AN ABSTRACT OF THE THESIS OF

Ole Nørregaard Jensen for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on October 20, 1994. Title: Characterization of Photochemically Cross-linked Protein-Nucleic Acid Complexes by Mass Spectrometry.

Abstract approved: _____

Douglas F. Barofsky

A novel protocol for the study of protein-nucleic acid interactions is presented and demonstrated to be feasible. The protocol combines photochemical crosslinking techniques and mass spectrometric methods into a new strategy for identifying protein domains or amino acid residues that are in close contact with nucleic acid in protein-nucleic acid complexes. Identifying nucleic acid binding domains in proteins provides a starting point for understanding structure-function relationships in protein-nucleic acid complexes.

The protocol can be divided into three parts: 1) Crosslinking of the protein-nucleic acid complex by irradiation with ultraviolet light and subsequently verifying the crosslinking by mass spectrometry; 2) Mass spectrometric peptide mapping of crosslinked protein-nucleic acid complexes to identify crosslinked peptide-nucleic acid hybrids; 3) Tandem mass spectrometric sequencing of peptide-nucleic acid hybrids to localize the crosslinked amino acid residue(s).

The experimental data described in this dissertation documents our efforts to establish and implement this analytical protocol. Using several different protein-nucleic acid systems and different crosslinking techniques, we have demonstrated the feasibility of a mass spectrometric based approach to structurally characterize UV-crosslinked protein-nucleic acid complexes. Matrix-assisted laser desorption/ionization mass spectrometry was for the first time demonstrated to be highly effective for detection and molecular weight determination of intact, UV-crosslinked protein-nucleic acid complexes and for molecular weight determination of synthetic and UV-crosslinked peptide-nucleic acid hybrids. Electrospray ionization mass spectrometry and tandem mass spectrometry was demonstrated to be effective for analysis of synthetic peptide-nucleic acid hybrids and, in conjunction with HPLC, for peptide mapping of a protein. The first application of MALDI mass spectrometry to the characterization of crosslinked peptide-nucleic acid hybrids isolated from a photochemically crosslinked protein-nucleic acid complex demonstrate that the new protocol can be used to identify nucleic acid binding domains in proteins.

Characterization of Photochemically Cross-linked Protein-Nucleic Acid Complexes by Mass Spectrometry

by

Ole Nørregaard Jensen

A THESIS
submitted to
Oregon State University

in partial fulfilment of
the requirements for the
degree of

Doctor of Philosophy

Completed October 20, 1994
Commencement June 1995

Doctor of Philosophy thesis of Ole Nørregaard Jensen presented on October 20, 1994

APPROVED:

Redacted for Privacy

Major Professor, representing Biochemistry & Biophysics

Redacted for Privacy

Acting Head 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 Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Ole Nørregaard Jensen, Author

To Vibeke

Acknowledgment

I would like to thank the following colleagues for providing UV-crosslinked protein-nucleic acid complexes:

Dr. Mark C. Young and Professor Peter H. von Hippel (University of Oregon)
Mr. Stephen Swenson and Professor Steven Seifried (University of Hawaii)
Mr. Samuel Bennett and Professor Dale Mosbaugh (Oregon State University)

I would also like to thank

Professor Jane Aldrich and colleagues (Oregon State University) for helping with peptide synthesis.
Mr. Dallas Connor and Professor Martin Shetlar (University of California, San Francisco) for sharing preliminary data on UV-crosslinking of protein-nucleic acid complexes.

Thanks to everyone in the OSU mass spectrometry laboratory, especially

Elisabeth 'Lilo' Barofsky, for excellent assistance with MALDI mass spectrometry
Donald A. Griffin, for assistance with electrospray mass spectrometry.

A special thanks to

Professor Douglas F. Barofsky, for his encouragement and advise throughout this study,

and to my beloved wife and best friend,

Vibeke.

I thank the following institutions for financial support during my study at Oregon State University:

Beckett Foundation, Copenhagen, Denmark
Carlsberg Foundation, Copenhagen, Denmark
Danish Research Academy, Aarhus, Denmark
Fulbright Commission/Institute of International Education, San Francisco, USA

Support for this project was also provided by:

Medical Research Foundation of Oregon
OSU Research Council (NIH Biomedical Research Support Grant RR07079)
The Environmental Health Sciences Center (OSU)

Table of Contents

List of Figures

List of figures, Continued

List of Tables

List of Appendix Figures

List of Appendix Tables

List of Abbreviations

bp	base pair (in DNA)
DIPEA	Diisopropylethylamine
DNA	Deoxyribonucleic acid
dT	Deoxyribothymidylic acid
E. coli	Escherichia coli
ESI	Electrospray ionization
FSC	Fused silica capillary
gp32	gene 32 protein from phage T4 (DNA helix destabilizing protein)
HOBt	Hydroxybenzotriazole
HPLC	High performance liquid chromatography
MALDI	Matrix-assisted laser desorption/ionization
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NTP	Nucleoside triphosphate
PyBOP	Benzotriazolyloxy-tris(pyrrolidino]-phosphonium-hexafluorophosphate
Rho	Transcription termination factor Rho (*E. coli*)
RNA	Ribonucleic acid
4sU	4-thio-uracil
TIC	Total ion current
TOF	Time-of-flight
Ung	Uracil-DNA glycosylase (*E. coli*)
UV	Ultraviolet

Nomenclature

Photochemically crosslinked protein-nucleic acid complexes are denoted using the symbol "×" between the crosslinked species, i.e. protein × nucleic acid.

Preface

What started out as a 6-page research proposal for the BB655 *DNA-protein interactions* class, taught by Professors Chris Mathews and Ken Van Holde, now has turned into the present thesis, several grant proposals, and several publications. This is partly due to the initial encouragement to "check out the idea" from Professor Van Holde and to the guidance and support from my advisor, Professor Douglas F. Barofsky.

Characterization of Photochemically Cross-linked Protein-Nucleic Acid Complexes by Mass Spectrometry

Chapter 1: Introduction

Pulchra sunt quae videntur,
Pulchroria quare scientur,
Longe pulcherrima quea ignorantur

Beautiful are the things we see,
More beautiful those we understand,
Much the most beautiful those we do not comprehend

Niels Steensen (Steno)

Ever since it was realized that deoxyribonucleic acid (DNA) is the carrier of genetic information, researchers have been trying to elucidate the mechanisms by which the linear DNA code is interpreted and transformed into the awesome complexity of the living cell. Vital processes, such as cell division and cell differentiation, are orchestrated and controlled in large parts by the interaction of proteins with nucleic acids. Some proteins bind to DNA to regulate the expression of cell specific genes. Enzymes bind to DNA to replicate the genome in preparation for cell division. Other proteins bind to DNA to arrange the linear chain into the tightly packed structures of chromatin and chromosomes. Genes are transcribed into ribonucleic acid (RNA) by means of enzymes that read off the DNA-sequence and produces a complementary RNA strand. This RNA is subsequently processed and translated into a protein gene-product by the ribosome – a large protein-nucleic acid complex. Even though the processes of replication, transcription, and translation are fairly well understood, we still have a long way to go to get a detailed picture of the molecular features of these processes.

How do proteins and nucleic acids recognize each other in an often highly specific

manner? What are the structural features of the protein that make it match up with a particular nucleic acid substrate? Which amino acids are lining the nucleic acid binding site of nucleic acid binding proteins? The scope of the present work is to develop a novel analytical approach that might help answer these questions.

The new experimental strategy is based on using photochemical crosslinking techniques in combination with mass spectrometric methods to characterize protein-nucleic acid contacts. As will be explained in detail in the next chapters, irradiation of protein-nucleic acid complexes by ultraviolet light induces the formation of a covalent photochemical bond between the two species. The bond is formed between an amino acid in the nucleic acid binding site of the protein and a nucleic acid base that is in close proximity. In effect, the nucleic acid binding site of the protein has been tagged by the nucleic acid substrate, causing an increase in molecular weight of the protein. Identification of the protein domain or the actual amino acids that are tagged will therefore localize the nucleic acid binding domain of the protein.

Recent developments in the field of mass spectrometry have made sensitive analysis of peptides, proteins, and nucleic acids possible. Characterization of intact, modified proteins is possible, because a modification causes a mass increase relative to the unmodified protein. Peptide mixture analysis (peptide mapping) by mass spectrometry is a powerful tool for monitoring proteolytic digests of proteins and for localizing modified stretches of amino acids.

Given these premises it seems obvious to attempt to combine UV-crosslinking techniques and mass spectrometry into a new sensitive analytical approach for investigating protein-nucleic acid interactions, specifically for identifying contact points at the protein-nucleic acid interface.

Improving the speed and sensitivity by which information regarding protein-nucleic acid contacts can be obtained, should allow experiments to be performed in cases where limited sample is available and it should allow many new experiments to be performed in cases where sample amounts is not a limiting factor. Knowing the nucleic acid binding domain of a protein is important for model building and for design of other experiments, such as site-directed mutagenesis, binding studies, and conformational studies, to further characterize the nucleic acid binding features of the protein.

Furthermore, the devised strategy should, in the long term, allow analysis of interactions between multiprotein complexes and nucleic acids – a class of molecular aggregates where structural information on the molecular level is hard to obtain by other means. The emerging role and importance of multiprotein complexes in cell biology make this outlook particularly interesting.

The aim of the present study is to demonstrate the feasibility of a mass spectrometry based protocol for localizing contact points between proteins and nucleic acids after photochemical crosslinking. This work also identifies the difficulties and problems associated with

the proposed strategy and lays a foundation for future lines of research.

In the present work we have demonstrated for the first time that intact, UV-crosslinked protein–nucleic acid complexes can be analyzed by mass spectrometry. The molecular weight information provided by the analyses directly demonstrates that the expected crosslinked species were formed. The purity and stoichiometry of a purified complex was ascertained by mass spectrometric analysis.

Direct mass spectrometric mapping of peptide mixtures from a purified UV-crosslinked protein–oligonucleotide complex was attempted but was not successful. However, we have shown that peptide-nucleic acid hybrids isolated from a UV-crosslinked protein–nucleic acid complex, after proteolytic digestion, can be analyzed by mass spectrometry. Accurate molecular weight determination of a mixture of crosslinked peptide–nucleic acid hybrids allowed identification of the tryptic peptides that were crosslinked to a DNA oligomer. This in turn located amino acid stretches of the protein that are involved in DNA binding and for the first time demonstrates the utility of the proposed strategy using a biologically relevant system.

In order to optimize experimental conditions for mass spectrometric analysis of covalent peptide-oligonucleotide conjugates we have synthesized larger amounts of defined sequence peptide-linker-oligo(dT) hybrids by chemical methods. The results of mass spectrometric studies of these species indicate that molecular weight information of low picomole amounts of purified peptide-oligonucleotide hybrids can be obtained. Tandem mass spectrometry of these species produced useful structural information regarding the nucleotide sequence but no amino acid sequence information.

The results obtained using photoaffinity-labeled Rho demonstrate how powerful the hyphenated method of liquid chromatography-mass spectrometry (LC-MS) is for peptide mapping. However, peptide-nucleic acid complexes could not be unambigously identified, probably due to the unknown nature of the photo-adduct or due to instability of the photoaffinitylabel.

We feel confident that photochemical crosslinking techniques in combination with mass spectrometric methods is a feasible approach for characterization of protein-nucleic acid interactions. However, further optimization of several analytical steps of the protocol is necessary to increase sensitivity and specificity for identification of amino acid-nucleic acid contact points.

Some of the data obtained during this study have been published (1, 2, 3, 4).

Chapter 2: Background

Protein-nucleic acid interactions

A central goal of molecular biology is to elucidate the molecular interactions between proteins and nucleic acids. Full knowledge of the structural details, as well as of the nature and magnitude of the driving forces, of protein-nucleic acid interactions is considered essential to understanding the basic cellular processes involved in packaging, replication, recombination, dynamic alteration and repair of DNA, transcription and processing of RNA, synthesis of proteins, and regulation of enzyme activities. Some examples of processes that involve protein-nucleic acid interactions are listed in table I.

Table I. Examples of protein-nucleic acid interactions and the processes they orchestrate and control.

Protein-DNA interactions	Replication/repair
	Transcription
	Gene regulation
	Chromosome structure
Protein-RNA interactions	Transcription
	RNA-processing
	RNA-transport
	Translation
Protein-NTP interactions	Signal transduction
	Enzyme regulation
	Metabolic control

For example, RNA transcription involves the binding of initiation and transcriptor factors (enhancers, activators) to specific DNA regions near a promoter region. This in turn allows RNA polymerase to bind to the promoter and initiate transcription. An understanding of the structural features that govern the specificity of the above interactions is needed because many of the processes where protein-nucleic acid interactions are involved are critical to the survival of a cell, an organ, or an organism.

A cell contains millions of basepairs of DNA. The integrity of DNA is maintained by proteins that can bind more or less specifically to DNA strands. Structural proteins, such as the histones, bind non-specifically to DNA, that is, the proteins do not require a particular DNA-sequence to be present in order to bind (however, recent data indicate that certain structural features of DNA, such as bends or kinks, might be necessary for histones to bind). Regulatory proteins, such as transcription factors, bind to specific DNA sequences, usually 6-12 bp long, that are present in only a few copies in the whole genome.

How do proteins bind to DNA? What are the features of the protein-nucleic acid interfaces that allows either specific or non-specific binding to take place? How do sequence-specific DNA binding proteins discriminate between "regular" DNA and a specific DNA sequence motif? Although full answers to these questions are not known, enough structural data on DNA and sequence-specific DNA-binding proteins is available to allow some conclusions to be drawn as discussed below. There is virtually no detailed structural information available for non-specific protein-nucleic acid interactions.

Let us first take a look at the structure of DNA. The DNA double helix contains a major and a minor groove. In B-DNA, the *in vivo* structure of DNA, the major groove is big enough to allow a protruding part of a protein to 'probe' nucleic acid bases within the groove. In some cases, DNA conformation allows nucleobases in the minor groove to be probed by amino acids. Formation of two hydrogen bonds between a DNA basepair and an amino acid side chain of the protein is enough to provide sequence discrimination and specificity (5). Thus, sequence specific binding can be obtained by forming hydrogen bonds between donor/acceptor groups on amino acid sidechains of the nucleic acid binding site of the protein and complementary acceptor/donor groups in the DNA (6). Sequence independent hydrogen bonds can form between the DNA phosphate backbone and amino acid residues that are positioned properly relative to DNA. The latter interactions further stabilize protein-DNA complexes.

Other types of interactions are also in effect, for example, electrostatic interactions between the phosphate backbone of DNA with positively charged amino acid sidechains of the protein. Electrostatic interactions to the phosphate backbone are obviously less specific than interactions via hydrogen bonds to the nucleobases. Electrostatic interactions are believed to be

involved in histone-DNA interactions, since histone proteins are highly basic, and, as mentioned above, bind nonspecifically to DNA.

Interchelation of aromatic sidechains of amino acids with nucleobases in single-stranded DNA has been proposed as a means for (non-specific) binding of single-stranded DNA binding proteins (7, 8).

It has been realized that the conformations of both DNA and protein change upon binding of the protein to DNA. The conformational change is brought about to protect non-polar groups at the protein-nucleic acid interface from the surrounding solvent. This thermodynamically favorable process serves to further stabilize protein-DNA complexes (9, 10). The conformational change might also serve to 'report' that the protein is bound, thereby serving as a physiological signal for other proteins. A dramatic example of conformational change of DNA induced by a protein is the structure of the TATA-binding protein with DNA. The protein induces a 90 degree kink in the DNA that presumably allows other transcription factors and RNA polymerase to bind (11).

Let us look in more detail at the known features of some sequence-specific DNA binding proteins to get a better understanding of the structural features needed to direct a protein to bind to DNA.

STRUCTURAL MOTIFS FOR BINDING DNA

Over the last 6-7 years the structures of several DNA-binding proteins have been determined by X-ray crystallography and NMR spectroscopy. Identical structural motifs have reoccurred in certain classes of proteins (12, 13, 14). These include helix-turn-helix, homeodomain, β-ribbon, zinc-finger, leucine zipper, and helix-loop-helix motifs.

In many procaryotic repressors/activators (for example, cro, λ, 434, P22, Trp, CAP) a helix-turn-helix (HTH) motif is responsible for recognition of the DNA target site (6, 15, 16, 17, 18). This motif consists of two α-helices, helix 1 and 2, which are positioned at an angle of 120 degrees relative to each other. The second α-helix of the motif, also known as the recognition-helix, is responsible for forming specific hydrogen bonds with nucleobases in the major groove of the DNA. Sequence specific DNA-binding is accomplished by binding of a protein dimer to the palindromic target sequence (each protein monomer provides one helix-turn-helix motif) (Figure 1). Alignment and anchoring of the recognition-helix in the major groove is accomplished mainly by (non-specific) hydrogen bonds between the DNA phosphate backbone and the protein. The fact that the protein binds as a dimer illustrates how simple complementarity to the palindromic DNA sequence can be obtained.

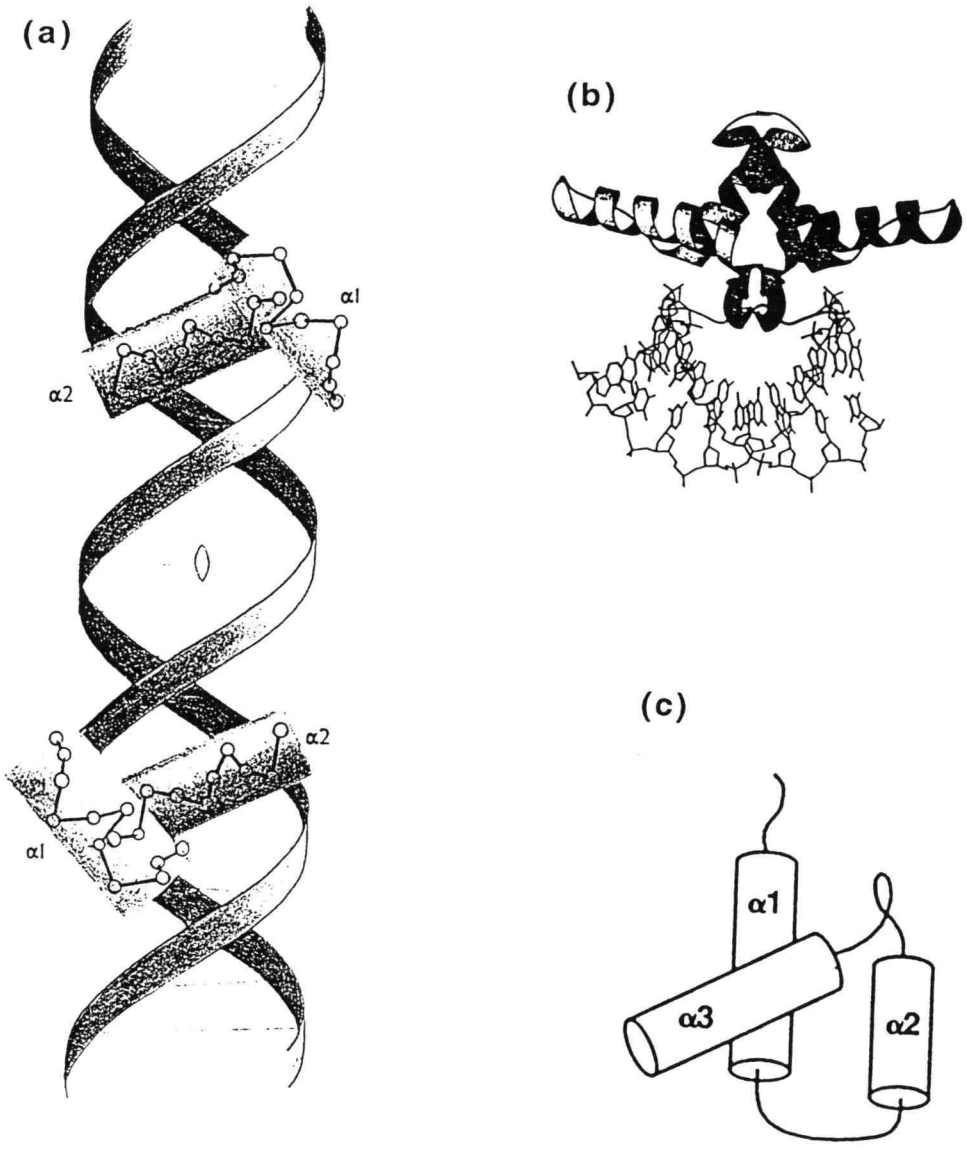


Figure 1. Three dimensional structures of DNA binding motifs.
The λ cro repressor homodimer binds to DNA using the HTH motif (a). The MetJ repressor homodimer beta-ribbon motif bound to the major groove in DNA (b). The monomeric eucaryotic homeodomain motif (c) binds to DNA using an α-helix for recognition. Adapted from refs. (17) and (19).

The HTH motif is not the only DNA binding motif in procaryotes. For example, the *E. coli* MetJ repressor and the phage P22 arc repressor employ an antiparallel β-ribbon motif to recognize their DNA targets (13, 19). The antiparallel β-sheet fits into the major groove of DNA (Figure 1b).

Eucaryotes use an α-helix structure similar to the procaryotic HTH motif, called a homeodomain (13, 20, 21) (Figure 1c). The difference between the homeodomain and HTH motif is that transcription factors containing the former motif bind to DNA as monomers and that the alignment of the homeodomain-helices is different from the alignment of the HTH motif relative to the DNA (13).

