5**, 691-8.
131. Tang, C.M., Tomkinson, A.E., Lane, W.S., Wold, M.S., and Seto, E. (1996) Replication protein A is a component of a complex that binds the human metallothionein IIA gene transcription start site. *J. Biol. Chem.* **271**, 21637-44.
132. Tang, X.J. and Boyd, R.K. (1992) An investigation of fragmentation mechanisms of doubly protonated tryptic peptides. *Rapid Commun. Mass Spectrom.* **6**, 651-7.
133. Tang, X.J., Thibault, P., and Boyd, R.K. (1993) Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Anal. Chem.* **65**, 2824-34.
134. Thayer, M.M., Ahern, H., Xing, D., Cunningham, R.P., and Tainer, J.A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *Embo J.* **14**, 4108-20.
135. 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**, 14547-55.
136. van Holde, K.E., Johnson, W.C., and Ho, P.S., *Principles of Physical Biochemistry*. 1998, Upper Saddle River: Prentice-Hall. 657.
137. van Oostveen, I., Ducret, A., and Aebersold, R. (1997) Colloidal silver staining of electroblotted proteins for high sensitivity peptide mapping by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal. Biochem.* **247**, 310-8.
138. Varshney, U., Hutcheon, T., and van de Sande, J.H. (1988) Sequence analysis, expression, and conservation of Escherichia coli uracil DNA glycosylase and its gene (ung). *J. Biol. Chem.* **263**, 7776-84.
139. Varshney, U. and van de Sande, J.H. (1991) Specificities and kinetics of uracil excision from uracil-containing DNA oligomers by Escherichia coli uracil DNA glycosylase. *Biochemistry* **30**, 4055-61.

140. Vassilyev, D.G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. (1995) Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **83**, 773-82.
141. Veenstra, T.D. (1999) Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes. *Biochem. Biophys. Res. Commun.* **257**, 1-5.
142. Walther, A.P. and et al. (1999) Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit. *Biochemistry* **38**, 3963-73.
143. Wang, Q., Shoeman, R., and Traub, P. (2000) Identification of the 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**, 6645-51.
144. Wang, Y. and Adzuma, K. (1996) Differential proximity probing of two DNA binding sites in the Escherichia coli recA protein using photo-cross-linking methods. *Biochemistry* **35**, 3563-71.
145. Wang, Y., Taylor, J.S., and Gross, M.L. (1999) Nuclease P1 digestion combined with tandem mass spectrometry for the structure determination of DNA photoproducts. *Chem. Res. Toxicol.* **12**, 1077-82.
146. Williams, 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**, 516-39.
147. Wilm, M. and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1-8.
148. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano- electrospray mass spectrometry. *Nature* **379**, 466-9.
149. 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**, 1834-8.
150. 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**, 2523-7.

151. 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**, 3643-7.
152. Wong, D.L., Pavlovich, J.G., and Reich, N.O. (1998) Electrospray ionization mass spectrometric characterization of photocrosslinked DNA-EcoRI DNA methyltransferase complexes. *Nucleic Acids Res.* **26**, 645-9.
153. 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* **35**, 13-24.
154. Xu, N., Pasa-Tolic, L., Smith, R.D., Ni, S., and Thrall, B.D. (1999) Electrospray ionization-mass spectrometry study of the interaction of cisplatin-adducted oligonucleotides with human XPA minimal binding domain protein. *Anal. Biochem.* **272**, 26-33.
155. Yates, J.R., 3rd, Eng, J.K., and McCormack, A.L. (1995) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal. Chem.* **67**, 3202-10.
156. Yates, J.R., 3rd, Eng, J.K., McCormack, A.L., and Schieltz, D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**, 1426-36.
157. Yates, J.R., 3rd, McCormack, A.L., and Eng, J. (1996) Mining genomes with MS. *Anal. Chem.* **68**, 534A-40A.
158. Zhang, S.P., Zubay, G., and Goldman, E. (1991) Low-usage codons in Escherichia coli, yeast, fruit fly and primates. *Gene* **105**, 61-72.