Another eucaryotic motif, the zinc-finger, has been found in many transcription factors (13, 22, 23). This structure is stabilized by complexation with a Zn^{2+} ion. The zinc-finger family can be divided into three subgroups. The first one (class 1) is a 30 residue stretch of amino acids with one Zn^{2+} ion coordinated by 2 His and 2 Cys residues (Figure 2a). The structure consists of an antiparallel β-sheet (hairpin) followed by a 12-residue α-helix that is packed against the β-sheet. One Cys present in each β-strand and two His residues on the inward-facing side of the α-helix stabilize the zinc-finger structure by binding themselves to a Zn^{2+} ion. This is the type of zinc finger originally found in TFIIIA from *Xenopus laevis*.

The second zinc finger motif (class 2, Figure 2b) is a 70 amino acid stretch, found in some steroid receptors, that contains two Zn^{2+} ions, each coordinated by 4 Cys residues. The third class contains two closely spaced Zn^{2+} ions that share 6 Cys residues and is found in some yeast activators, for example, GAL4 (Figure 2c).

Zinc-finger proteins usually contain repeats of three or more of these motifs. It was recently shown that a triplet zinc-finger domain (Zip268, class 1) binds in the major groove of DNA and partly wraps around the DNA (24) as had been predicted (25). Each finger is in contact with nucleobases and the phosphate backbone, and each consecutive finger/DNA contact is spaced 3 bp apart. Mainly the N-terminal residues of the α-helix form hydrogen bonds to the nucleobases in the major groove. More recently, a 5-finger structure (class 1) complexed with DNA was solved (26). In this case, only fingers 2-4 were in contact with DNA while finger 1 interacted mainly with finger 2. Fingers 4 and 5 contacted nucleobases while fingers 2 and 3 mainly interacted with the phosphate backbone. Much remains to be said about zinc-fingers.

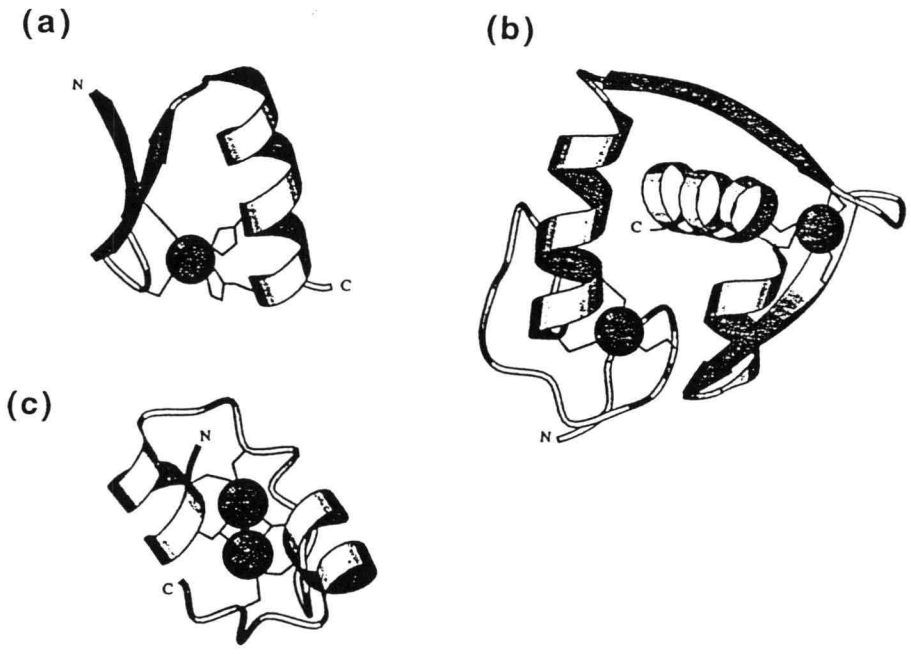


Figure 2. Zinc finger DNA binding motifs.
Several zinc-binding motifs exists. The Class 1 Zif268 zinc finger is shown in (a), the Class 2 glucocorticoid receptor zinc-finger is shown in (b), and the Class 3 GAL4 Zn_2Cys_6 zinc finger is shown in (c). Adapted from ref. (23).

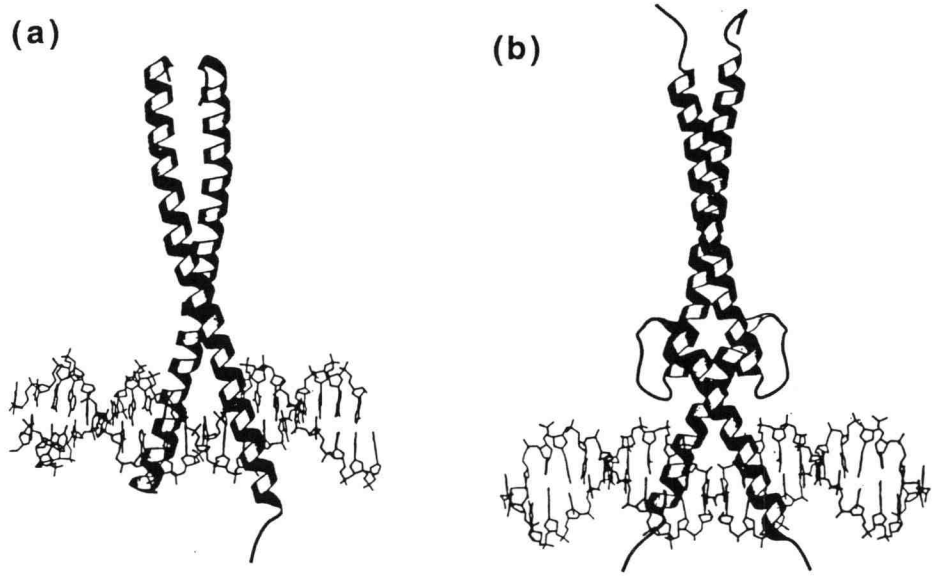


Figure 3. Basic coiled coil (bCC) and basic helix-loop-helix (bHLH) structures.
The GCN4 bZIP domain bound to DNA is shown in (a) and the Max bHLH domain is shown in (b). In both cases, binding to the DNA major groove is accomplished by α-helices. Protein-protein interactions (dimerization) are facilitated by hydrophobic interactions between α-helices (leucine zippers). Adapted from ref. (27).

The DNA binding domain of the leucine zipper structure, also known as the basic coiled coil (bCC) motif, consists of a leucine zipper domain that promotes dimerization, and a basic domain that binds to DNA (13, 27). The leucine zipper motif is an α-helical, heptad repeat of leucine residues over a stretch of 30-40 amino acids. It was proposed and recently demonstrated by using the DNA binding domain of the yeast transcription factor GCN4 (28) that leucine zipper regions can dimerize by forming an α-helical coiled coil of 2 parallel α-helices (Figure 3a). Dimerization induces a structural transition of the basic domain into an α-helical structure that allows DNA binding. Presumably, the dimer is Y shaped, with the coiled coil forming the stem of the Y. The model suggests that the forks/branches bind to a palindromic sequence on the DNA in a scissor-grip fashion.

The basic helix-loop-helix (bHLH) motif is similar to the bCC motif in that the structure is predominantly α-helical and a flexible α-helical region is used for DNA binding. However, dimerization is accomplished by hydrophobic interactions between loops and α-helices, forming a left-handed four-helix bundle (27) (Figure 3b).

Why is it that an α-helix is used so often in DNA-binding motifs? It was realized early on that an α-helix can fit into the major groove of B-DNA. Furthermore, an α-helix is very stable, a requirement for a structure that is exposed on the surface of a protein. In all the above cases, the α-helix is packed tightly against a supporting protein structure that helps present and align the recognition α-helix relative to the DNA. An overall picture emerges, in which there are at least three contributions to sequence-specific DNA recognition by proteins (12, 29): First, the supporting protein structure determines how the recognition helix is presented to DNA; Second, the orientation of the recognition α-helix determines which amino acids can interact with DNA; Third, specific amino acids that are exposed by the helix can interact with nucleobases in the DNA. These contributions allows a wide variety of DNA sites to be recognized.

Another feature of the above mentioned motifs is that they are usually found in several copies, either as protein-dimers (HTH, class 2 and class 3 zinc-fingers, bCC, bHLH) or as covalent repeats (class 1 zinc-fingers). The presence of multiple contact points increases the stability of the protein-DNA complexes and allows for an even larger variety of recognition motifs.

That nature cannot be expected to produce easily classified DNA binding proteins can be illustrated by the yeast transcriptional activator GAL4 (30). It binds DNA by means of a class 3 zinc-finger domain and dimerizes via a leucine zipper domain.

Other less easily classifiable binding domains exist. For example, interchelation of nucleic acid bases with aromatic amino acids has been suggested to play a role in single-stranded DNA binding proteins as mentioned above.

Which amino acids recognize which nucleobases – is there a one-to-one correspondance between certain amino acids and certain nucleobases? Unfortunately, no patterns have emerged

from the protein-DNA complexes for which structures are available, although a slight preference for interaction with purines (especially guanine) might be apparant (13). In all the known structures, half of the protein-nucleic acid contacts involve hydrogen bonds between the protein and the nucleobases, while most of the remaining contacts involve hydrogen bond interactions between the protein and the phosphate backbone of DNA (13).

OTHER NUCLEIC ACID BINDING MOTIFS.

Binding of mononucleotides seems to be accomplished in a similar way by a variety of proteins. By nucleotide we here mean NAD, NADP, FAD, NTP, NDP, NMP, and FMN. The β–α–β mononucleotide binding domain (the Rossmann fold) is found in many dehydrogenases, where it is a subset of a larger structure of (at least) three parallel β-sheets with two intervening α-helices (17). Both the β-sheets and the α-helices are involved in nucleotide binding. An amino acid sequence motif has been found in these mononucleotide binding proteins (31, 32). The motif varies slightly between protein-families but is usually built around the sequence GxxxxG and spans over 10-30 residues. Protein families that contain this motif are kinases, elongation factors, ras-proteins, myosin heavy chain, and ATP synthase (32).

METHODS TO STUDY PROTEIN-NUCLEIC ACID INTERACTIONS

What do we want to know about a nucleic acid binding protein? If it is a DNA binding protein we would like to know if it binds specifically or non-specifically. If it binds specifically, the localization, size, and sequence of its DNA binding motif should be investigated. Whether the binding be specific or non-specific, we would want to find out the affinity of the protein towards nucleic acid. Finally, we would want to know the structure of the protein, especially the structure of the nucleic acid binding domain of the protein and how it interacts with the nucleic acid substrate.

Many methods are available that will answer one or the other of these questions. Some of these methods are described briefly below.

The character and level of interaction between proteins and nucleic acids can be effectively revealed through indirect methods, such as gel retardation, filter binding, sedimentation equilibrium, and footprinting (33). These techniques do not expose any of the structural features of the interaction, but can provide qualitative and quantitative information on binding parameters, such as binding constants, binding site size, and binding site sequence. The amino acid-base pair contacts in protein-nucleic acid complexes can be indirectly identified by

site-directed mutagenesis (34): The influence of amino acid substitutions on nucleic acid binding is investigated on a trial and error basis. This is a widely used method for studying the effects of amino acid substitutions on substrate-binding. However, the influence of amino acid substitutions on the overall protein structure is hard to predict, and therefore, care must be taken when interpreting the results of this kind of experiment.

Direct methods for determining the topology of protein-nucleic acid interactions are sparse. X-ray crystallography and, more recently, multidimensional nuclear magnetic resonance (NMR) spectroscopy have yielded 3-dimensional, structural information on protein-nucleic acid complexes (6). These techniques, which reflect a static state of a dynamic complex, require large, homogeneous amounts of sample and are tedious to perform. However, persistent and skillful use of these techniques have provided us with the protein structural features for DNA binding proteins described in the preceding sections.

Visualization via electron microscopy has added qualitative, topological insights to many facets of protein-DNA interactions discovered by other biochemical and biophysical techniques (35, 36).

The emerging importance of 'molecular crowding' in many metabolic processes (37) warrants the development of new analytical tools for investigation of multiprotein-nucleic acid complexes. One example is the enzyme complex for nucleotide synthesis, which is now believed to be working as an assembly line, channeling nucleic acid constituents directly to DNA polymerase. Another example is the interaction between transcription factors, RNA polymerase, and DNA. Nucleosomes and ribosomes are other examples of large protein-nucleic acid complexes. Due to their large size these complexes are not easily amendable to analysis by NMR spectroscopy and X-ray crystallography. Genetic analyses and microscopic techniques (35) can be used to generate models of these assemblies. However, molecular detail of protein-nucleic interactions in such complexes is hard to obtain.

There is a need for methods that allow non-specific interactions to be investigated in molecular detail. X-ray crystallography and NMR need homogenous samples in order to produce structural data. These techniques are therefore not easily amendable to non-specific interactions where the protein can move relative to the DNA oligomer, supposedly because there is no "best" binding site, therefore creating a heterogenous mixture of complexes. However, certain DNA conformations might allow some non-specific protein-DNA complexes to be more stable than others. An alternative approach to study non-specific protein-DNA interactions is to "freeze" the complexes by crosslinking by irradiation with ultraviolet light, thereby producing a covalent complex that is amendable to structural analysis by protein chemical techniques. These photochemical crosslinking techniques are described in the following section.

Photochemical cross-linking of nucleic acids to proteins

In the following sections I will describe the concepts and techniques that establish the basis for developing a protocol for mapping nucleic acid binding sites in proteins by photochemical crosslinking and mass spectrometry. In the next chapter I will bring these concepts together into an overall analytical strategy.

It was discovered already in 1962 that ultraviolet light induces the formation of cross-links between DNA and protein (38, 39). Later on, it was proposed and demonstrated that UV induced crosslinking only takes place between nucleotide and amino acid residues that are in close contact during UV irradiation (40, 41). Before describing in too much detail how photochemical crosslinking can be used as a tool to investigate protein-nucleic acid interaction, I will review some of the photochemical properties of nucleic acids and amino acids.

PHOTOCHEMISTRY OF NUCLEIC ACIDS AND AMINO ACIDS

The heterocyclic bases of nucleic acids absorb light primarily at 260 nm wavelength. Aromatic amino acids absorb light primarily at 280 nm. Proteins contain only 5-10 percent of aromatic amino acids on average. This allows rather large doses of 250-270 nm UV-irradiation to be used for photochemical crosslinking of nucleic acids to proteins without significantly damaging the protein since most of the light is absorbed by the nucleic acids. Some nucleic acid analogs allow irradiation at wavelength above 300 nm to be used for crosslinking thereby minimizing photoinduced damage to both protein and nucleic acids.

Upon absorption of a photon the nucleic acid base is excited from the ground state S_0 to the singlet state S_1. The lifetime of this excited state is less than 100 ns (42). The molecule can then undergo a photochemical transformation to the ground state or a transition into the first triplet state T_1. From the triplet state the molecule can undergo a photochemical transformation or a transfer into the ground state. The triplet state can exist for up to 1 μs. During irradiation with a relatively low powered UV lamp ($<10^8$ W/m^2) only the S_1 and the T_1 states are populated. By irradiation with laser light at higher fluences ($>10^9$ W/m^2) the S_1 and T_1 states can absorb another photon and get excited to higher excited states (H_S and H_T respectively) with very short lifetimes (<100 fs).

As stated above, the lifetimes of the reactive states S_1 and T_1 are less than a microsecond. This is less than the timescale for macro-molecular motion (milliseconds to

seconds) and is the reason why pulsed UV-laser irradiation can be used to specifically "freeze" otherwise dynamic protein-nucleic acid complexes, that is, without creating non-specific complexes. However, the H_S and the H_T states can be highly energetic, in some cases exceeding the ionization energy of the nucleic acid base in solution, and therefore might create longer lived ions or radical species that can react non-specifically with the protein. Therefore, care must be taken to optimize and characterize the laser irradiation conditions for a given system.

The short lifetime of the reactive nucleic acid species and the formation of a zero-length bond require the nucleic acid to be in close contact with the protein. In other words, it has to be bound in the nucleic acid binding site of the protein. The efficiency of the crosslinking reaction is characterized by the quantum yield: The ratio of reacted molecules to the total number of excited molecules. This parameter is dependent on several factors (43): i) the excitation efficiency of the base; ii) the lifetime of the excited state; iii) the reactivity of the excited base species; iv) the reactivity of the amino acid residue in closest proximity; v) the relative positions and the mobilities of the components to be crosslinked. It follows that the crosslinking yield is therefore dependent on the conformation and composition of the protein-nucleic acid complex and on the solution conditions.

What is the chemical mechanism by which nucleic acids crosslink to amino acids? Due to the low crosslinking yields and, in some cases, heat and acid lability of the crosslinked products, it is a difficult analytical task to characterize crosslinking mixtures. However, model studies using purine and pyrimidine bases and individual amino acids or their analogs have provided some information (44, 45). Pyrimidine bases, especially thymine, are the best crosslinkers. The crosslink is formed in the C5 or C6 position of pyrimidines presumably via a free radical mechanism. A pyrimidyl radical is generated by photoexcitation of the nucleobase followed by hydrogen abstraction from a nearby amino acid residue. The pyrimidyl radical then recombines with the amino acid radical to form a covalent bond (44).

$$P + h\nu \longrightarrow P^*$$

$$P^* + RH \longrightarrow PH\cdot + R\cdot \longrightarrow PRH$$

Shetlar and coworkers have shown that all amino acids have the potential to crosslink to DNA, with Cys, Lys, Phe, Trp, and Tyr being the most reactive (46). More recently, Shetlar and collegues have characterized UV-crosslinked complexes of thymine-(*N*-acetyltyrosine) and thymine-peptide (angiotensin I) using HPLC, NMR, and mass spectrometry techniques. They demonstrated, for the first time, the formation of a thymine-peptide crosslink between the thymine C6 and the ortho position of a Tyr residue in the peptide (47). In this case, they propose a reaction mechanism that involves photoionization of tyrosine followed by deprotonation to form a

radical cation (Figure 4). The ejected electron is captured by the nearby thymine base yielding a radical anion, followed by protonation at C6 to produce the C5-yl radical. Combination of the radicals would produce the thymine-tyrosine adduct.

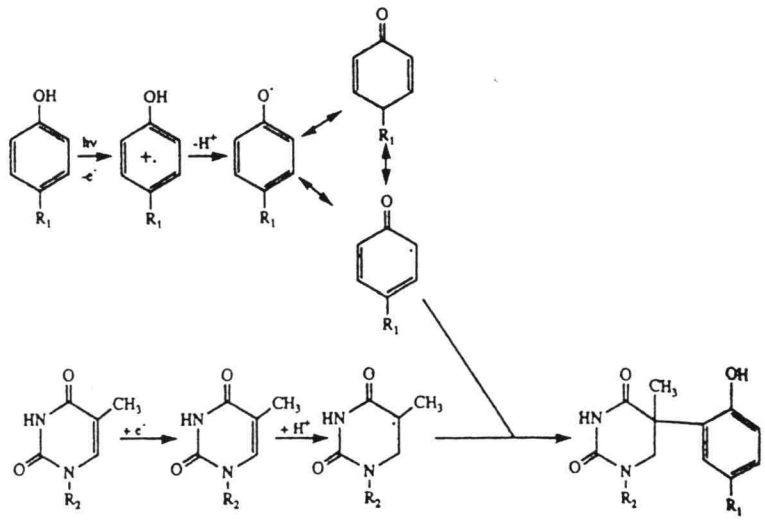


Figure 4. Proposed mechanism by Shetlar et al. for photochemical crosslinking of thymine to N-acetyltyrosine.
Photoionization of Tyr yields the radical cation, which rapidly deprotonates to form the oxygen-centered radical. The ejected electron is captured by thymine to yield a radical anion, which protonates at C6 to form the C5-yl radical. Combination of the two radicals yield the thymine-tyrosine adduct. Adapted from ref. (47).

To date, there have been no reports on the detailed characterization of photochemical crosslinks formed between proteins and nucleic acids. The identification of amino acids that are crosslinked to poly(dT) has been accomplished by looking for a sequence gap during Edman degradation of crosslinked peptides (48) as described below. This approach identifies the crosslinked amino acid, but it does not provide any information on the type of linkage between the nucleic acid end the amino acid residue.

UV CROSS-LINKING OF NUCLEIC ACIDS TO PROTEINS

In order to understand the mechanisms of UV-induced formation of covalent bonds between nucleic acid and protein, a variety of relatively simple model systems have been studied

(44, 45). Results from these studies indicate that the amino acids all have comparable crosslinking potentials but that the nucleic acids have distinctly different crosslinking potentials. Thymine and uridine respectively are by far the most efficient deoxynucleotide and ribonucleotide crosslinkers (49). Although not fully understood, the UV crosslinking technique has proven useful in exploring both static and dynamic protein-nucleic acid complexes (43, 48, 49). Over the last two decades UV-crosslinking has been widely used to study protein-nucleic acid interactions *in vivo* and *in vitro* (43, 44, 48). Some applications are listed in Table II to illustrate the variety of nucleoprotein complexes that can be studied by this technique.

General requirements for crosslinking (chemical as well as radiation induced) to be biologically relevant are: i) Crosslinking should produce stable products; ii) Crosslinks should be formed only between amino acid residues and nucleobases that are in close proximity to one another; iii) Crosslinks should reflect the native conformation of the protein-nucleic acid complex.

Photochemical crosslinking presumably meets these requirements, e.g. it takes place only between amino acid and nucleic acid residues that are in very close proximity – it maintains the original spacing between nucleic acid and polypeptide. This is where the term "zero-length" cross-linking originates. This high specificity of the crosslink can be used as a tool to explore the molecular details of protein-nucleic acid interactions *in vivo* and *in vitro*. This feature also implies that only specific protein-nucleic acid complexes are crosslinked and sets photochemical crosslinking apart from chemical crosslinking. The latter employes molecules that contain two crosslinkable groups separated by a spacer/linker. In contrast to chemical crosslinking reactions, UV-induced crosslinking depends only weakly on the temperature, pH and ionic strength of the solution, thus allowing independent optimization of the conditions for formation and stability of the nucleoprotein complex.

The utility of the photochemical crosslinking technique has been substantially extended through the use of nanosecond pulses of monochromatic UV light generated by a laser (49) and through the incorporation of photosensitive analogs into the nucleic acid reactants (photoaffinity labeling) (48). These techniques will be described in the following sections.

LASER GENERATED PULSED UV CROSS-LINKING

Pulsed (nanosecond) UV laser irradition has advantages over continuous UV-irradiation for crosslinking of protein-nucleic acid complexes. It allows dynamic aspects of protein-nucleic acid interactions to be studied because the irradiation time (single 5-8 nsec pulse) is less than the time-scale of macromolecular motion (49, 50). Due to the power of modern lasers, the same number of photons can be delivered to the experimental system in a split second, as would

otherwise take hours, days, or even weeks of continuous irradiation. This feature should significantly decrease the amount of non-specific labeling and sideproducts (although little is known about the lifetimes of the photoactivated intermediates).

The pulsed UV crosslinking technique was first demonstrated to be useful for studying a RNA polymerase/DNA complex (51) and has since been applied to studies on nucleosomes (52, 53, 54, 55), ribosomes (56), DNA-hybridization (57), and the phage T4 replication machinery (58, 59, 60), and the interactions between RNA and the *E. coli* rho protein (49). Binding site sizes and binding constants of nucleic acids to proteins can be determined since the amount of crosslinked complex is proportional to the equilibrium concentration of the protein-nucleic acid complex.

PHOTOAFFINITY LABELING

A photoaffinity label is an analog of the normal substrate that can be activated by UV-light. Upon irradiation with UV-light the photoreactive group on the substrate will react to form a covalent bond to the nearest molecule. Photoaffinity labeling can greatly increase the crosslinking efficiency of nucleic acids; consequently, this technique can significantly reduce the photon flux required to achieve crosslinking and concomitantly minimize the occurrence of unwanted photoreactions. Photosensitive nucleic acid analogs used in photoaffinity labeling studies include azido derivatives (48), 4-thiouracil (4sU) (61), 4-thiothymidine (62), and 5-halo-pyrimidines (5-bromo-uracil, 5-iodo-uracil) (63, 64, 65). The use of halo- or thio-substituted pyrimidines allows irradiation at higher wavelength (>300 nm) to be used which in turn decreases irradiation damage to proteins. It was recently demonstrated that the combination of photoaffinity labeling with laser irradiation can produce crosslinking yields of 60-70% relative to bound oligonucleotide due to the high reactivity of the photoactive nucleic acid analog (65).

In the present work, 4-thio-uridine-diphosphate (4sUDP) was used to label the nucleic acid binding sites of transcription termination factor Rho.

Table II. Systems where photochemical crosslinking has been used to study protein-nucleic acid interactions.
The use of laser generated pulsed ultraviolet (pUV) irradiation, continous UV (cUV) irradiation, or photoaffinity-labeling is indicated.

System	Technique	Comment	Ref.
Phage fd gene 5 protein	cUV	DNA Binding site	(66)
Phage T4 gene 32 protein	cUV	DNA Binding site	(67)
Phage T4 gene 32 protein	pUV	DNA binding	(58, 59)
E.coli SSB	cUV	DNA Binding site	(68)
Human A1 hnRNP	cUV	DNA Binding site	(69)
HIV Reverse transcriptase	cUV	Binding site	(70)
RNA pol III	Photoaffinity	DNA contact	(62)
T4 DNA pol + Acc. proteins	cUV	Complex formation	(71)
RNase A	cUV	Binding site	(41)
CRP	cUV	Protein-DNA interaction	(72)
Poly(dA)/poly(dT)	pUV	Hybridization	(57)
Uracil-DNA glycosylase	cUV	DNA Binding site	(3)
Histone	pUV	DNA-protein interaction	(52, 53)
Rat DNA pol β	cUV	DNA Binding site	(73)
RNA pol II	Photoaffinity	RNA binding site	(74)
β-tubulin	cUV	GTP binding site	(75)
Transfer RNA/synthase	cUV	Protein-tRNA contacts	(76)
Ribosome (*E.coli*)	cUV	Assembly	(43, 77)
GCN4 (Yeast)	Photoaffinity	DNA contact	(63)
Myc (AMV)	Photoaffinity	DNA contact	(64)
DNA pol I (*E. coli*)	Photoaffinity	Template-Primer contact	(78)

CHARACTERIZATION OF UV-CROSSLINKED PROTEIN-NUCLEIC ACID COMPLEXES

By labeling nucleotides with [^{32}P], UV-crosslinked protein-nucleic acid complexes can be monitored by gel-electrophoresis and autoradiography, and they can be quantified by gel-scanning. Locating the nucleic acid crosslinking site in a reacted protein relies on prior knowledge of its entire primary structure. The classical procedure fo identifying nucleic acid binding sites after UV-crosslinking consists of digesting the purified, crosslinked complex with trypsin or some other proteolytic enzyme, isolating the [^{32}P]nucleotide-labeled peptide via anion-exchange and/or reversed-phase HPLC, and amino acid (Edman) sequencing the crosslinked peptide (48). A site of crosslinking can usually be inferred from the absence of an identifiable phenylthiohydantoin derivative at a particular position in the peptide sequence. This approach consumes considerable time, labor, and sample, and because it is indirect, the results can be inconclusive. Nevertheless, a wealth of protein-nucleic acid complexes have been characterized by this approach (see Table II).

In a few cases, fast-atom-bombardment (FAB) mass spectrometry has been used to characterize photoaffinity labeled peptides (47, 75, 79) demonstrating the usefulness of mass spectrometry for studying modified peptides. However, rather large amounts of peptides were consumed in these experiments. Modern mass spectrometric ionization techniques are 100-1000 fold more sensitive than FAB for peptide analysis (see below).

How does structural information obtained by UV-crosslinking experiments compare to known three dimensional structures of protein-nucleic acid complexes? Only a few systems have been studied by both UV-crosslinking and X-ray crystallography or NMR .

The first experimental evidence that UV-crosslinking of protein-nucleic acid complexes creates covalent bonds in the nucleic acid binding site of the protein was provided by Havron and Sperling in 1977 using ribonuclease A (RNase A) and pUp (41). The latter is a competitive inhibitor of RNase A enzyme. Residues 80, 81, and 82 of the protein were identified as being UV-crosslinked to pUp. The 3-dimensional structure of RNase A revealed that residues 77-82 form the bottom of the binding site for the pyrimidine ring. Thus, UV-crosslinking specifically labels the residues in the nucleic acid binding site of RNase A.

The DNA binding properties of the phage fd gene 5 protein were studied by NMR (8, 80) , and the DNA-binding site was studied by UV-crosslinking to fd 5 DNA and $(dT)_4$ (66). The putative DNA binding groove of the protein is defined by Tyr^{26}, Tyr^{34}, Phe^{73}, and Tyr^{41} based on the NMR data obtained from the protein in the presence of DNA ligand. The UV-crosslinking experiment identified Cys^{33} as the only crosslinked residue. This residue is very close to the Tyr residues of the putative DNA binding site of the gene 5 protein.

The structure of rat DNA polymerase β, in the absence and presence of DNA and dideoxyCTP, was recently solved (81, 82). Previously, UV-crosslinking of the 8 kD N-terminal fragment of the polymerase to $(dT)_{16}$ indicated that Ser^{30} and His^{34} were involved in DNA binding. This region of the DNA polymerase did not provide good structural information by X-ray diffraction in the absence of DNA indicating that it is flexible. In the presence of DNA, this region changed conformation so as to bring it closer to the single stranded DNA primer. However, no direct contact to DNA was observed and the authors speculate that this is due to the short primer region that was used in the experiment. Although not conclusive, the structural data indicate that the UV-crosslinking experiment produced relevant data.

More detailed information was obtained by crosslinking Myc, a basic helix-loop-helix transcription factor, to DNA by using a consensus oligonucleotide with 5-bromo-uracil incorporated instead of thymine (64). A histidine that is conserved between bHLH proteins was crosslinked to the DNA. The structure of Max, a similar HLH protein, shows that this histidine forms hydrogen bonds to a guanine base positioned next to a thymine residue in the DNA consensus sequence (11).

Although the above mentioned examples indicate that UV-crosslinking indeed takes place at the protein-nucleic acid interface of complexes, more research is needed to demonstrate the relevance of UV-crosslinking and photoaffinity-labeling experiments to nucleoprotein structure under physiological conditions. This can be accomplished by UV-crosslinking of protein-nucleic acid complexes of known structure (assuming that these are correct) and then correlating the data obtained by both approaches. Methods that facilitate fast and sensitive analysis of UV-crosslinked complexes would help in bringing about such information.

Mass spectrometry

Mass spectrometry (MS) refer to the measurement of mass-to-charge (m/z) ratios of gas-phase ions derived from a sample, and it can therefore be used for molecular weight determination of sample constituents. In some cases, structural information can be obtained if the analyte molecule fragments in the ion source or in the mass analyzer. In the last decade, tremendous progress in the application of mass spectrometry to the analysis of larger and and more complex compounds has come about due to the development of new ionization techniques and more accurate and sensitive mass analyzers (83, 84, 85, 86, 87). This evolution of techniques has now made it an almost routine task to analyze peptides, proteins and oligonucleotides by mass spectrometry.

Matrix-assisted laser desorption/ionization (MALDI) (88), which is best combined with time-of-flight mass analysis, and electrospray ionization (ESI) (89), which can be readily coupled on-line to micro-chromatographic (μLC) or electrophoretic systems and is usually (but not necessarily) combined with quadrupole mass filtering, both permit the molecular weights of proteins as large as 40 kD to be determined to within 1 part in 10^4 from sample amounts on the order of 0.1-10 pmol. Both the forms and amounts of samples required by these two techniques are consistent with existing practices in biological research laboratories. In the following I will describe these two new ionization methods as they pertain to the topic of this thesis.

MATRIX-ASSISTED LASER DESORPTION/IONIZATION

By using crystals of an ultraviolet (UV) light absorbing compound as a matrix for sample preparation and a pulsed UV-laser for desorption/ionization, Karas and Hillenkamp discovered that gas-phase ions can be produced from polypeptide and protein samples (88, 90, 91). The new ionization technique was coined *matrix-assisted laser desorption/ionization* (MALDI). MALDI has now matured into a routine MS technique. Sub-picomole levels of peptides (86, 92, 93), proteins (84, 86, 91, 94, 95), and oligonucleotides (96, 97, 98, 99, 100, 101, 102, 103), can be analyzed, either as purified samples or as mixtures. This latter application is known as MALDI peptide mapping when applied to, e.g., a proteolytic digest of a protein (see below). A mass accuracy of 0.01% for peptides (mass < 10,000) and 0.1% for protein molecular weight determination (mass > 20,000) has been demonstrated using time-of-flight (TOF) mass analyzers (104, 105).

The 266 nm line from a Nd:YAG (Neodynium:Yttrium-Aluminum-Garnet) laser was used in combination with a nicotinic acid matrix in the original work by Karas and Hillenkamp. More recently, the use of 337 nm (N_2) and 355 nm (Nd:YAG) UV-lasers have been demonstrated (106), and a wealth of matrix-materials have been discovered (100, 102, 107, 108, 109) (see Table III). During crystallization of the matrix/analyte solution, the analyte has to be incorporated into the crystal (in effect creating a solid solution of analyte molecules in the matrix) for the crystal to produce ions upon UV laser irradiation. As can be seen from Table III the matrix requirements for protein and peptide analysis differ inversely from those required for oligonucleotide analysis. This is an important fact to keep in mind for later discussions of MALDI mass spectrometric analysis of protein-nucleic acid complexes. The matrix has to be tolerant to buffers and solvents usually used for protein work, and it has to be fairly volatile and, at the same time, stable in the high vacuum in the mass spectrometer. The role of the matrix is presumably to absorb the UV light and convert the photon energy, via electronic excitation, into heat that causes a rapid evaporation and ionization of the matrix and the embedded analyte material (110). Physical chemical studies of this phenomena may one day produce a definitive explanation as to why the MALDI process works as well as it does.

Table III. Some commonly used matrices for MALDI mass analysis of proteins, peptides, and oligonucleotides.
The suitability of a given matrix for a given compound class is indicated with + (Not good) to +++++ (Excellent).

Matrix	Peptide and protein	Oligo-nucleotide	Ref.
Sinapinic acid	+++++	+	(107)
4-hydroxy-α-cyano-cinnamic acid	+++++	+	(108)
2,5-dihydroxybenzoic acid	++++	++	(109)
Ferulic acid	+++	+++	(107)
2,4,6-trihydroxyacetophenone	++	+++++	(100)
3-Picolinic acid	+	+++++	(102)

ELECTROSPRAY IONIZATION

In the late sixties, Dole and coworkers characterized the properties of charged particles generated at atmospheric pressure by spraying polymer solutions from a metal needle held at a high electric potential (3-6 kV) and positioned close to a conducting plate at ground potential (111). Dole attempted to determine the mass of the electrosprayed particles by a charge retention technique. The problem with this technique was that the charge state of the particles was an unknown and therefore it was not straightforward to use the technique for molecular weight determination of large molecules. In the mid-eighties, Fenn and coworkers at Yale set out to reevaluate the electrospray technique using a quadropole mass spectrometer to analyze the charged particles generated by electrospray. As mentioned above, a mass spectrometer always combines both mass and charge information into the form of an m/z ratio without distinguishing between the two. Fenn and colleagues soon realized that the electrospray technique is a convenient and simple way of generating ions from polymer solutions (89, 112, 113). Interestingly, the molecular ions that are generated by electrospray ionization (ESI) carries approximately one charge for every one-thousand mass units. For example a molecule of mass 20,000 usually carries between 10 and 22 charges, bringing the m/z ratios into the range of a quadropole mass spectrometer (10-2500 amu).

The rediscovery of electrospray and its use as an ionization technique for mass spectrometry has revolutionized the field of biological mass spectrometry over the last 5-6 years. The use of ESI-MS as a tool in protein chemistry is particularly fascinating. High accuracy molecular weight determination of peptide, protein, and oligonucleotide samples at the low picomole to femtomole level is possible (114). An extremely powerful approach to mixture-analysis is to combine a separation technique with ESI-MS. ESI-MS is routinely used as a sensitive and selective on-line HPLC detector that not only generates an elution profile of the separated molecules but also provides molecular weight information on each eluting compound. This hyphenated mode of operation (liquid chromatography-mass spectrometry, LC-MS) is a powerful and sensitive technique for analyzing proteolytic digest mixtures derived from proteins (LC-MS peptide mapping, see below). LC-MS analysis for selective identification of phosphopeptides and glycopeptides has been reported (115, 116). A more recently developed separation technique, capillary electrophoresis (CE), can also be interfaced to ESI-MS and is a promising technique to further increase the sensitivity for mass spectrometry of biological compounds (117, 118, 119).

MASS SPECTROMETRIC PEPTIDE MAPPING

Peptide mapping refers to the analysis of peptide mixtures subsequent to chemical or enzymatic cleavage of a protein. Mass spectrometric peptide mapping is a proven procedure that couples chemical and enzymatic procedures with mass analysis to obtain information about the contribution of specific stretches of the amino acid backbone to the structure and function of proteins and polypeptides (120). This technique has been used to analyze peptides and proteins for molecular weight, sequence, location of mutated amino acids, location of post-translational modifications, character of cystine bridges, and differences between native and recombinant species (83, 86, 121, 122). MALDI peptide mapping (92, 95) and LC-MS peptide mapping (123, 124, 125, 126, 127) are particularly sensitive tools for protein digests. Modifications of proteins or peptides can be characterized by monitoring the molecular weight increase caused by covalent attachment of the ligand. Localization of phosphorylation sites and glycosylation sites in proteins can now be performed by LC-MS peptide mapping by taking advantage of the unique capabilities of the mass spectrometer for detection of specific reporter ions (115, 116).

TANDEM MASS SPECTROMETRY

Tandem mass spectrometry (MS/MS) (128) is a means of obtaining detailed structural information on molecules. A tandem mass spectrometer is a linear arrangement of two mass spectrometers (MS1 and MS2) separated by a collision cell (Figure 5). The collision cell can be filled with an inert gas, usually Argon or Helium. A molecular ion (parent ion), selected by MS1, is allowed to collide with gas molecules in the collision cell between MS1 and MS2. Collisions will induce thermal excitation of the molecular ion, in turn causing it to fragment. The charged fragments (daughter ions) can be analyzed by scanning MS2 over the mass range of interest.

By using the tandem MS mode, structural (sequence) information on peptides can be obtained (86, 129, 130). It has also been demonstrated that sequence information can be obtained from intact proteins by tandem mass spectrometry (131, 132, 133). The nomenclature for peptide fragment designations is given in Figure 6 (134).

Collisional activation of peptide ions produces mainly fragment ions resulting from cleavage of the peptide amide bond. Charge retention on the N-terminal fragment results in B ion series, while charge retention on the C-terminal fragment results in Y ion series. Tryptic peptides produce particularly nice B and Y ion series, because both the N-terminal and the C-terminal

fragment can retain a charge: The B ions are formed by protonation of the N-terminal amino group and Y ions by protonation of the basic sidechain of the C-terminal residue (lysine or arginine).

Peptides can be sequenced by tandem mass spectrometry either by off-line analysis of a purified peptide, by direct analysis of mixtures (selecting one peptide at a time for sequencing), or by on-line sequencing of peptides as they elute off an HPLC column (LC-MS/MS) (124, 125). The latter procedure is extremely sensitive and powerful as has been demonstrated recently by separation and sequencing of antigenic peptides isolated from major histocompatibility complexes, by capillary HPLC interfaced to ESI-MS/MS (135, 136, 137).

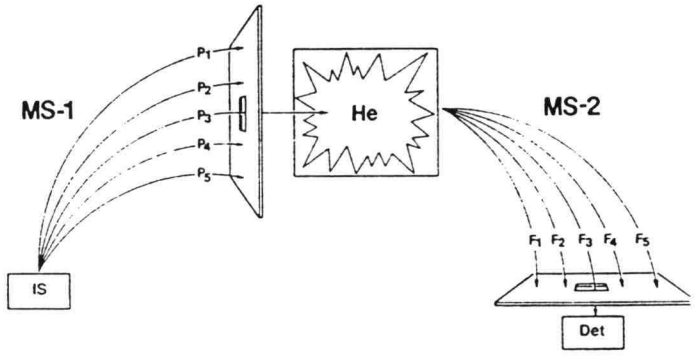


Figure 5. Principle of tandem mass spectrometry.
A parent ion selected in MS1 undergoes collision induced fragmentation in the gas filled region (collision cell) between MS1 and MS2. The resulting fragment (daughter) ions are analyzed by scanning MS2. Adapted from ref (130).

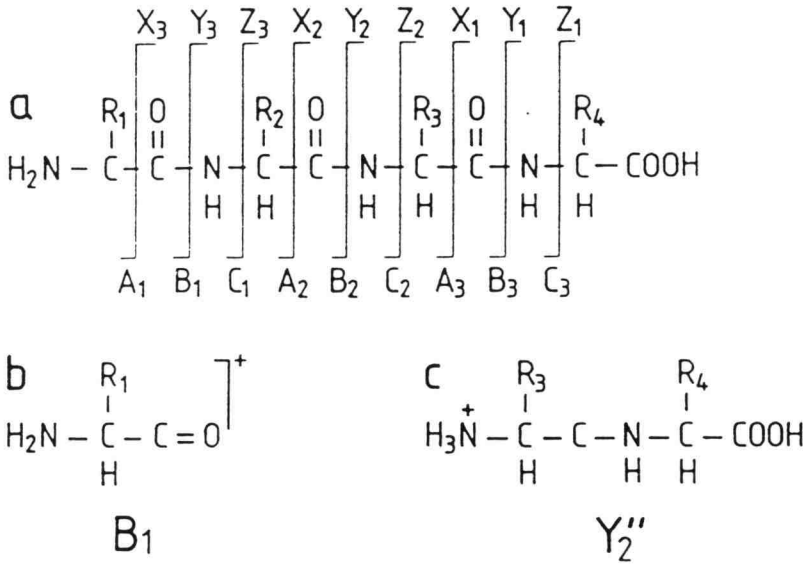


Figure 6. Nomenclature for peptide fragments generated by mass spectrometric fragmentation.
The three possible cleavage points of the peptide backbone are called, A, B, and C, when the charge is retained on the N-terminal fragment, and X, Y, and Z, when the charge is retained on the C-terminal fragment. The most common fragment ions are A, B, and Y. Adapted from ref. (134).

Chapter 3: Strategy and aims

As touched upon in earlier sections, characterization of protein-nucleic acid complexes will provide us with important information that might one day allow us to understand the nature of specific and, maybe more importantly, non-specific protein-nucleic acid interactions in greater detail.