APPENDIX

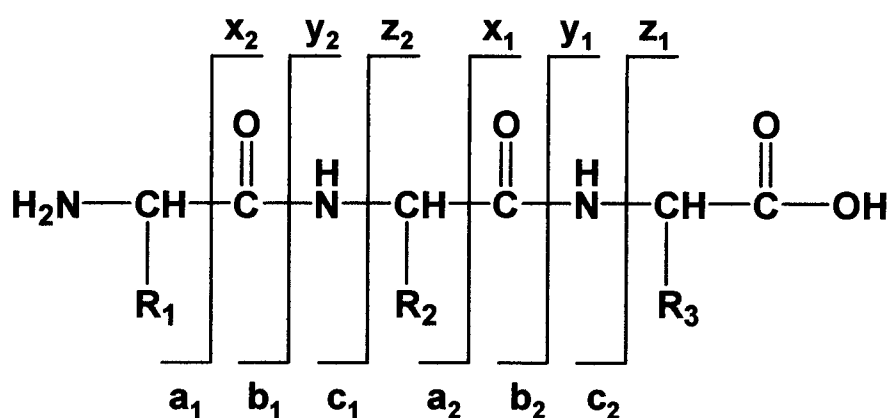


Figure A.1. Nomenclature for common peptide backbone fragments produced from collisionally induced dissociation. As described by Beermann (10), fragment ions that retain charge on the N-terminal portion of the original peptide are labeled a_n , b_n , and c_n while fragment ions that retain charge on the C-terminal portion of the original peptide are labeled x_n , y_n , and z_n . As described by Jensen *et al.* (59), peptide fragments of a nucleopeptide are also designated a_n , b_n , c_n , x_n , y_n , and z_n , while peptide fragments of a nucleopeptide that contain covalently crosslinked oligonucleotide are designated a_n^* , b_n^* , c_n^* , x_n^* , y_n^* , and z_n^* .

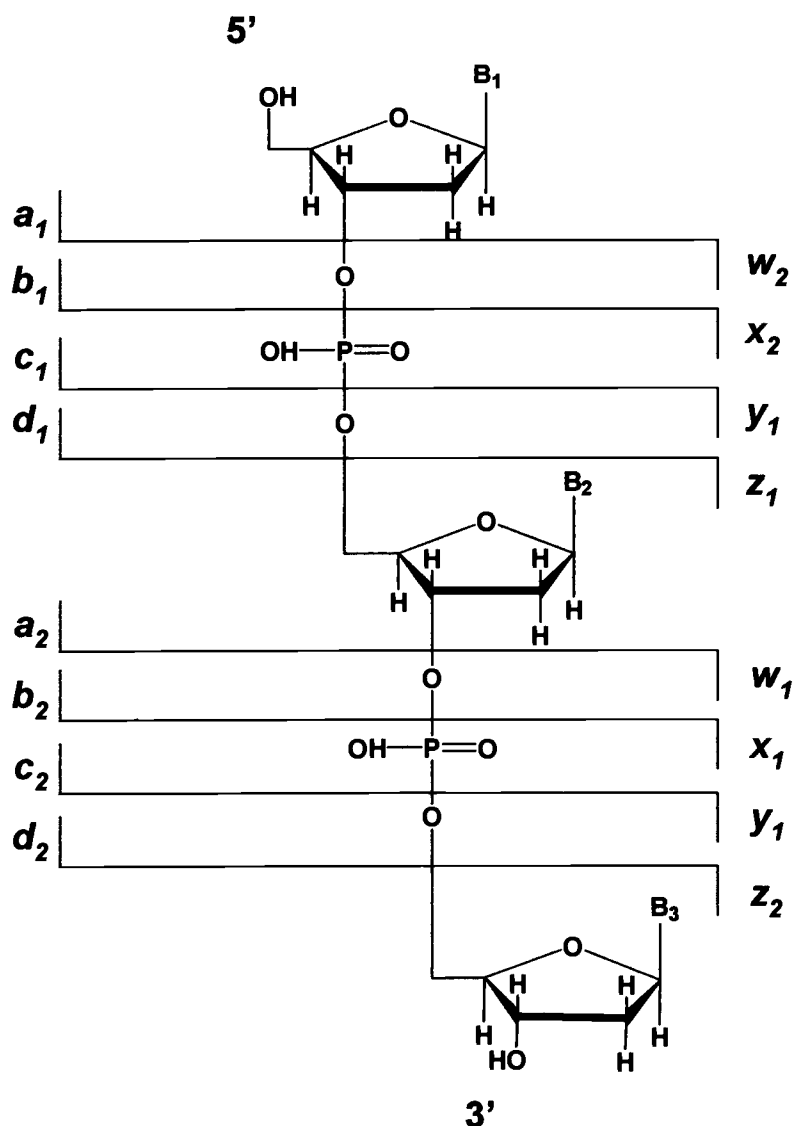


Figure A.2. Nomenclature for common oligonucleotide backbone fragments produced from low energy collisionally induced dissociation. Oligonucleotide fragment ions were labeled using the nomenclature proposed by McLuckey, *et al* (83). Due to the similarities between the nomenclature of peptide and oligonucleotide fragmentation, oligonucleotide fragmentation products are italicized. The four possible cleavages along the phosphodiester backbone are indicated by *a_n*, *b_n*, *c_n*, and *d_n* for fragments containing the 5'-OH group and *w_n*, *x_n*, *y_n*, and *z_n* for fragments containing the 3'-OH group. Oligonucleotide fragments of a nucleopeptide are also designated *a_n*, *b_n*, *c_n*, *d_n*, *w_n*, *x_n*, *y_n*, and *z_n*, while oligonucleotide fragments from a nucleopeptide that contain covalently crosslinked peptide are designated *a_n^{*}*, *b_n^{*}*, *c_n^{*}*, *d_n^{*}*, *w_n^{*}*, *x_n^{*}*, *y_n^{*}*, and *z_n^{*}* (59).