The classical techniques have provided considerable information, but in order to probe more deeply, we feel that new and more sensitive techniques must be developed.

The recent surge in applications of UV-crosslinking and photoaffinity-labeling techniques to study protein-nucleic acid interactions provides a new tool that, in combination with protein chemical techniques, allow contact points between proteins and nucleic acids to be identified under the assumption that a zero-length covalent bond is formed at the protein-nucleic acid interface. This investigation has been carried out on the premise that mass spectrometric characterization of photochemically crosslinked protein-nucleic acid complexes, when fully developed and implemented, will be a fast, sensitive, and selective method for identifying crosslinked amino acid residues in proteins and, presumably, crosslinked nucleic acid base residues in oligonucleotides. Mass spectrometry has recently proven to be more sensitive than Edman degradation for peptide sequencing, and the hyphenated method of LC-MS and LC-MS/MS allows proteolytic digest mixtures to be separated, mass analyzed, and sequenced in considerable less time than by the classical approach. Other advantages of mass spectrometry over Edman degradation are that N-terminally blocked peptides can be sequenced and that post-translational or chemical modifications can be identified.

Our strategy is depicted in Figure 7. In the first stage, the nucleic acid binding protein and its nucleic acid substrate or a photoactivatable analogue thereof are incubated under proper conditions to form the protein-nucleic acid complex. The sample is subsequently irradiated with UV-light for a given period of time in order to create photochemically crosslinked protein-nucleic acid complexes. The irradiated reaction mixture is directly analyzed by MALDI MS to monitor the crosslinking reaction and to assess the identity of the crosslinked product based on accurate molecular weight determination. If the crosslinking yield is low, the crosslinked protein-nucleic acid complex can be separated from the unreacted reagents before further characterization.

In the second stage, the crosslinked complex is fragmented by an enzyme (for example trypsin) or by chemical treatment, and the peptide mixture is analyzed by MALDI and HPLC-ESI MS in order to generate a mass specific peptide map.

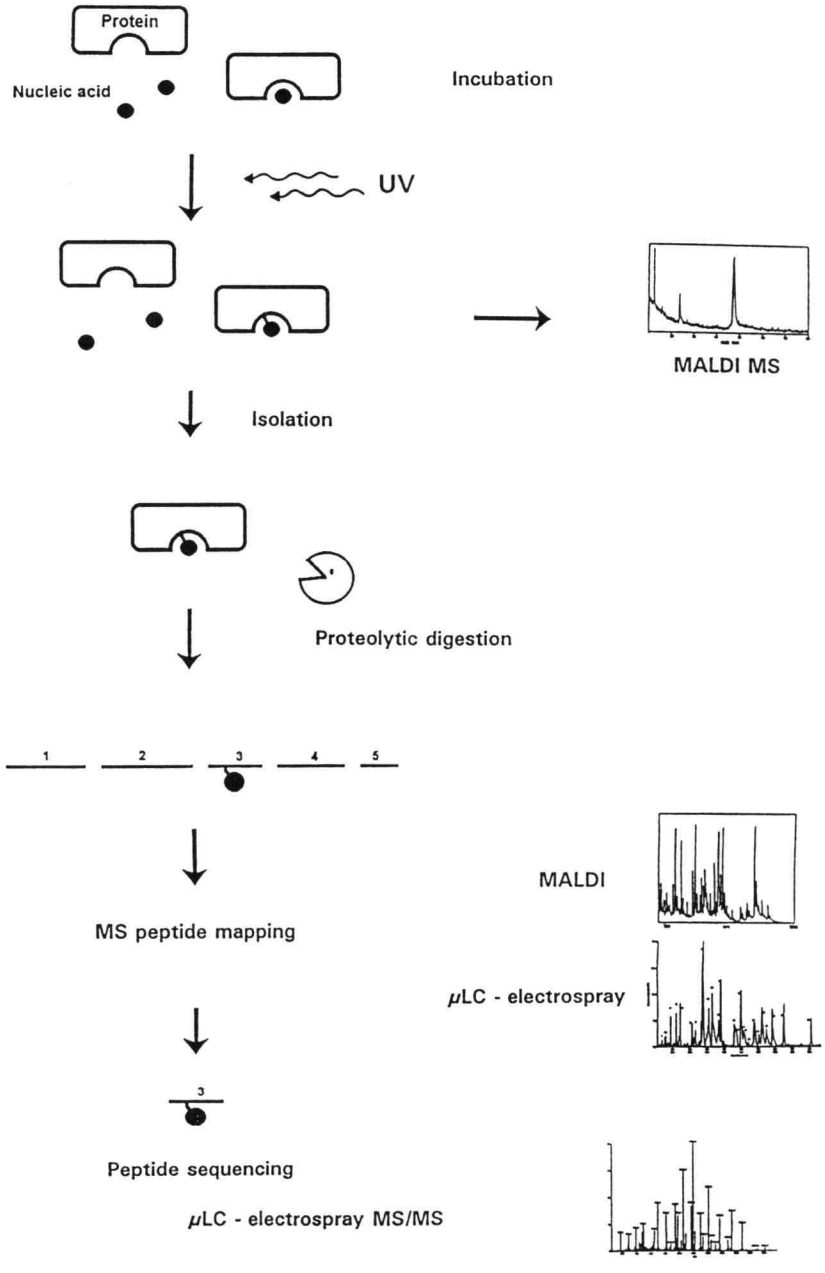


Figure 7. Strategy for characterization of protein-nucleic acid contacts using photochemical crosslinking and mass spectrometry.
The protocol consists of three parts: i) Photochemical crosslinking of protein to nucleic acid; ii) Peptide mapping to identify crosslinked peptide-nucleic acid complexes; iii) Identification of crosslinked amino acid residues by sequencing of covalent peptide-nucleic acid complexes.

Since the UV-crosslinked protein-nucleic acid complex is a covalently modified protein, comparison of mass spectrometric peptide maps of the protein produced before and after UV-crosslinking to the nucleic acid can be used to identify peptides that deviate from their expected molecular weight by a mass increase that corresponds to a nucleic acid addition (the modification in this case). This strategy relies on the fact that, in most cases, the amino acid sequence of the nucleic acid binding protein will be known and that it will be possible to predict the respective molecular weights of the peptides resulting from cleavage by a specific enzyme. However, even in cases where the protein sequence is unknown, a change of retention time and mass of the adducted vs. non-adducted peptides might serve to identify the crosslinked peptide.

In the third stage, the modified peptides identified in the previous stage are sequenced by ESI tandem mass spectrometry (MS/MS) in order to identify the amino acid residues involved in forming the photochemical crosslink to the nucleic acid, and thereby, the nucleic acid binding sites of the protein. The use of fast atom bombardment MS/MS for the final characterization (sequencing) of UV-crosslinked peptide-nucleic acid complexes has been demonstrated (47, 75, 79). However, fast atom bombardment ionization lacks the sensitivity of MALDI and ESI. Since these latter techniques require only femtomole to low picomole amounts of sample for protein and peptide analysis, the amount of material needed for the full characterization of a UV-crosslinked protein-nucleic acid complex via our protocol should not exceed a few tens of picomoles. Sensitivity at this level is essential for UV laser crosslinking and in vivo studies.

BIOLOGICAL PERSPECTIVES

Due to the improved sensitivity and speed available through implementation of the above strategy more experiments can be performed on a given amount of protein and nucleic acid. For example, crosslinking sites can be studied as a function of salt concentration or in the presence of other macromolecules. Such experiments might provide information on how conformational changes affect protein-nucleic acid complexes. Because our protocol has the potential to be faster than the traditional approach (that is, radiolabeling, HPLC peptide mapping, and Edman degradation), it becomes feasible to apply the strategy to a larger number of proteins and, in turn, possibly identify common or new nucleic acid binding motifs.

Employing mass spectrometry to identify photo-products can allow us to learn more about the chemical changes that macromolecules undergo during UV-irradiation. Specifically, it is possible to characterize the photochemical process and the type of bond that is formed between protein and nucleic acid. Such information might relate to other fields of research, such as aging

and cancer, where DNA or protein modifications induced by external factors degrade or alter the cellular control mechanisms.

In many cases it is easier to obtain structural information of an unliganded protein than of the corresponding protein-nucleic acid complex by techniques such as X-ray diffraction and NMR spectrocopy. Using molecular modeling software and graphical workstations it is possible to fit the nucleic acid to the putative binding site in the protein. Photochemical crosslinking and identification of the crosslinked amino acids by mass spectrometry can be used to develop or establish such models. The proposed protocol might therefore be used to supplement and improve data obtained by the more established techniques.

As mentioned in a previous section, the importance of macromolecular assemblies in biological processes is emerging. Using our strategy to dissect multiprotein-nucleic acid complexes might provide insights into how individual components of the complexes carry out their respective roles in nucleic acid binding.

Ultimately, it might be possible to irradiate living cells or tissues, isolate specific crosslinked protein-nucleic acid complexes, and identify the contact sites by using this sensitive protocol.

ANALYTICAL CHALLENGE

Successful detection of crosslinked amino acid-nucleotide adducts demands sensitive and gentle analytical techniques. The crosslinking yield can vary from below 1% to above 70% depending on the system under study, the irradiation conditions, and the type of nucleic acid that is used. The protein $\times$ nucleic acid complex[1] might be unstable to acid, base, high temperature, or salt. Very little is known about the covalent bonds that are induced by UV-irradiation of protein-nucleic acid complexes as mentioned previously, and this adds some uncertainty to this pursuit. More information should be forthcoming by implementation of our protocol since mass spectrometry can reveal structural information about molecules. The crosslinked species are often present in complex mixtures, a condition that necessitates high sensitivity and dynamic range of the analytical techniques.

Mass spectrometry fulfills these requirements. The absolute sensitivity for individual peptides and nucleic acids is better than 1 picomole by several orders of magnitude. Specificity is obtained by accurate molecular weight determination. The ultimate analytical method, chromatographic separation with online mass spectrometric detection, is now available and provides high separation efficiency coupled to specific and sensitive detection of analytes.

---

[1]The nomenclature A $\times$ B indicates a photochemically crosslinked A-B complex.

AIMS

The general aim of the present investigation has been to establish the feasibility of the above mentioned strategy and to identify problems associated with it. We have approached this by:

1) Analyzing intact UV-crosslinked protein-nucleic acid complexes by mass spectrometry. Molecular weight data can directly establish that a complex has been formed and can provide information on the stoichiometry of the components of the complex.

2) Demonstrating that covalent peptide-nucleic acid complexes, derived from UV-crosslinked protein-nucleic acid complexes by protease digestion, can be analyzed by mass spectrometry. Accurate molecular weight determination will help identify the peptide portion of the complex because the peptide mass can be calculated by subtraction of the known nucleic acid mass from the experimentally obtained mass of a peptide-nucleic acid complex. The molecular weight of the peptide along with the cleavage specificity of the protease is enough to map the peptide in the protein sequence.

3) Demonstrating that mass spectrometry can be used in combination with chromatography to identify covalent peptide-nucleic acid complexes in peptide mixtures. Chromatographic separation of peptides and peptide-nucleic acid complexes with on-line mass spectrometric detection will be a powerful method to rapidly localize the nucleic acid domain of a protein.

4) Demonstrating that tandem mass spectrometry can be used to sequence peptide-nucleic acid complexes, thereby localizing the amino acid residue(s) involved in forming the covalent bond to the nucleic acid.

5) Demonstrating that molecular weight information obtained by mass spectrometry of UV-crosslinked protein-nucleic acid complexes and peptide-nucleic acid complexes can be used to localize nucleic acid binding domains in proteins.

Chapter 4: Model systems

In order to establish the feasibility of the proposed protocol we have used a variety of biological systems consisting of nucleic acid binding proteins and their respective ligands. The gene 32 protein is involved in DNA replication, recombination and repair, transcription termination factor Rho is an important component in transcription, whereas uracil-DNA glycosylase is a DNA repair enzyme. These three biological systems and a chemical model system are described below.

Phage T4 gene 32 protein

Phage T4 gene 32 protein is the single-stranded DNA-binding protein in a 5 protein system involved in the elongation phase of leading strand DNA synthesis (138, 139, 140). Its apparent role is to bind to the transiently formed single-stranded DNA-structures at the replication fork and stabilize them. The gene 32 protein (gp32) is also involved in recombination and repair of DNA where it serves to protect and stabilize single-stranded DNA (ssDNA).

The gene 32 protein is present at a concentration of 3 μM in T4 infected *E. coli* cells (approximately 10,000 molecules/cell). It controls its own translation by specific, cooperative binding to gp32 mRNA (141). The gp32 polypeptide contains 301 amino acid residues (Figure 8) and has a molecular weight of 33,502. Hydrodynamic studies have shown the protein to be a prolate ellipsoid with a 4:1 axial ratio and a 12 nm length.

During replication multiple gp32 molecules bind to ssDNA like pearls on a string. It has been demonstrated that DNA-binding induces a conformational change of gp32 (142). DNA-binding activity resides in the central domain of the protein as shown by limited proteolysis (143), photoaffinity-labeling (144), and UV-crosslinking (67), whereas protein-protein interactions are mediated by the acidic C-terminal domain of the protein (145, 146). The central DNA-binding region of gp32 has a zinc-finger like motif (3 Cys, 1 His) that presumably binds Zn^{2+}. The presence of one Zn^{2+} per gp32 is necessary for cooperative binding to DNA (147) and for translational autoregulation of gp32 mRNA (148). The amino-terminal region is necessary for cooperative binding of gp32 to ssDNA. A structural model has been proposed in which the acidic

C-terminal domain of gp32 folds into a four-helix bundle that might interact with other proteins (146).

The DNA binding site of gp32 covers 8 bases in ssDNA (49). NMR studies performed in the absence and presence of ssDNA indicates that aromatic amino acids (Phe, Trp) are involved in protein-DNA interactions (149, 150). It has been proposed that a stretch of amino acids from Tyr^{73} to Tyr^{115} define part of the DNA binding surface. Subsequent UV-crosslinking experiments have identified Phe^{183} as the crosslinked residue (67). Although a lot of effort has been put into characterizing DNA-binding properties of gp32 over the last 25 years the molecular details are still not known – and there is still no 3-dimensional structure available on gp32 although both NMR and X-ray studies were initiated years ago (141).

Single-stranded DNA binding by gp32 is suggested to be mediated by hydrophobic interactions between aromatic amino acid sidechains and nucleobases (stacking) and charge-charge interactions between positively charged amino acid sidechains and negatively charged DNA backbone phosphate groups (7).

Leading strand DNA synthesis proceeds at a rate of several hundred nucleotide residues per second. Since a number of sequential reactions are required to insert a single nucleotide residue at the 3' end of the nascent DNA chain (151), each template-dependent reaction cycle must go to completion in milliseconds. As is typical in cases where several proteins are involved, the protein-nucleic acid system in the T4 DNA replication system is an extremely labile complex that cannot survive rigorous biochemical analysis without the intervention of some technique like UV induced crosslinking. Given that the bacteriophage T4 leading strand replication system has been well characterized using other methods (138, 139) and that the amino acid sequence for gp32 is known, von Hippel and coworkers used gp32 to calibrate and characterize laser generated pulsed UV crosslinking as a tool for studying dynamic aspects of protein-nucleic acid interactions that take place on the millisecond time scale (49, 58, 152).

The work on the T4 gp32 protein has shown that a dose of approximately 10^{17} photons (266 nm) delivered in a single 5 ns pulse (generated by a frequency-quadrupled Nd:YAG laser) is optimal for inducing crosslinking between DNA and proteins. Under these experimental conditions ~4% of $(dT)_{10}$ oligomers put into the reaction vessel covalently bind to the reactant protein; by contrast, only 0.08-0.16% of cytidine or purine oligomers crosslink to the protein reactant. Typically, pulsed UV-crosslinking at 266 nm with a single pulse of $2\text{-}5 \cdot 10^{16}$ photons/sample works well for protein concentrations ranging from 10^{-8} to $5 \cdot 10^{-5}$ M and for nucleic acid concentrations of 10^{-8} to 10^{-4} M (in units of nucleotide residues).

The gene 32 protein crosslinked to $(dT)_{20}$ oligonucleotide by pulsed UV irradiation was used in our initial mass spectrometric studies of protein-nucleic acid complexes in order to establish the feasibility of the project.

Transcription termination factor Rho

NTPases, specific and non-specific single-stranded nucleic acid binding proteins, translocases, and helicases have activities that are critical to genetic recombination, replication, and critical aspects of gene expression, such as transcriptional control and message processing. Transcription termination factor Rho from *Escherichia coli* (Rho) has been widely adopted as a model system for the physical study of nucleic acid binding protein structure and activity (153, 154, 155, 156). The protein is a single-stranded RNA binding protein with an ATPase activity. It binds a wide subset of sequences defined within a particular RNA secondary structure (or non-structure). Rho will bind single-stranded RNA and DNA equally well, yet its ATPase activity is not viable when DNA is bound. Furthermore, the affinity of nucleic acid to the nucleic acid binding site of rho is determined by base content – not consensus sequence.

TRANSCRIPTION TERMINATION AND RHO

Transcription of a mRNA from a DNA-template involves three processes: initiatiation, elongation and termination. In *E. coli*, termination of the transcriptional process can be accomplished by two mechanisms called Rho-dependent and Rho-independent transcription termination, respectively. In Rho-independent transcription termination a DNA region of dyad symmetry followed by a stretch of A's usually encodes the transcription termination site. The former region allows a hairpin loop structure to form in the nascent RNA-transcript while the latter region weakens the RNA:DNA interaction. This causes instability of the transcriptional "bubble", facilitating the dissociation of the RNA-polymerase/DNA/RNA ternary complex and subsequent release of the RNA-transcript (157). This is the most common termination mechanism in procaryotes. The transcript is always terminated at the same site (within 1-2 nucleotide units), that is, the mRNA size for a given gene is constant. In contrast to this sequence-dependent termination event, Rho-dependent transcription termination requires a protein factor to release the nascent RNA-transcript while the RNA-polymerase is paused at specific Rho-dependent termination sites. There is no obvious consensus sequence that dictates Rho-dependent termination events as will be explained below.

Transcription termination factor Rho is a RNA-dependent ATPase. Rho-dependent transcription termination is presumably initiated by the Rho hexamer binding to a (rC)-rich recognition site on the nascent RNA-transcript. This event is followed by an ATP-driven extension of the interaction towards the 3' end of the transcript. If the RNA-polymerase is stalled

at a pause-site, it will allow Rho to catch up and this eventually causes the release of RNA from the ternary transcription complex. The sizes of RNA transcripts terminated by this mechanism vary because the process is dependent on the kinetics of RNA-polymerase pausing and Rho movement (ATPase activity). The details of the Rho-dependent termination mechanism are not yet fully understood, although the large amount of information accumulated over the last few years, have spawned several models. Rho apparently acts as an ATP-driven DNA-RNA helicase and therefore it belongs to an important class of proteins that couple NTP-hydrolysis with physical action on a nucleic acid substrate (155). This class includes DNA-helicases involved in replication, recombination and repair and translational initiation factors (such as eIF4A and eIF4F) that promote the binding of ribosomes to eucaryotic mRNA and the search for initiation sites. Elucidating the molecular details of Rho-action therefore can provide insight into mechanisms involved in important cellular processes. The following overview of our current knowledge of Rho structure and function is based on reviews by Bear and Peabody (154), Richardson (155) and Platt and Richardson (156). More recent references are included, where appropriate.

RHO STRUCTURE

The Rho polypeptide is composed of 419 amino acids (Figure 9); the molecule has a molecular weight of 47,004 and an isoelectric point close to 9. Rho contains only one cysteine residue and therefore no disulphide bridges are present. The active form of Rho is a hexamer made of three functional units of asymmetric Rho dimers. It has a planar hexagonal geometry (closed ring or lock-washer appearance) with D_3 symmetry as revealed by physical-chemical studies (158, 159) and cryo-electron microscopy (160). Each Rho subunit has one RNA-binding site and one NTP binding site. The RNA-binding site is located in the 151-residue N-terminal domain of the Rho polypeptide whereas the NTP-binding site is located in the central domain of the polypeptide. An amino acid sequence of proposed purine nucleotide binding sites [A/G _ _ G _ G K T/S] can be found in several prokaryotic and eucaryotic helicases and also exists in Rho [^{178}App(K)aGKT]. Site-directed mutagenesis studies also indicate that this region of Rho is involved in NTP binding and furthermore that Asp^{265} is required for ATPase activity (161). Photoaffinity-labeling studies using azido-ATP identified Lys^{181} as the labeled residue (162, 163). Another, although weaker, NTP-binding sequence motif can be found in the region of amino acid 345-359.

NTP-BINDING SITE OF RHO

The ATPase activity of Rho is only present when RNA is bound. Work performed by von Hippel and coworkers indicates that the functional units in the Rho hexamer are the dimers (164, 165). The Rho hexamer has six nucleoside-triphosphate (NTP) binding sites, one on each subunit. Three of these are considered high-affinity sites and the other three are low-affinity sites (164). However, all binding sites have a binding constant that allows binding of ATP, in principle allowing all ATP-binding sites to be involved in the catalytic ATPase cycles. Physical-chemical studies and limited trypsinolysis experiments indicate that the Rho hexamer undergoes a conformational change upon binding of ATP, in effect, reducing the symmetry of the Rho hexamer from D_3 to "pseudo-D_3". This might have implications for the mechanism by which Rho moves uni-directionally (5' to 3') along the RNA-transcript, in that unidirectional movement by association-dissociation events can be achieved only by an asymmetric structure (164). Also, conformational change in one functional dimer might change the ATP binding constants of the neighbouring dimers, allowing cooperativity or allosteric regulation between functional dimers.

RNA-BINDING SITE OF RHO

As stated previously, each Rho subunit has a RNA-binding site making a total of six in the hexamer. Three are high-affinity and three are low-affinity RNA-binding sites, that differ 10-fold in binding constant (165, 166). All six RNA-binding sites of Rho have to be occupied in order to maintain the ATPase activity (165). The bound Rho hexamer covers approximately 70 nucleotides on the substrate RNA-transcripts and RNA-binding induces a conformational change of Rho. Apparently, Rho has no RNA-sequence or -structure requirements for binding to RNA. Rather, a lack of RNA secondary and tertiary structure and the presence of rC-clusters in the RNA are necessary for binding. Both poly(rC) and poly(dC) can bind to Rho, but the latter inactivates the Rho ATPase, indicating that Rho is able to discriminate between RNA and DNA. Using synthetic (rU, rC)-octamers and -decamers, it was shown that each RNA-binding site covers 8-9 nucleotides (166, 167). It was demonstrated that Rho ATPase activity is oligonucleotide-sequence and -composition dependent in that stretches of 5 or more rC units provide the largest association constant for the RNA-binding sites of Rho, while rU units situated at the 5' end of (rU, rC)-octamers provide the greatest stimulation of the ATPase activity of Rho (166, 168). This demonstrates that RNA-binding by Rho is 'polarized', that is, it requires proper 5' to 3' orientation of the RNA-transcript in the binding sites, where rC residues promotes the binding of the oligonucleotide and rU residues stimulates the ATPase activity. This requirement for "orientation"

may provide another means for uni-directional movement of the rho hexamer along the RNA-transcript. Occupancy of the NTP binding sites by ATP, ADP or a non-hydrolyzable ATP analog had no effect on the binding of (rU, rC)-oligonucleotides, whereas RNA-binding is necessary for the ATPase activity of Rho. Furthermore, the RNA binding constant does not correlate directly with ATPase activity, and therefore, it is proposed that one has to consider the RNA-binding event and the ATPase driven movement of Rho along the RNA-transcript as separate aspects (166). As for the ATP-binding site, RNA-binding to one functional dimer might affect binding in the neighbouring dimers, that is, cooperative or allosteric effects might be present in the hexamer.

AIMS

The aim of the photoaffinity labeling and mass spectrometric studies presented in the following chapters is to establish the proposed strategy and to locate the NTP binding site of Rho in order to get a better understanding of the mechanisms of Rho function. Previously published data on photoaffinity-labeling of the ATP-binding site of Rho (162, 163) and site-directed mutagenesis studies (161) indicate that the region around the Lys^{181} residue is involved in ATP binding. We want to confirm this finding using 4sUDP as a photoaffinity label and our analytical protocol, and subsequently use RNA with 4sU incorporated to label the RNA binding site of Rho.

The ATP-binding site of Rho is well suited to study by photoaffinity labeling with the ATP-analog 4-thio-uridine diphosphate (4sUDP). UV crosslinking experiments performed with 4sUDP at 335 nm are selective and relatively devoid of extraneous photoreactions because the excitation wavelength is far from the absorbance maxima of amino acids or other nucleic acids (61). Compared to azido labels (48), 4sUDP is much easier to work with. Since 4sUDP occurs in nature with biological properties very similar to uracil, it can be readily incorporated anywhere within an oligonucleotide reactant making it possible, for example, to map an entire DNA- or RNA-binding site on a protein.

Uracil-DNA glycosylase

This short overview is based on a recent and extensive review by Mosbaugh and Bennett (169) of the uracil-excision repair pathway and the role of uracil-DNA glycosylase in this process.

Uracil can be introduced or incorporated into DNA by several routes: Deamination of cytosine is believed to be the predominant pathway, but high levels of dUTP in the cell can result in misincorporation of U instead of T during DNA synthesis. The implications of the presence of U in DNA is still not fully understood. In some cases cells can survive U incorporation under conditions where uracil excision repair is absent. However, in the presence of uracil excision repair activity, excess uracil-incorporation is lethal to cells. This effect is caused by the activity of the uracil-DNA glycosylase (Ung) enzyme.

Ung catalyzes the cleavage of the N-glycosidic bond between the uracil base and the deoxyribose backbone of DNA and thereby initiates the uracil excision repair pathway. This pathway causes the apyrimidinic (AP) site to be repaired by cleavage of the damaged DNA strand (incision) next to the AP-site, followed by a one-nucleotide excision repair process that includes removal of the AP-site, DNA synthesis, and ligation. This is in contrast to (oligo)nucleotide excision-repair pathways where a longer patch (12-15 bases) is removed by an exonuclease enzyme, followed by DNA synthesis and ligation. This implies that the uracil excision repair pathway is concerted, that is, binding of Ung allows only one-nucleotide repair enzymes access to the DNA – the regular nucleotide excision repair enzymes are excluded from reaching the AP site. It has therefore been suggested that Ung serves as a nucleation point for other enzymes, although no evidence for protein-protein interactions has been provided.

Ung has been isolated from many species, indicating that it is an enzyme of major importance to cells and organisms. This is also suggested by the large degree of homology and amino acid identity (40-50%) that is found between Ung enzymes isolated from different species. The *E. coli* enzyme, which was used in the present work, consists of 228 amino acids (Figure 10), excluding an N-terminal Met that is posttranslationally removed. Ung has a molecular weight of 25,563 and a pI of 6.6, contains one cysteine residue, and has a "normal" amino acid composition (9.2% acidic residues, 8.3% basic residues).

Ung binds to single stranded DNA with a two-fold specificity relative to double stranded DNA and has no cofactor requirement. The enzyme is inhibited by uracil and DNA containing AP-sites. The phage PBS2 uracil-DNA glycosylase inhibitor (Ugi) is a noncompetitive inhibitor of Ung. It forms a very stable 1:1 complex with the enzyme.

The aim of the present study is to locate the DNA binding region in the Ung polypeptide. This will provide information as to which amino acids are important for DNA binding. Assuming that

the catalytic site is near the DNA binding site, identification of the latter might also provide information on the catalytic site and in turn shed light on the mechanism involved in cleavage of the N-glycosidic bond. Following identification of the DNA binding domain of Ung site-directed mutagenesis studies will be initiated to further characterize the binding site.

Synthetic peptide-oligonucleotide hybrids

During the course of the present study we found it necessary to synthesize peptide-oligonucleotide hybrids in order to learn more about the chromatographic and mass spectrometric behavior of these species. We were particularly interested in testing the applicability of tandem mass spectrometry for structural elucidation of peptide-oligonucleotide hybrids.

We synthesized peptide-linker-oligonucleotide hybrids by coupling a peptide carboxyl group to an amino group on an aminolinker-5'-oligonucleotide. The two species were coupled by formation of an amide bond using a catalyst. The purified peptide-linker-oligonucleotides were analyzed by reversed phase HPLC, MALDI, and ESI mass spectrometry.

MFKRKSTAEL AAQMAKLNGN KGFSSEDKGE WKLKLDNAGN 40
GQAVIRFLPS KNDEQAPFAI LVNHGFKKNG KWYIETCSST 80
HGDYDSCPVC QYISKNDLYN TDNKEYSLVK RKTSYWANIL 120
VVKDPAAPEN EGKVFKYRFG KKIWDKINAM IAVDVEMGET 160
PVDVTCPWEG ANFVLKVKQV SGFSNYDESK FLNQSAIPNI 200
DDESFQKELF EQMVDLSEMT SKDKFKSFEE LNTKFGQVMG 240
TAVMGGAAAT AAKKADKVAD DLDAFNVDDF NTKTEDDFMS 280
SSSGSSSSAD DTDLDDLLND L 301

Figure 8. Amino acid sequence of phage T4 gene 32 protein.

MNLTELKNTP VSELITLGEN MGLENLARMR KQDIIFAILK 40
QHAKSGEDIF GDGVLEILQD GFGFLRSADS SYLAGPDDIY 80
VSPSQIRRFN LRTGDTISGK IRPPKEGERY FALLKVNEVN 120
FDKPENARNK ILFENLTPLH ANSRLRMERG NGSTKDLTAR 160
VLDLASPIGR GQRGLIVAPP KAGKTMLLQN IAQSIAYNHP 200
DCVLMVLLID ERPEEVTEMQ RLVKGEVVAS TFDEPASRHV 240
QVAEMVIEKA KRLVEHKKDV IILLDSITRL ARAYNTVVPA 280
SGKVLTGGVD ANALHRPKRF FGAARNVEEG GSLTIIATAL 320
IDTGSKMDEV IYEEFKGTGN MELHLSRKIA EKRVFPAIDY 360
NRSGTRKEEL LTTQEELQKM WILRKIIHPM GEIDAMEFLI 400
NKLAMTKTND DFFEMMKRS 419

Figure 9. Amino acid sequence of E. coli transcription termination factor Rho.

MANELTWHDV LAEEKQQPYF LNTLQTVASE RQSGVTIYPP 40
QKDVFNAFRF TELGDVKVVI LGQDPYHGPG QAHGLAFSVR 80
PGIAIPPSLL NMYKELENTI PGFTRPNHGY LESWARQGVL 120
LLNTVLTVRA GQAHSHASLG WETFTDKVIS LINQHREGVV 160
FLLWGSHAQK KGAIIDKQRH HVLKAPHPSP LSAHRGFFGC 200
NHFVLANQWL EQRGETPIDW MPVLPAESE 229

Figure 10. Amino acid sequence of E. coli uracil-DNA glycosylase.

Chapter 5: Experimental

High performance liquid chromatography

SEMI-PREPARATIVE HPLC

A Beckman Model 100A HPLC system equipped with a Vydac C_{18} column (The Separations Group, Hesperia, CA; 4.6 mm x 250 mm, 5 μm particles, 300 Å pores) was used for semi-preparative purification of peptide and synthetic peptide-nucleic acid hybrids. Peptides were detected by UV-monitoring at 220 nm using a Waters Model 486 UV detector. Nucleotides and synthetic peptide-linker-nucleotide hybrids were detected by UV-monitoring at 256 nm.

MICROBORE HPLC SYSTEM

An ABI 140B syringe pump (Perkin-Elmer/Applied Biosystems Division) equipped with a Rheodyne 8125 injector (5 μL sample loop) and a 1 mm x 250 mm Vydac C_{18} column (The Separations Group, Hesperia, CA) was used for all separations. An ABI/Kratos model 783 UV-detector equipped with a microbore flow cell (1 μL dead volume) was used to detect eluting compounds. For microbore HPLC-MS applications the column was connected to the electrospray mass spectrometer by a 50 μm ID fused silica capillary. Further details of the microbore HPLC-MS setup can be found in the mass spectrometry section below.

CAPILLARY HPLC SYSTEM

Nanoscale separations were carried out by using custom packed 320 μm x 110 mm or 250 μm x 250 mm fused silica capillary columns.

In order to obtain a low flowrate, while still getting reproducible gradients, the ABI 140B pump was run at a flow rate at 40-80 μL/min and the flow split to approximately 5 μL/min before the injector. The splitting tee was fabricated from 50 μm ID fused silica capillary and an Upchurch tee (~100μm ID). The splitting ratio was set by the length (backpressure) of the open-ended capillary.

Mass spectrometry

THE OSU MALDI-TOF MASS SPECTROMETER

A custom-built time-of-flight mass spectrometer (Figure 11) equipped with a frequency-tripled (355 nm) Nd:YAG laser (Spectra-Physics GCR-11) was used for positive and negative ion MALDI mass spectrometric analysis. Acceleration potentials of 24 kV and 18 kV were applied to the probe and the first extraction lens respectively. A negative potential of 200-600 V was applied to an ion guide-wire spanning the 60 cm flight path to improve ion transmission to the detector. A negative potential of 3,500 V was applied to the front of the instrument's detector, which is a hybrid consisting of a microchannel plate mounted in front of a mesh electron multiplier (Hamamatsu R2362).

The data acquisition system consists of a LeCroy TR8828 transient recorder controlled by a LeCroy 6010 Magic Controller that is interfaced to a PC via a National Instruments GPIB PCII board. Recorded time-of-flight mass spectra consist of the summed data generated from 30-50 consecutive 5-20 $\cdot$ 10^6 W/cm^2 laser pulses. Mass calibration and data analysis was performed using M-over-Z software (170) running on a PC or TOFMA software (171) running on an Atari Mega 4 computer.

SAMPLE PREPARATION FOR MALDI MASS SPECTROMETRY

The following compounds were purchased from Aldrich Chemical Co. for use as MALDI matrices: 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), 4-hydroxy-α-cyanocinnamic acid (HCCA), 4-hydroxy-3-methoxycinnamic acid (ferulic acid), 2,4,6-trihydroxyacetophenone (THAP), and 2,5-dihydroxybenzoic acid (gentisic acid or DHB).

Formic acid (98%) was purchased from Fluka. Diammoniumhydrogencitrate, trifluoroacetic acid and 1,1,1,3,3,3-hexafluoroisopropanol were purchased from Aldrich. All water was Millipore filtered. Acetonitrile was from Malinckroth (UltimAR grade). Further details of the solvent combinations used are mentioned in figure legends and tables in the Results chapter.

The matrix compounds were dissolved to a final concentration of 10 g/L (saturated solution) except HCCA ,which was dissolved to a final concentration of 5 g/L (saturated solution). Matrix solutions were centrifuged prior to use to precipitate insoluble matter. Typically, 1 μL sample was mixed into 3-9 μL of matrix solution to obtain a 1-10 μM analyte concentration. For

each analysis, 0.5-1 μL of the sample/matrix solution was deposited onto a stainless steel probe. Prior to insertion into the mass spectrometer, the deposit was dried under a stream of air at ambient temperature, washed by dipping into cold (4°C) water, and dried again.

Alternatively, the mass spectrometer target was prepared by depositing 1 μL of matrix solution which was gently wiped off after 10-15 seconds using a Kimwipe tissue. This allows a microcrystalline "seed-layer" to form on the probe tip. Subsequently, 1 μL of the analyte/matrix solution, prepared as above, was deposited on top of the seed layer and allowed to dry. In some cases the analyte/matrix crystals were rinsed by gently dipping the probe tip into water before the sample had dried totally ("wet-rinse"). The probe was then inserted into the mass spectrometer and mass spectra acquired as outlined above.

Equine cytochrome c (MW 12,360), equine myoglobin (MW 16,951), chymotrypsinogen A (MW 25,656), and bovine serum albumin (MW 66,430) were purchased from Sigma for use as internal mass calibrants.

THE OSU ELECTROSPRAY IONIZATION TANDEM MASS SPECTROMETER

In the spring of 1994 a Sciex API III Plus triple quadropole mass spectrometer (Perkin-Elmer/Sciex, Canada) was installed in the Department of Agricultural Chemistry/Environmental Health Sciences Center. This instrument uses pneumatically assisted electrospray (ion-spray) to create gas-phase ions at atmospheric pressure (Figure 12).

Calibration was performed in the positive ion mode using a mixture of polypropylene glycols. Positive ion mass spectra were collected using an electrospray potential of 5,000V and orifice potentials in the range of 60 to 110 V. In the negative ion mode the electrospray potential was set at -4100 V with the orifice potential set at -110 V. A curtain gas (nitrogen) flow of 0.6 L/min and a nebulizer gas (oxygen) flow of 0.6 L/min were used for flow injection analysis and capillary LC-MS.

Flow injection analyses at 3.3 μL/min were performed using a Harvard syringe pump connected to the electrospray interface via a Rheodyne 8125 injector (5 μL sample injection loop). For peptide analysis the solvent was an aquous solution of 0.1% formic acid/50% methanol. For oligonucleotide and peptide-oligonucleotide hybrid analysis the solvent was 5 mM ammonium acetate/50% acetonitrile (172). The mass range of interest, typically 400-2,200, was scanned by Q1 in approximately 4 seconds using a dwell time of 1 msec and a step size of 0.5 Da.

Sequencing by tandem mass spectrometry (MS/MS) was accomplished by selecting a doubly charged parent ion emerging from Q1 and scanning Q3. A collision cell gas (argon) pressure of 100-200 (arbitrary units, laboratory frame) was used.

Online detection of HPLC eluents by ESI-MS, i.e. LC-MS, was accomplished by connecting the above mentioned microbore HPLC system to the ESI interface with a 50 μm fused silica capillary. A flow rate of 40 μL/min was used. Negative ion mass spectra were collected as outlined above. The conditions for microbore and capillary HPLC-ESI MS peptide mapping of Rho are described in the Rho section of this chapter.

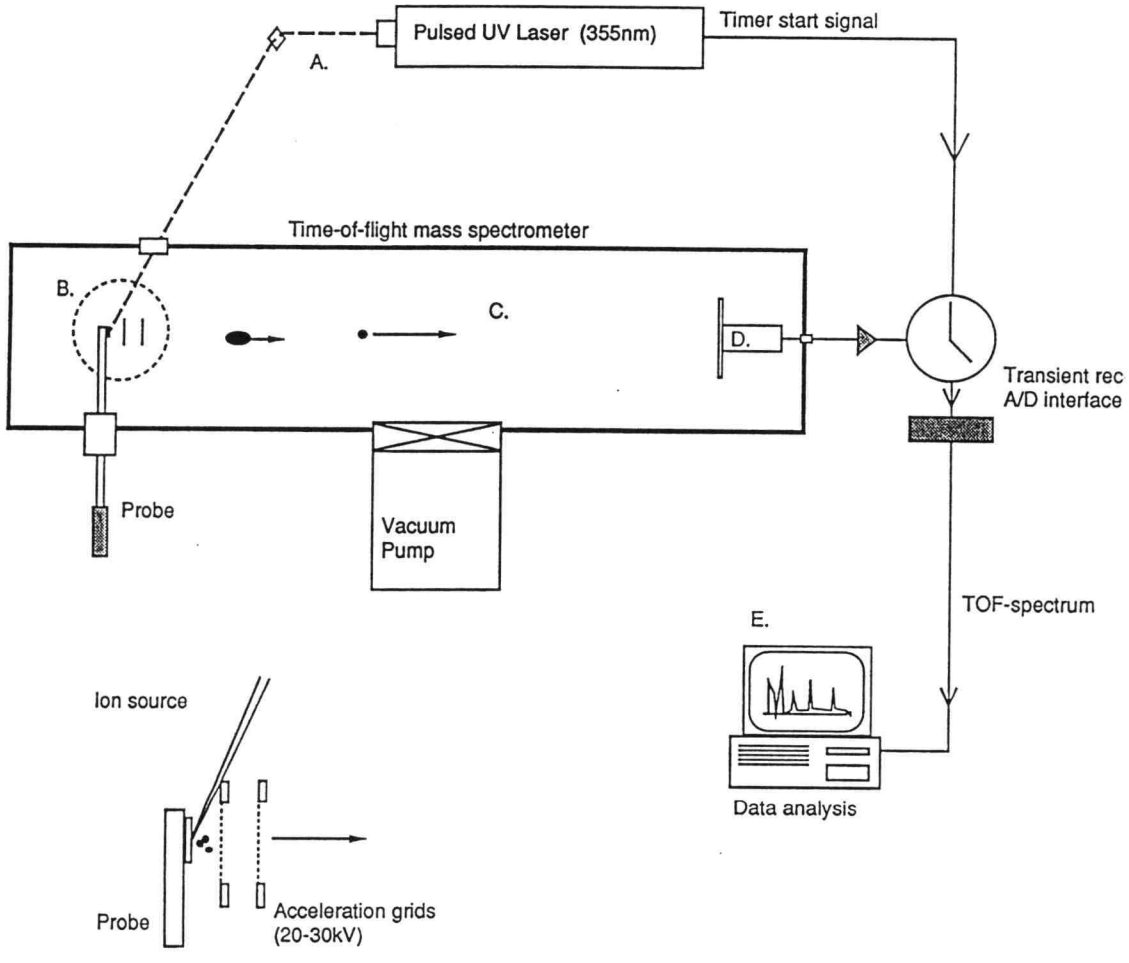
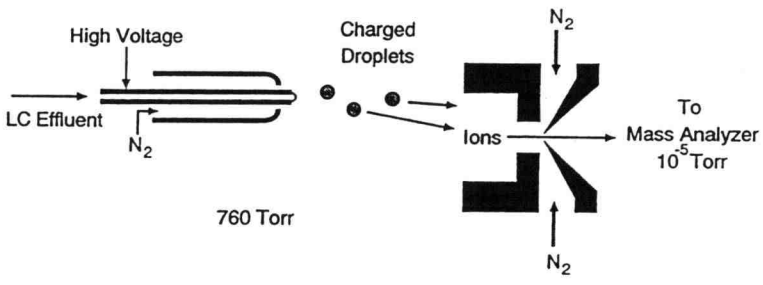
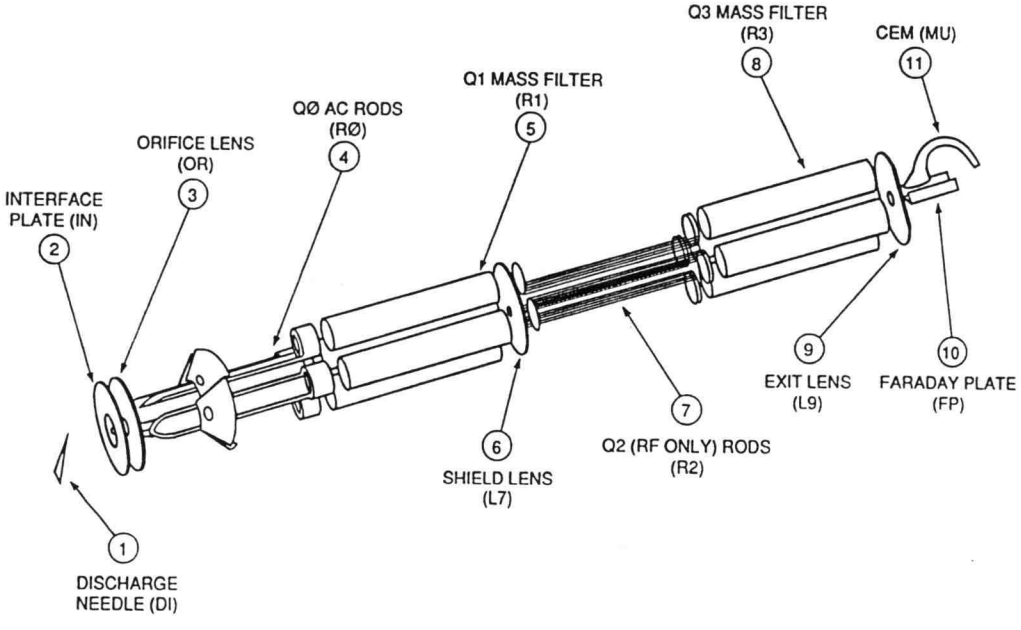


Figure 11. The OSU matrix-assisted laser desorption/ionization time-of-flight mass spectrometer.





Pictorial Illustration of the Ion Optics Path

Figure 12. The Sciex API III Plus electrospray ionization triple-quadropole mass spectrometer.

gp32 project

PULSED UV LASER CROSSLINKING OF GP32 TO dT_{20}

Oligomeric dT_{20} was purchased from Pharmacia. Purification of gp32 followed the method of Nossal (173), except that a phenyl Sepharose column step was added after the DEAE resin step specified in the original procedure.

Pulsed UV-laser crosslinking of gp32 to dT_{20} oligomers was performed on 10 µL mixtures containing 22.5 µM gp32 and 10 µM dT_{20} in 20 mM HEPES, pH 7.5, 100 mM KOAc, 1 mM EDTA, and 0.5 mM DTT. 25% glycerol was present in some of the samples. In order to monitor the efficiency of crosslinking, a mixture containing a trace amount of 5'-[^{32}P]-$(dT)_{20}$ was crosslinked at the same time. Samples of dT_{20} were radioactively 5'-labeled with T4 polynucleotide kinase (174). After crosslinking, radioactive mixtures were subjected to electrophoresis on 10% polyacrylamide-SDS gels. Yields of crosslinking were determined by scanning the dried gel on an AMBIS 4000 radiodetection unit (AMBIS, Inc.). The protocol for irradiating gp32-dT_{20} mixtures was as described by Hockensmith *et al.* (58), with the following modifications: i) the laser was operated in the single-shot mode in order to obtain an 8 ns pulse; and ii) between samples, the laser was switched to continuous pulsed mode to stabilize the pulse energy. The energy of each single laser pulse was measured to be 22.9 ± 3.0 mJ (3.0 ± 0.4 · 10^{16} photons) using a pyroelectric detector (Oriel, model 70825) connected to a digital storage oscilloscope (Tektronix, model 2201).

The UV-laser crosslinking experiments were performed by Dr. Mark Young in the laboratory of Peter von Hippel, Institute of Molecular Biology, University of Oregon.

SAMPLE PREPARATION FOR MALDI MASS SPECTROMETRY OF UV-CROSSLINKED GP32-dT_{20} COMPLEXES

Samples were prepared as outlined above. Further details on matrix and solvent combinations are given in the Results chapter.

Ung project

UV-CROSSLINKING OF URACIL-DNA GLYCOSYLASE TO dT_{20}

Preparations of Ung and Ung x dT_{20} complex were provided by Samuel E. Bennett and Professor Dale W. Mosbaugh. The oligonucleotide dT_{20} was UV-crosslinked to the Ung protein as described in detail elsewhere (3). Briefly, Ung was incubated with an 83-fold molar excess of dT_{20} and UV-irradiated at 254 nm for 30 minutes. The crosslinking yield was calculated to be approximately 5% relative to Ung. Crosslinked Ung x dT_{20} complex was isolated by anion-exchange chromatography and DNA-Sepharose chromatography. Tryptic digestion of purified, crosslinked Ung x dT_{20} complex and purification of peptide x dT_{20} hybrids by anion-exchange chromatography was performed as described in detail elsewhere (3). Prior to mass spectrometric analysis, the samples were desalted using Nensorb cartridges and eluted with water/methanol (1:1).

MASS SPECTROMETRY OF PURIFIED, UV-CROSSLINKED URACIL-DNA GLYCOSYLASE X dT_{20} COMPLEX.

A 10 g/L solution of 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) dissolved in 33% formic acid/33% acetonitrile was used as a matrix for MALDI mass analysis of purified UV-crosslinked Ung-dT_{20}. For each analysis, 0.5 µL of the matrix solution was deposited on the probe and dried under a stream of air. Another 0.5 µL of the matrix solution was deposited on top of the initial application and followed immediately by addition of 0.5 µL of a solution containing the purified crosslinked Ung-dT_{20} complex. The sample was allowed to dry by slow evaporation and the dried deposit was washed by rinsing in cold (4°C) Millipore-filtered water. The sample was then inserted into the mass spectrometer and analyzed by MALDI.

All time-of-flight mass spectra were generated by summing the signals generated by 30-50 individual laser pulses and were calibrated by using ion-signals from the matrix and dT_{20} as internal mass calibrants. The molecular weights reported in the Results chapter are averages of several independent measurements.

PROTEOLYTIC DIGESTION OF UNG AND UNG X dT_{20} BY TRYPSIN

The experimental conditions for over night digestion of *S*-alkylated Ung and Ung x dT_{20} by trypsin are given by Bennett *et al.* (3).

MASS SPECTROMETRIC PEPTIDE MAPPING OF UNG AND UNG X dT_{20} TRYPTIC DIGESTS.

A 5 g/L solution of 4-hydroxy-α-cyanocinnamic acid in 0.1% trifluoroacetic acid/33% acetonitrile was used as a matrix for MALDI mass spectrometric mapping of tryptic peptides derived from Ung and Ung x dT_{20}. The samples were prepared by mixing 0.5 µL of the peptide mixture into 2-3 µL of the matrix solution. For each analysis, 1 µL of this solution was deposited on the mass spectrometric probe and dried under a stream of air. The crystallized sample was rinsed prior to MALDI mass analysis by dipping the probe tip into cold (4°C) water in order to remove urea and other interfering components present in the sample.

MASS SPECTROMETRY OF PURIFIED PEPTIDE X dT_{20} COMPLEXES.

MALDI mass analyses of desalted samples of peptide-dT_{20} complexes were performed using 10 g/L solutions of 2,4,6-trihydroxyacetophenone (THAP) or 4-hydroxy-3-methoxycinnamic acid (ferulic acid) in 50 mM diammoniumhydrogen citrate/50% acetonitrile as the matrix. The samples were mixed 1:1 or 1:2 with the matrix solution on the mass spectrometric probe and allowed to dry at room temperature.

Rho project

UV-INDUCED CROSSLINKING OF TRANSCRIPTION TERMINATION FACTOR RHO TO 4SUDP

Rho protein was prepared and stored as described by Gogol *et al.* (160); 4-thio-uridine diphosphate (4sUDP) was purchased from Sigma. Due to the size of the Rho protein (47 kD), gel electrophoresis cannot discriminate between the unlabeled and photoaffinity-labeled protein; therefore, an ATPase inactivation assay is employed to monitor the UV crosslinking reaction

between 4sUDP and Rho. The protein was assayed for NTPase activity by utilizing 50 nM final Rho monomer concentration and 20 μM poly(rC) (in nucleotides) as described by Seifried *et al.* (175). Photolabeling experiments were carried out in NTPase solution with 5 μM Rho and 50 μM 4sUDP. Solutions were irradiated by a 450 W or 150 W Xenon lamp whose output was passed through a monochromator to yield 335 nm light. Filtering the output for infrared radiation was found to be unnecessary. Aliquots were removed at various times to determine the rate of crosslinking by assaying the photo-inactivation of the NTPase activity of Rho. The rate of photo-inactivation of Rho was found to be linear with time and directly related to the slit-width of the monochromator (Swenson and Seifried, U. of Hawaii, *manuscript in preparation*). This is an indication of a well-behaved photochemical reaction with no dead-end products or unproductive photochemistry, which is in accordance with our knowledge of the 4sUDP photosystem. Inactivation of Rho by 4sUDP does not occur in the absence of light.

Addition of ADP or ATP to the photolabeling solution decreased the rate of photolabeling, a clear indication that the adenine-containing nucleotide was competing with the 4sUDP for the NTP binding site of Rho. Complex behavior of the photolabeling rate was observed when polymeric nucleotides were bound to Rho during irradiation indicating the lack of 4sUDP binding to the polynucleotide cofactor binding site of Rho.

Photolabeling of Rho by 4sUDP was carried out by Stephen Swenson in the laboratory of Dr. Steven Seifried, Department of Biochcemistry and Biophysics, University of Hawaii at Manoa.

Samples containing respectively Rho, UV-irradiated Rho, Rho + 4sUDP (non-irradiated), UV-irradiated Rho + 4sUDP, and UV-irradiated 4sUDP were prepared. Rho concentrations were 20 μM and 4sUDP concentrations were 200 μM in ATPase buffer (0.1 M KCl, 20 mM HEPES, pH 7.5, 10 mM MgOAc).

REDUCTION AND *S*-ALKYLATION OF RHO AND RHO X 4SUDP

For *S*-carbamidomethylation, 10 μL of each of the above samples was added to 12 μL solution containing 0.1 M Tris-HCl (pH 8), 5 mM dithiothreitol (DTT) and 2 mM $CaCl_2$ and incubated for 30 minutes at 50°C. After cooling the samples to room temperature, 1 μL of 300 mM iodoacetamide was added, and the samples were incubated for 15 minutes at room temperature.

PROTEOLYTIC DIGESTION OF RHO AND *S*-ALKYLATED RHO

In preparation for proteolytic digestion, non-alkylated samples were mixed 1:1 with 0.1 M Tris-buffer, pH 8. Alkylated samples were prepared as described above.

For tryptic digestion of Rho (cleavage after Lys and Arg), 1 μL of an aquous 0.5 g/L solution of trypsin (Boehringer-Mannheim, Sequencing Grade) was added to each sample. This corresponds to an enzyme to substrate ratio of 1:20 (w/w).

For proteolytic cleavage N-terminal to aspartic acid residues, 1 μL of a 0.1 g/L Endoproteinase Asp-N solution (Boehringer-Mannheim, Sequencing Grade) was added to each sample (enzyme/substrate ratio of 1:100 w/w).

Digestion with endoproteinase Lys-C (cleavage C-terminal to Lys) was accomplished by adding 1 μL of a 0.5 μg/μL solution of enzyme (Wako) to the non-alkylated protein samples.

All enzymatic digestion reactions were allowed to proceed for 20-22 hours at room temperature.

Chemical cleavage by cyanogen bromide (C-terminal to Met) was performed by mixing a 5 μL aliquot of protein stock solution with 98% formic acid (1:1) followed by addition of 1 μL of a 5 M solution of CNBr in acetonitrile. After incubation for 22 hours in the dark at room temperature, 70 μL 0.1 M ammonium acetate was added and the samples were dried in a SpeedyVac. Prior to mass spectrometry, the samples were redissolved in 10 μL 0.1% TFA.

MALDI MS PEPTIDE MAPPING OF RHO

A saturated solution of 4-hydroxy-α-cyanocinnamic acid in 0.1% trifluroacetic acid/33% acetonitrile was used as the matrix. Analyte/matrix solutions were prepared by mixing a 0.5 μL aliquot of Rho peptide mixture sample into 2-3 μL matrix solution in a 0.6 mL microcentrifuge tube.

The mass spectrometer target was prepared by depositing 0.5 μL matrix solution, incubation for 15 seconds, followed by gently wiping off the remaining liquid with a Kimwipe tissue. This created a crystalline seed-layer on the probe. The analyte/matrix solution was then deposited on top of the seed layer and allowed to dry. The crystals were rinsed by dipping the target into water. After drying of the probe by a stream of air, the sample was inserted into the mass spectrometer.

MICROBORE LC-MS PEPTIDE MAPPING OF RHO

Microbore HPLC was done on the ABI Model 140B syringe pump at a flow rate of 40 μL/min. The 1 mm x 250 mm Vydac C_{18} column was used for peptide separation. Solvents were 5 mM ammonium acetate + acetonitrile (tryptic peptides, 0% to 80% acetonitrile in 50 minutes) or 0.1% formic acid + acetonitrile (tryptic, Lys-C, and Asp-N peptides, 0% to 65% acetonitrile in 60

minutes). In cases where UV- and MS-detection were used simultaneously, 0.1% trifluoroacetic acid was used as the aqueous mobile phase.

Mass spectrometric detection was accomplished using the Sciex API III Plus triple quadropole mass spectrometer. The column eluent was transfered to the ion source by a 50 μm ID fused silica capillary at a rate of 40 μL/min. Electrospray ionization was performed using a needle potential of 5,000 V, interface potential of 650 V and orifice potential of 110 V. The nebulizer gas (oxygen) flow rate was 0.6 L/min and the curtain gas (nitrogen) flow rate was set to 0.9-1 L/min.

Positive ion mass spectra were acquired by scanning the first quadropole (Q1) from 200 to 2,400 Da in 4 seconds.

Synthetic peptide-nucleic acid hybrid

SEMIPREPARATIVE REVERSE-PHASE HPLC PURIFICATION OF PEPTIDE PRIOR TO SYNTHESIS OF HYBRID

The peptide was synthesized by Professor Jane Aldrich and colleagues using a solid-phase synthesis protocol.

The peptide was purified by RP-HPLC using the Vydac C_{18} column and 0.1% trifluoroacetic acid and acetonitrile as solvents. The gradient used for peptide-purification was programmed as follows: 5%-35% B in 10 min.; 35%-65% B in 30 min.; 60% to 90% B in 5 min. A flowrate of 1 mL/min was used. The peptide was detected by UV-monitoring at 220 nm.

SYNTHESIS OF LINKER-OLIGONUCLEOTIDES

The linker-oligonucleotides were a gift from Oligos Etc., Inc. (Wilsonville, Oregon). An amine linker chain {-$(CH_2)_6$-NH_2} was attached to the 5' end of the oligothymidylic acid by a phosphodiester linkage.

SYNTHESIS OF COVALENT OLIGONUCLEOTIDE-PEPTIDE HYBRIDS.

Approximately 800 μg of oligonucleotide was used in each experiment. All other reagents were added in 2-3 times molar excess over nucleotide. The reactions were performed at

room temperature in 200 μL dimethylformamide (DMF). The DMF stock solution was dried by sparging with N_2 for 2 hours. Diisopropylethylamine (DIPEA) was prepared as a 10% solution in DMF. Benzotriazolyloxy-tris(pyrrolidino]-phosphonium hexafluorophosphate (PyBOP) was prepared as a 100 g/L solution in DMF and hydroxybenzotriazole (HOBT) as a 50 g/L solution in DMF.

Coupling of peptide to the dT_6 oligonucleotide was accomplished as follows. Dry peptide (1.57 mg) was dissolved in 100 μL DMF in an 1.5 mL microcentrifuge tube and sonicated briefly. 8.5 μL of 10% DIPEA was added followed by 11.8 μL of PyBOP solution and 7 μL of HOBT solution. The peptide solution was incubated for 10 minutes. This solution was then added to another microcentrifuge vial containing 820 μg of dT_6 oligonucleotide dissolved in 73 μL DMF. The reaction mixture was vortexed regularly over the incubation period of 4 hours to 3 days at room temperature. Aliquots were removed and analyzed by RP-HPLC and by MALDI mass spectrometry in order to monitor the time course of the reaction.

The coupling of peptide to the dT_{10} oligonucleotide was accomplished by an analogous method.

MICROBORE HPLC MONITORING OF REACTION MIXTURES OF PEPTIDE-NUCLEIC ACID HYBRIDS

The coupling reactions between peptide and nucleic acid was monitored by microbore HPLC. The solvent system consisted of 5 mM ammonium acetate and acetonitrile. The components of the reaction mixture were separated using a gradient from 5% to 90% acetonitrile over 45 minutes at a flowrate of 40 μL/min. Eluting compounds were collected, dried, and analyzed by MALDI mass spectrometry.

HPLC PURIFICATION OF COVALENT PEPTIDE-NUCLEIC ACID HYBRID

The peptide was purified using an aquous solution of 0.1% trifluoroacetic acid and 0.08 % TFA in acetonitrile as solvents. A gradient from 5% to 90% acetonitrile over 45 minutes at a flowrate of 1 mL/min was used. Purification of the hybrid was accomplished using solvents and a 1 mL/min gradient identical to those described for the microbore system.

MALDI MASS SPECTROMETRY OF PEPTIDE-NUCLEIC ACID HYBRIDS

Peptide was dissolved in 0.1% trifluroacetic acid and mixed with matrix solution (saturated solution of HCCA in 0.1% TFA /acetonitrile (2:1)).

Aliquots from reaction mixtures were diluted in water (1:99) and then mixed 1:4 with matrix solution (saturated solution of THAP in 50 mM diammonium citrate/50% acetonitrile).

Linker-oligo(dT)'s and HPLC purified hybrids were redissolved in water + acetonitrile (1:1) and an aliquot mixed with THAP matrix as described above.

0.5-1 μL of matrix/analyte solution was deposited onto the probe and allowed to dry.

ESI-MS AND -MS/MS OF SYNTHETIC PEPTIDE

A 50 μM solution of HPLC purified peptide in 50% methanol/1% formic acid was prepared. Flow injection at 3.3 μL/min was accomplished by using a Rheodyne 8125 injector with a 5 μL loop connected to a Harvard syringe pump. Nebulizer gas (oxygen) flow and countercurrent gas (nitrogen) flow were 0.6 L/min. The electrospray voltage was 5,000 V and the interface voltage was 650 V. The optimal orifice voltage was 60 volts. For collisional excitation an Argon pressure of approximately 130 (laboratory frame) was used to induce fragmentation of the doubly charged peptide ion at m/z 621. This is about half the pressure normally used for CID of tryptic peptides.

ESI-MS AND LC-ESI-MS OF PEPTIDE-NUCLEIC ACID HYBRIDS

HPLC purified hybrid samples were dried in a speedvac, redissolved in 5 mM ammonium acetate and acetonitrile (1:1), and subsequently analyzed by negative ESI MS by flow injection using a 1:1 mixture of 5 mM ammonium acetate and acetonitrile as solvent. The electrospray potential was -3,800 V to -4,100 V and the orifice potential was -110 V. Microbore HPLC-MS analyses of reaction mixtures were performed under similar ion source conditions at a liquid flow rate of 40 μL/min.

ESI-MS/MS OF PEPTIDE-NUCLEIC ACID HYBRIDS

Tandem mass spectrometry of peptide-linker-dT_6 hybrid in the positive ion mode was performed under the same conditions as peptide fragmentation, except that the argon gas pressure in the collision cell was increased to 150-180 (laboratory frame collision gas thickness)

In the negative ion mode, MS/MS experiments were performed using an electrospray potential of -4,100 V and an orifice potential of -120 V. The collision gas pressure was set at 150.

Chapter 6: Results

The following sections describe the experimental results obtained by mass spectrometric analysis of photochemically crosslinked protein-nucleic acid complexes. Starting by using samples of UV-irradiated protein/nucleic acid mixtures, the presentation of experimental data follows the strategy outlined in Chapter 3 (Figure 7) from the top down. The first part of this chapter describes the data obtained by MALDI mass spectrometric analysis of intact complexes of gp32 × dT_{20}, Ung × dT_{20}, and Rho × 4sUDP. Then a description of attempts to directly analyze proteolytic peptide mixtures derived from the crosslinked complexes by MALDI mass spectrometry is given, followed by results obtained by MALDI analysis of synthetic peptide-linker–$dT_{6/10}$ hybrids. The next section describes the analysis of anion-exchange purified Ung peptide × dT_{20} complexes. The last sections describe how MALDI and HPLC-ESI-MS was used to characterize proteolytic digest mixtures of Rho and Rho × 4sUDP in an attempt to identify peptide × 4sUDP complexes.

Analysis of intact, UV-crosslinked protein–dT_{20} complexes

Our initial experiments concentrated on demonstrating that intact, UV-crosslinked protein–DNA complexes could be detected by MALDI mass spectrometry. This work is described in the following section. Proteins that bind to single-stranded DNA (gp32 and Ung) were UV-crosslinked to dT_{20} and analyzed by MALDI mass spectrometry.

MALDI MASS ANALYSIS OF GP32 UV-CROSSLINKED TO dT_{20}

Laser generated pulsed UV crosslinking has been used extensively to characterize the DNA-binding properties of gp32 (49, 58, 59). We used gp32 crosslinked to dT_{20} by pulsed UV irradiation for our first mass spectrometric studies. The UV laser irradiated reaction mixtures of gp32 and dT_{20} were analyzed by SDS-polyacrylamide gel electrophoresis and MALDI mass spectrometry without any prior cleanup of the samples. The crosslinking yields were estimated by

gel scanning to be 10-16% relative to the oligonucleotide concentration (Figure 13). This corresponds to gp32 × dT_{20} covalent complex yields of 4.4-7.0% or, equivalently, to gp32 × dT_{20} concentrations of 1-2 μM in the reaction vials.

Experimenting with several matrix-solvent combinations for MALDI (summarized in Table IV), we found that previously reported matrices did not yield spectra of the gp32 × dT_{20} covalent complex. However, we discovered by trial and error that good MALDI mass spectra of the photo-induced gp32 × dT_{20} complex along with the native gp32 are produced from a system of sinapinic acid in formic acid/acetonitrile (2:1) containing 1% (v/v) glycerol after washing the glycerol off the matrix crystals with cold water (Figure 14a). About 1 time in 4, we encounter a reaction mixture from which we cannot produce a spectrum of the gp32 × dT_{20} complex with this modified sinapinic acid sample preparation procedure.

A non-irradiated mixture of gp32 and dT_{20} did not produce any signals that could correspond to a covalent or non-covalent complex (data not shown).

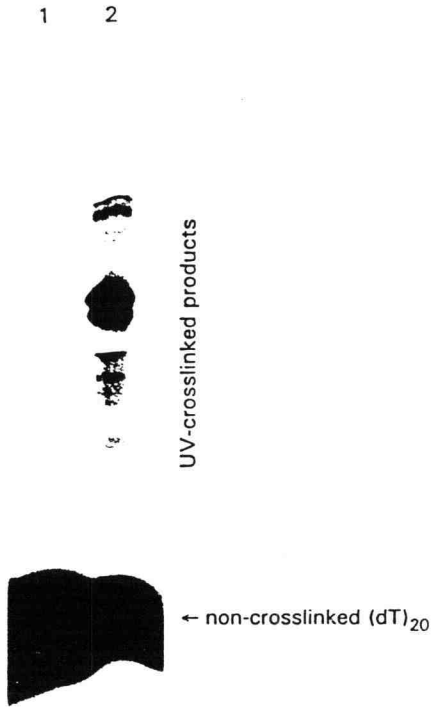


Figure 13. Autoradiogram of SDS polyacrylamide gel electrophoresis of a crosslinking reaction mixture of gp32 and dT_{20}.

A trace amount of 5'-[^{32}P]-labeled dT_{20} was added to visualize the DNA. The pulsed UV laser crosslinking procedure was as described in the Experimental section. Lane 1: An aliquot of a mixture of gp32 and dT_{20} before exposure to a UV laser pulse. Only non-crosslinked dT_{20} is visible in this lane. Lane 2: Aliquot of the mixture after exposure to a single 266 nm UV laser pulse. The higher molecular weight species in this lane result from the dT_{20} crosslinked to gp32. The crosslinking yield in this case is 12% of the DNA.

Table IV. Matrices surveyed for MALDI analysis of the UV-crosslinked gp32 × dT_{20} complex

Matrix	Solvent	Ion signals corresponding to
Sinapinic acid	0.1% TFA/acetonitrile (2:1)	-
	0.1% TFA/acetonitrile (1:1)	gp32 (weak)
	0.1% TFA/acetonitrile (1:3)	-
	Formic acid/acetonitrile (2:1)	gp32 and gp32 × dT_{20}[a]
	66% formic acid/ 33% acetonitrile/1% glycerol	gp32 and gp32 × dT_{20}[b]
	Formic acid/acetonitrile (1:1) with or without 1% glycerol	-
	Formic acid/2-propanol (1:1)	gp32
	Formic acid/1,1,1,3,3,3-hexafluoroisopropanol (1:1)	-
	1,1,1,3,3,3-hexafluoro-isopropanol	dT_{20}
4HCCA	0.1% TFA/acetonitrile (2:1)	gp32
	0.1% TFA/2-propanol (1:1)	-
	0.1% TFA/2-propanol (3:1)	-
	0.1% TFA/2-propanol (1:3)	-
Ferulic acid	0.1% TFA/acetonitrile (2:1)	-
	Water/acetonitrile (2:1)	-
Gentisic acid (2,5-DHB)	0.1% TFA/acetonitrile (2:1)	-
	Water/ethanol (1:1)	-

[a]Signals observed only if glycerol present in sample.

[b]Matrix-solvent system works well about 3 times in 4.

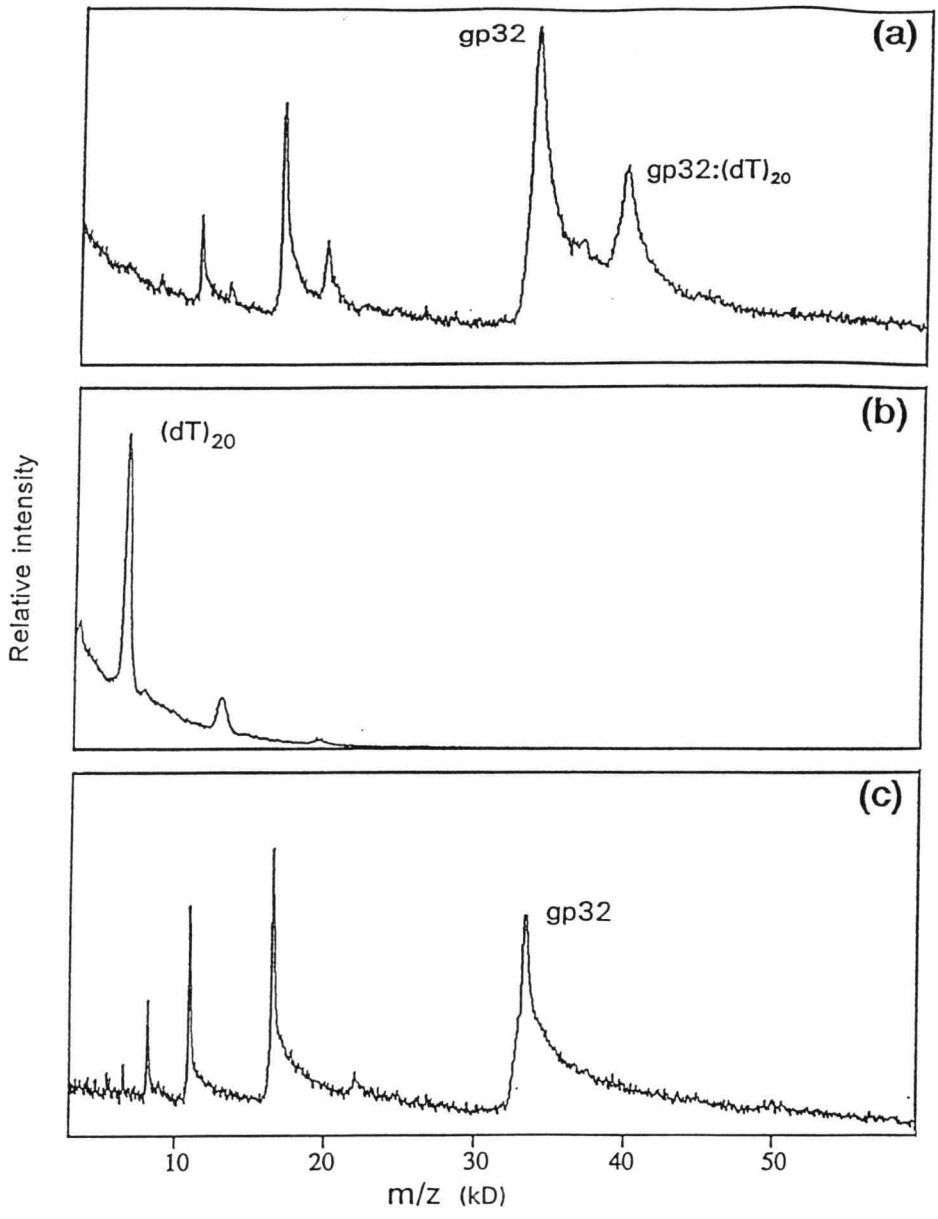


Figure 14. Positive ion MALDI mass spectra of samples taken from a mixture of gp32 and dT_{20} exposed to a single, 266 nm laser pulse.

Labeled peaks are singly charged species; peaks at lower m/z are multiply charged (e.g. +2, +3, etc.) species; and peaks at higher m/z are clusters of molecular units. (a) Matrix-solvent: sinapinic acid (10 g/L) in 66% formic acid/33% acetonitrile/1% glycerol. Non-crosslinked gp32 (calculated MW 33,502) and crosslinked gp32 x dT_{20} (calculated MW 39,603) ions are detected (m/z between the two peaks corresponds approximately to the MW of dT_{20}). (b) Matrix-solvent: sinapinic acid (10 g/L) in 1,1,1,3,3,3-hexafluoroisopropanol. Only ions corresponding to dT_{20} are detected. (c) Matrix-solvent: 4-hydroxy-α–cyanocinnamic acid (5 g/L) in 0.1% TFA/acetonitrile (2:1). Only ions corresponding to gp32 (calculated MW 33,502) are detected.

Unreacted dT_{20} oligonucleotide is not detected with the same matrix-solvent combination that reveals the gp32 × dT_{20} complex. However, by preparing the sample in sinapinic acid dissolved in 1,1,1,3,3,3-hexafluoroisopropanol, we found that the unreacted dT_{20} can be detected in the absence of any protein signals (Table IV; Figure 14b). We have also discovered two sets of sample preparation conditions under which the unreacted gp32 can be detected in the absence of any signals corresponding to dT_{20} or gp32 × dT_{20} (Table IV; Figure 14c).

The average measured molecular weight of gp32, 33,508 ± 3, agrees with the molecular weight calculated from the amino acid sequence, 33,502, to within 1 part in 5,500 (0.018%). This result is a good indication of the precision and accuracy usually obtained with pure proteins of this size on our instrument. The average experimental molecular weight of the gp32 × dT_{20} complex, 39,637 ± 69, is only slightly larger than the molecular weight (39,603) calculated on the assumption that a zero-length photochemical bond exists between the protein and the nucleotide (Table V).

MALDI MASS ANALYSIS OF UV-CROSSLINKED UNG × dT_{20}

The molecular weight of Ung was determined by MALDI MS to be 25,558 ± 6 (Table V), which, as in the case of gp32, is in excellent agreement with the calculated mass, 25,563. It was demonstrated that Ung can be photo-crosslinked to dT_{20} in a site-specific manner by continous UV irradiation for 30 minutes (3). The crosslinking efficiency was dependent on salt concentration and denatured Ung did not crosslink to dT_{20}. Presence of the Ung inhibitor (Ugi) prevented crosslinking and UV-crosslinked Ung × dT_{20} did not bind Ugi.

The purified, UV-crosslinked Ung × dT_{20} protein-nucleic acid complex was mass analyzed under experimental conditions similar to those used above for MALDI analysis of the UV-laser crosslinked gp32 × dT_{20} complex, using sample preparation conditions that employ a good protein solubilizing solvent, such as formic acid. In the present case, we used a matrix consisting of sinapinic acid dissolved in a 33% aquous solution of formic acid containing 33% acetonitrile. MALDI mass spectra exhibited signals corresponding to dT_{20} and to covalently crosslinked Ung × dT_{20} (Figure 15). A minor peptide contaminant with a molecular weight of approximately 3,500 was also detected (data not shown). The average molecular weight of the UV-crosslinked protein-DNA complex, 31,477 ± 42, is in good agreement with the MW, 31,587, obtained by adding the MW of the protein (25,563 Da) and the MW of the oligonucleotide (6,024 Da). Implicit in the calculated MW is the assumption that a "zero-length" covalent bond is formed. A relatively high laser fluence was used to obtain the spectrum, a fact which might explain the deviation of the

experimentally obtained mass from the calculated mass. However, the molecular weight is consistent with the hypothesis of a 1:1 Ung : dT_{20} ratio in the complex. It also provides direct evidence for the formation of a covalently crosslinked protein-nucleic acid complex because a non-covalent complex would dissociate during sample preparation. There is no signal that could correspond to non-crosslinked protein, indicating that the Ung × dT_{20} preparation is pure.

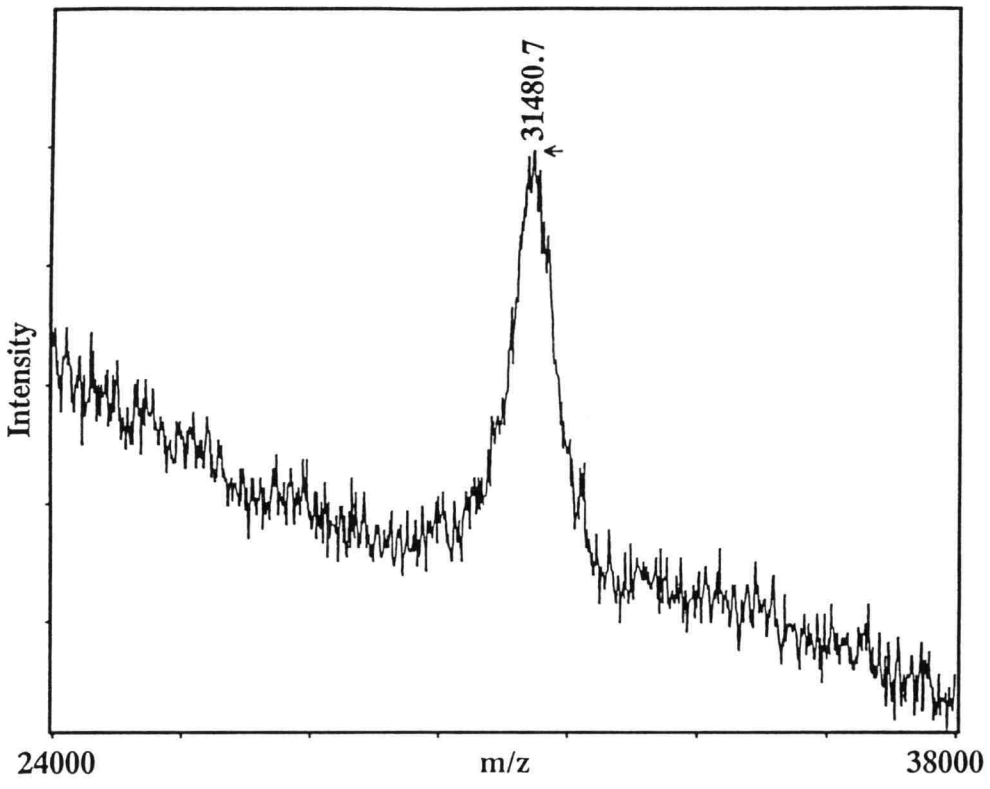


Figure 15. MALDI mass spectrum of purified UV-crosslinked Uracil-DNA glycosylase-dT_{20} complex.

The spectrum reveals a peak at m/z 31,480.7 corresponding to the UV-crosslinked protein-nucleic acid complex with a 1:1 stoichiometry (calculated MW 31,587). There are no signals corresponding to non-crosslinked Ung (MW 25,563). Sinapinic acid dissolved in 33% formic acid and 33% acetonitrile was used as the matrix. The spectrum is the sum of 30 individual laser pulses.

Table V. Molecular weight determinations of nucleic acid binding proteins and UV-crosslinked protein-nucleic acid complexes by MALDI mass spectrometry

Analyte	Molecular weight	
	Measured[a]	Calculated
gp32	33,508 ± 3	33,502[b]
gp32 × dT_{20}	39,637 ± 69	33,603
Ung	25,558 ± 6	25,563
Ung × dT_{20}	31,477 ± 42	31,587
Rho	47,004 ± 4	47,003
Rho × 4sUDP	47,560 ± 37	47,422

[a]Average ± standard error.

[b]Assuming that the four cysteine residues are in the cystine (oxidized) form.

Mass spectrometry of intact, photoaffinity-labeled Rho × 4sUDP complex

The UV-irradiated mixture of Rho and 4sUDP, which contained mainly Rho × 4sUDP protein-nucleic acid complex as measured by the ATPase inactivation assay, was successfully mass analyzed by MALDI using a matrix/solvent combination consisting of sinapinic acid dissolved in 1,1,1,3,3,3-hexafluoroisopropanol or sinapinic acid dissolved in 10% formic acid and applying the wipe/rinse technique. The MALDI mass spectrum exhibits signals corresponding to the singly, doubly, and triply charged ions of Rho × 4sUDP as well as signals corresponding to equine cytochrome c, the internal mass standard (Figure 16). The absence of any signal corresponding to non-crosslinked Rho could be consistent with a crosslinking yield close to 100%; before coming to such a conclusion however, we would require independent knowledge that there is no significant difference between the desorption/ionization efficiencies of Rho and Rho × 4sUDP under the conditions used. Also, the resolution of the instrument is not sufficient to fully resolve the two species if both were present.

The experimentally determined average of Rho's molecular weight, 47,004 ± 4, using HCCA as matrix, is in excellent agreement with the value, 47,003, calculated from the amino acid sequence (Table V). This high accuracy was obtained by using two internal mass calibrants, one with mass lower than Rho and the other with mass higher than Rho. The molecular weight of Rho × 4sUDP was determined on the average to be 47,560 ± 37 (Table V) which indicates that a modification of the protein has taken place. The difference between the measured masses of Rho × 4sUDP and Rho, 556 ± 37, is substantially larger than the molecular weight, 420, of 4sUDP.

MALDI mass spectrometric analysis of UV-irradiated Rho and of a mixture of Rho and 4sUDP resulted in molecular weights consistent with the presence of native Rho.

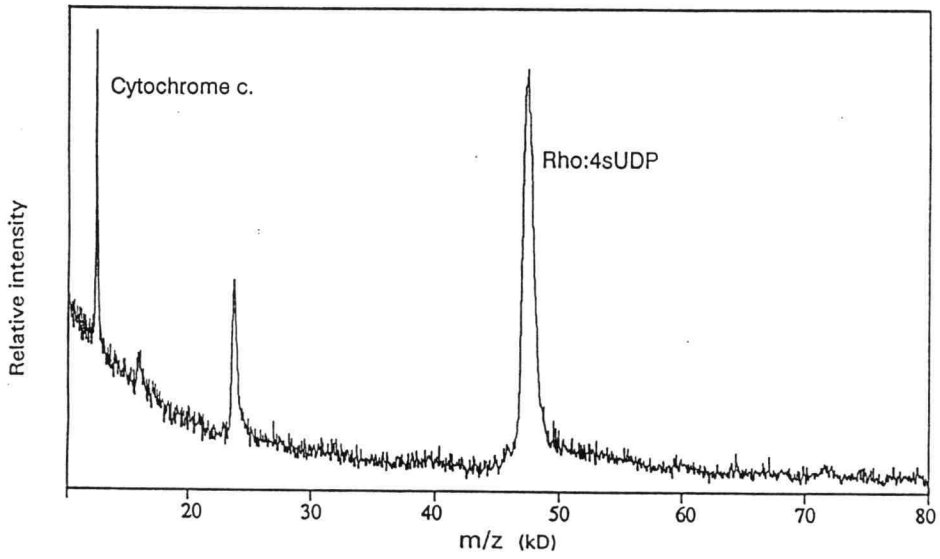


Figure 16. Positive ion MALDI mass spectrum of Rho × 4sUDP. *The molecular weight was measured to be 47,560 ± 37. The matrix/solvent combination was sinapinic acid in 1,1,1,3,3,3-hexafluoroisopropanol. Equine cytochrome c (MW 12,360.1) was used as an internal mass standard. The spectrum is the sum of signals generated by 50 laser pulses.*

Analysis by MALDI mass spectrometry of crude peptide mixtures derived from photochemically crosslinked complexes

Proteolytic digestion of UV-irradiated protein-nucleic acid mixtures produces a mixture of nucleic acids, peptides, and peptide × nucleic acid hybrids. The amount of the latter is determined by the crosslinking yield. Thus, by improving the crosslinking yield the chances of detecting the peptide × nucleic acid complex in the mixture should increase. Using the samples at hand, Rho × 4sUDP and purified Ung × dT_{20} complexes, we produced MALDI peptide maps after proteolytic digestion of the intact complexes. The results of the experiments involving Ung

and Ung × dT_{20} are presented in this section whereas the experimental data obtained by using Rho and Rho × 4sUDP is presented in a later section.

MALDI PEPTIDE MAPPING OF UNG AND UNG × dT_{20}

In order to ensure that tryptic digestion of Ung went to completion, MALDI MS was used for analysis of Ung and Ung × dT_{20} peptide mixtures obtained after 24 hours of tryptic digestion in 1 M urea. A tryptic MALDI peptide map of *S*-alkylated Ung is shown in Figure 17. The spectrum exhibits many peaks, each corresponding to a tryptic peptide. The experimentally measured molecular weigths of tryptic peptides are correlated to the predicted masses calculated from the Ung amino acid sequence, and each peak is then tentatively assigned to particular Ung peptide. The mass assignments are shown in Table VI. Almost all predicted peptides above MW 500 were detected this way, except T5, T14, T15, and the slightly acidic C-terminal peptide T20. The latter could be detected in the negative ion mode (data not shown). The tryptic peptides T6-T7 (MW 3,915) and T8-T9 (MW 2,588) were not cleaved internally, because trypsin does not usually cleave Arg-Pro or Lys-Pro amide bonds. Thus, peptides accounting for 211 of 228 amino acids, corresponding to 92% of the whole Ung amino acid sequence, could be mapped by MALDI mass spectrometric analysis as shown schematically in Figure 19.

The MALDI peptide map of trypsin treated Ung × dT_{20} was almost identical to the MALDI peptide map of non-crosslinked Ung, that is, it did not directly reveal UV-crosslinked peptide × dT_{20} complexes nor did any of the peptide signals disappear (Figure 18). Attempts to develop signals from peptide × dT_{20} complexes by using a variety of matrix/solvent combinations or negative ion mode were not successful (data not shown). Interestingly, the mass of the T6//7 peptide seem to be shifted up from 3916.4 to 3927.5 accompanied by a broadening of the peak in the UV-irradiated sample. T6/7 contains one of two Met residues that are present in Ung. The other Met residue is located in peptide T20, which is not detected in positive ion mode. Partial oxidation of the Met residues in Ung could explain the peak broadening and the mass increase, since a heterogenous population of peptide T6/7 would be present. Complete oxidation should lead to a mass increase of 16, corresponding to addition of one oxygen molecule to Met. However, the peak intensities for this peptide in the two spectra are fairly low, so care should be taken in interpreting the mass change and peak width.

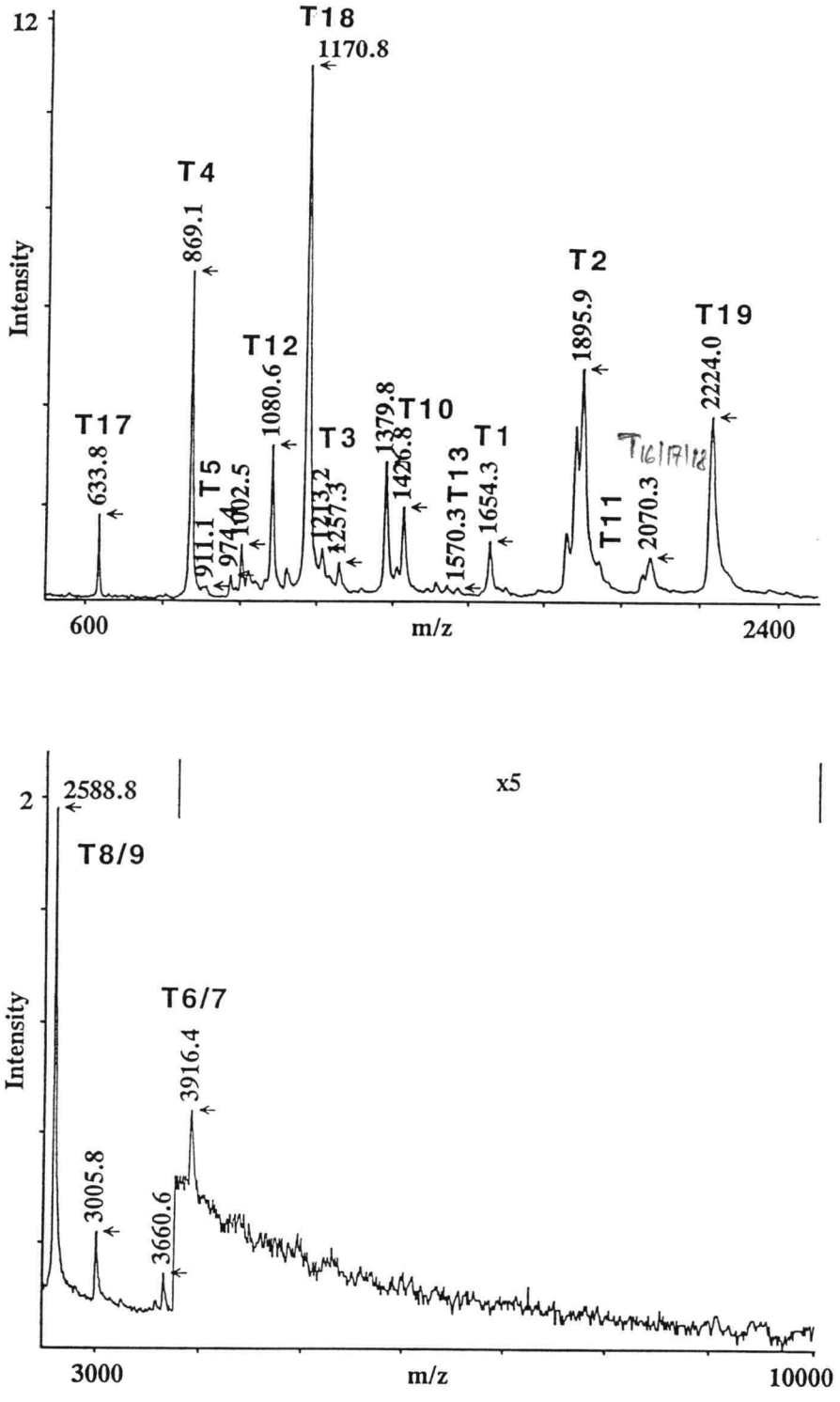


Figure 17. Positive ion MALDI tryptic peptide map of Ung.
The sample was mixed 1:7 into a saturated solution of 4-hydroxy-α–cyanocinamic acid in 0.1% trifluoroacetic acid/33% acetonitrile), deposited on the probe, dried, and rinsed in water prior to mass spectrometric analysis.. The spectrum is the sum of signals generated by 30 laser pulses.

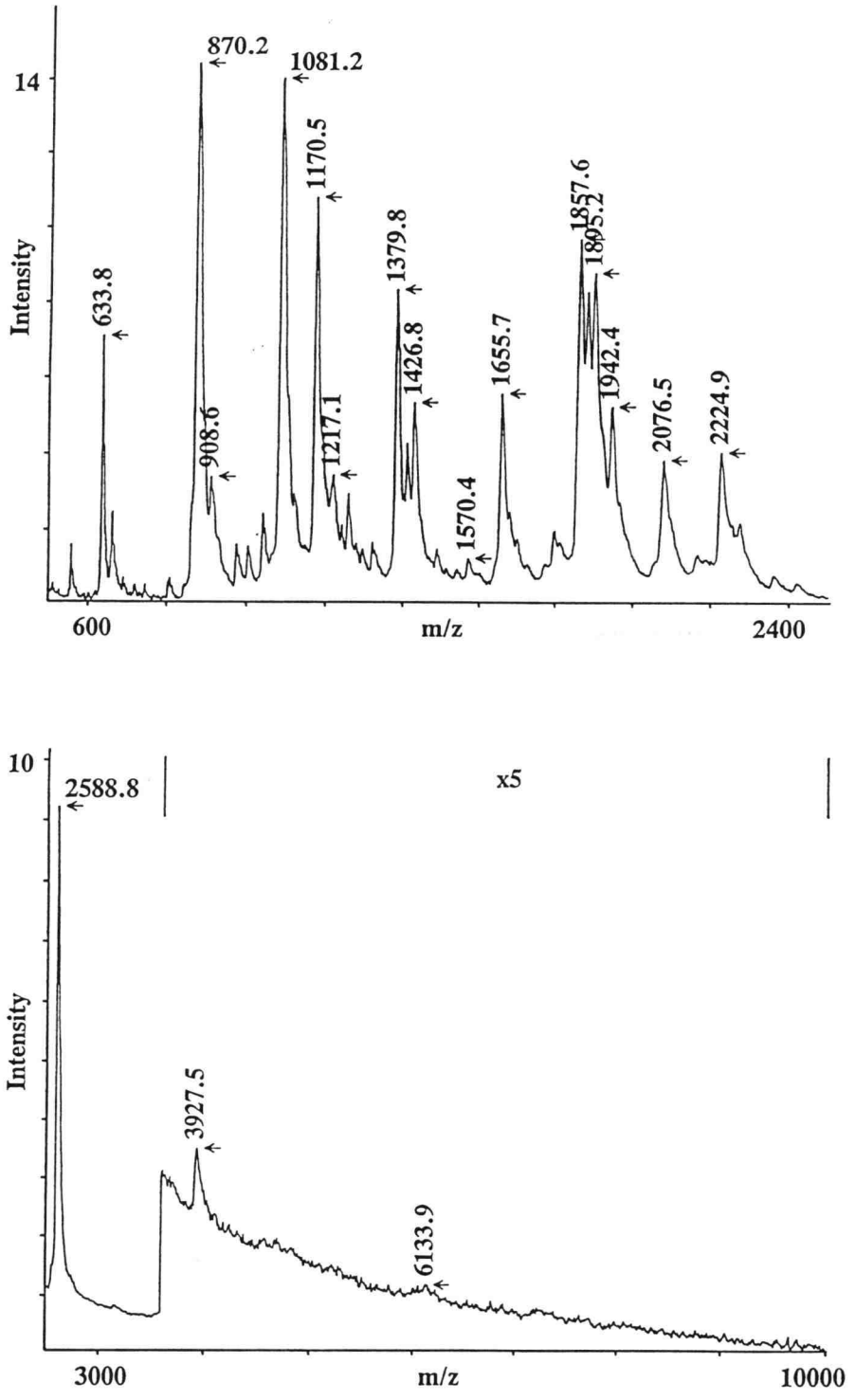


Figure 18. Positive ion MALDI tryptic peptide map of Ung × dT_{20}.
The sample was prepared and analyzed as described in the legend to the previous figure.

Table VI. Molecular weight determinations of Ung and Ung × dT_{20} tryptic peptides by MALDI mass spectrometric peptide mapping.

Ung tryptic peptide	Predicted Mass	*Experimental* MALDI peptide mapping
T1	1,655	1,655
T2	1,895	1,895
T3	1,217	1,217
T4	868	869
T5	908	908
T6	2,419	-
T6/7	3,915	3,916
T7	1,514	-
T8	1,279	-
T8/9	2,588	2,590
T9	1,329	-
T10	1,426	1,426
T11	1,943	1,943
T12	1,079	1,080
T13	1,571	1,571
T14	146	-
T15	616	-
T16	302	-
T17	633	633
T18	1,169	1,170
T18/19	3,376	-
T19	2,223	2,225
T20	1,771	1,771

MANELTWHDV LAEEK▾QQPYF LNTLQTVASE R▾QSGVTIYPP 40
T1 (1655) **T2 (1895)** **T3 (1217)**

QK▾DVFNAFR▾F TELGDVK▾VVI LGQDPYHGPG QAHGLAFSVR▾ 80
T4 (868) **T5 (908)** T6 **T6/7 (3915)**

PGIAIPPSLL NMYK▾ELENTI PGFTR▾PNHGY LESWAR▾QGVL 120
T7 T8 **T8/9 (2588)** T9

LLNTVLTVR▾A GQAHSHASLG WETFTDK▾VIS LINQHR▾EGVV 160
T10 (1426) **T11 (1943)** **T12 (1079)**

FLLWGSHAQK▾ K▾GAIIDK▾QR▾H HVLK▾APHPSP LSAHR▾GFFGC 200
T13 (1571) T14 T15 **T16/17/18 (2069)** **T17 (633)** **T18 (1169)**

NHFVLANQWL EQR▾GETPIDW MPVLPAESE 229
T19 (2223) **T20 (1771)**

Figure 19. Mapping of tryptic peptides detected by MALDI MS to the Ung amino acid sequence.
The peptides that were detected in the MALDI peptide map account for 211 of 228 (92%) of the amino acids in Ung.

Chemical synthesis and mass spectrometric analysis of covalent peptide-oligonucleotide hybrids

Using a novel approach that includes the coupling of a peptide carboxyl group (aspartic acid) to an aminolinker on a nucleotide, we have produced covalent peptide-linker-dT_6 and peptide-linker-dT_{10} hybrids. The reagents and the reaction products were characterized by RP-HPLC, MALDI mass spectrometry and ESI mass spectrometry.

A peptide **(I)** containing one free carboxyl group, in the form of an aspartic acid residue (underlined in **(I)**), was synthesized by a solid phase protocol. The N-terminal amino-group was

protected by acetylation and the C-terminal carboxyl group was protected by amidation. The calculated molecular weight of this peptide is 1240.4.

(I) CH_3CO-Gly-Phe-Phe-Gly-Ala-Asp-Arg-Phe-Val-Leu-Ala-NH_2

CHARACTERIZATION OF SYNTHETIC PEPTIDE AND AMINOLINKER-OLIGONUCLEOTIDES

Molecular weight determination by MALDI mass spectrometry and ESI mass spectrometry indicated that the desired peptide had been synthesized by virtue of the presence of a protonated molecular ion at *m/z* 1,241 (Table VII). The peptide was further characterized using ESI tandem mass spectrometry to sequence the doubly charged peptide ion (*m/z* 621) (Figure 20). The peptide was easily fragmented using only 1/2 to 1/3 of the collision gas pressure normally employed for collisional activation. It was also noted that a low orifice potential had to be used in order not to generate peptide fragments in the ion source region. This indicates that the peptide is very labile in the gas-phase. Almost all of the B and Y sequence ions are observed in the tandem mass spectrum (indicated in Figure 20) providing complementary confirmation of the amino acid sequence. These data confirm that the correct peptide had been synthesized.

The peptide was purified by RP-HPLC prior to being used in the reaction to remove low molecular weight contaminants leftover from the solid-phase synthesis reaction.

The two oligonucleotide species, NH_2-$(CH_2)_6$-5'-pdT_6 and NH_2-$(CH_2)_6$-5'-pdT_{10}, were analyzed by MALDI mass spectrometry and ESI mass spectrometry prior to being used in the coupling reaction to the synthetic peptide. The masses obtained were in agreement with the calculated molecular weights (Table VII). The spectra identified only minor contaminants so the oligonucleotides were used as they were received.

The peptide-linker-oligo(dT) hybrids were synthesized using PyBOP as a catalyst. The PyBOP activation of carboxyl groups is a recently developed approach for creating peptide amide bonds during peptide synthesis (176). In this case, an amide bond was formed by reaction of the aspartic acid sidechain carboxyl group with the amino-group of the amine-linker-oligonucleotide. The reaction is schematically shown in Figure 21.

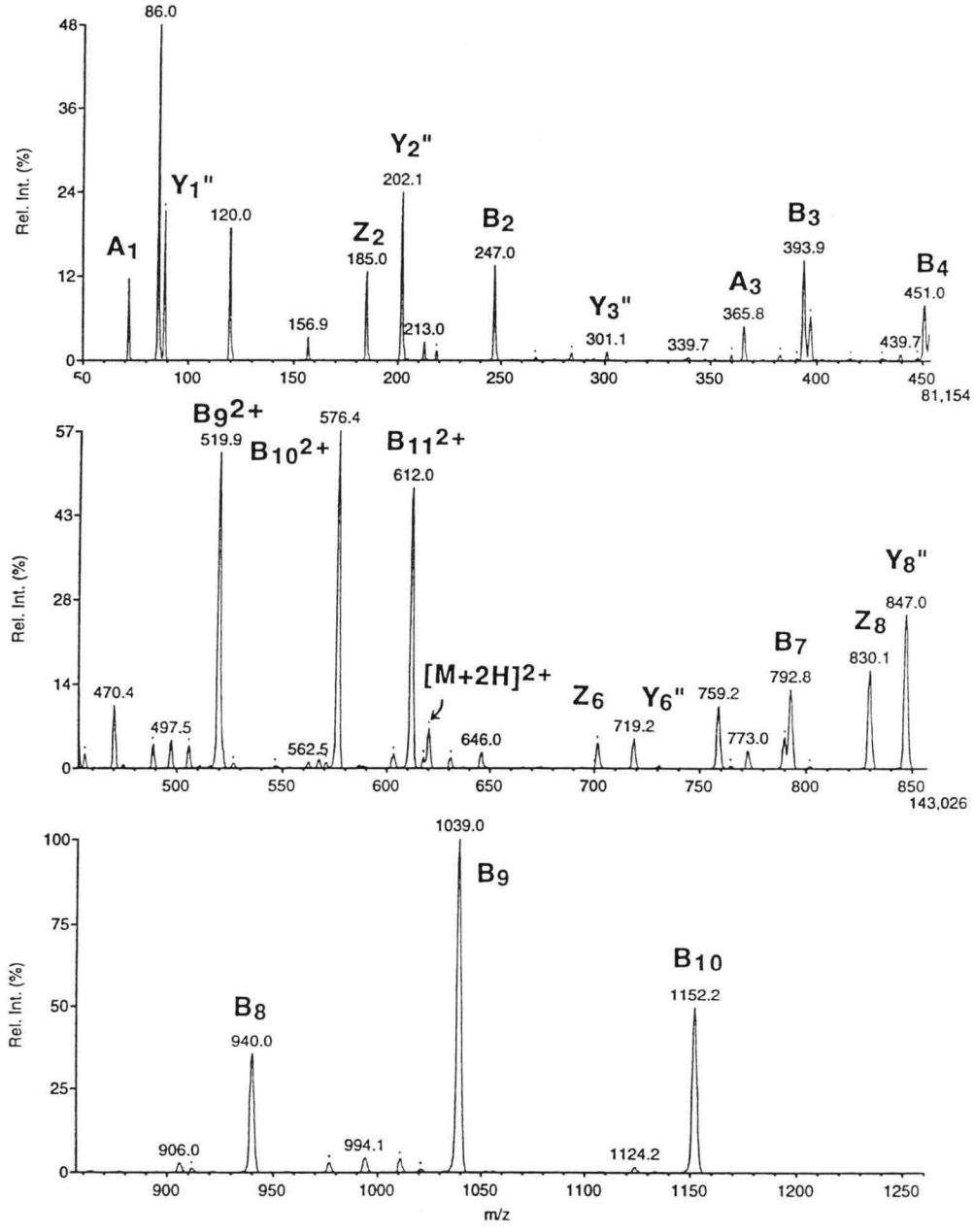


Figure 20. ESI tandem mass spectrum of the peptide (1).
The amino acid sequence and the corresponding B and Y sequence ions are indicated.

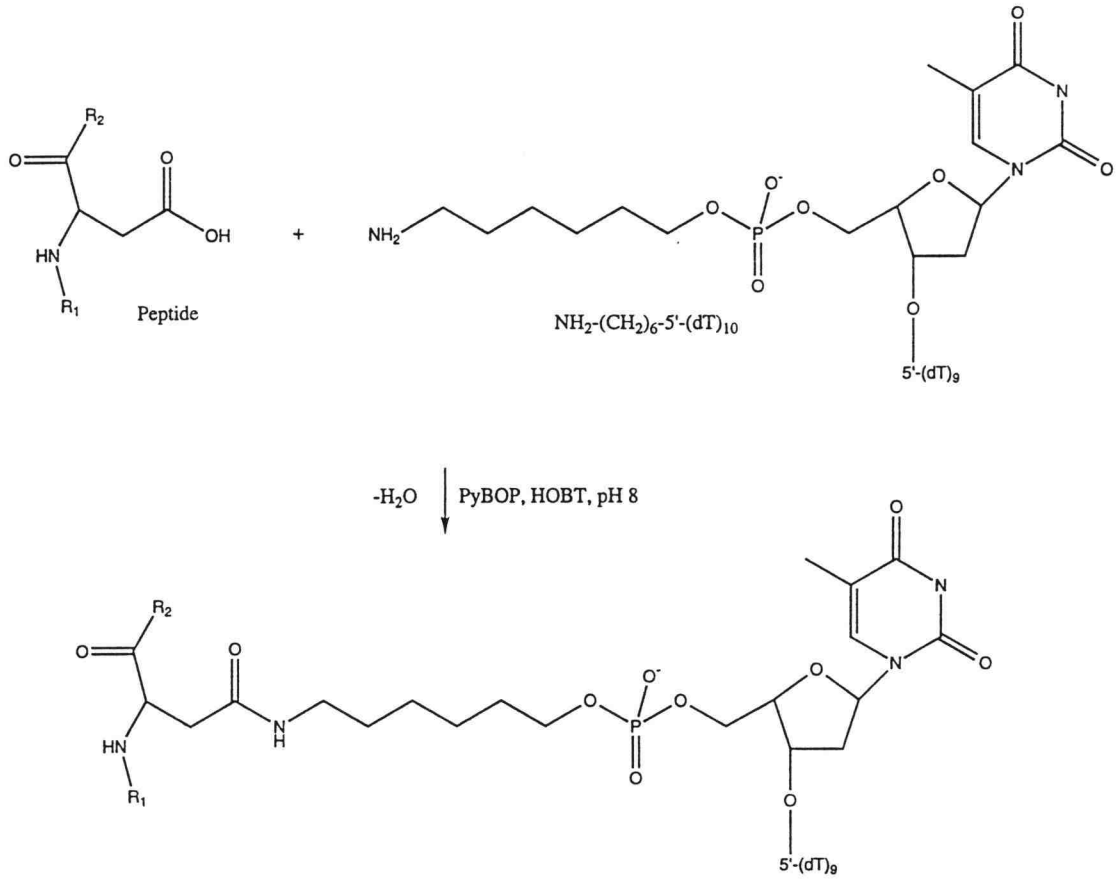


Figure 21. Reaction of peptide with aminolinker-oligothymidylic acid. *PyBOP was used to catalyze the coupling of the Asp carboxyl group to the aminolinker-oligo(dT) to produce an amide bond.*

MALDI MS OF REACTION MIXTURES AND HPLC PURIFIED HYBRIDS

The chemical reactions were monitored by MALDI mass spectrometry and RP-HPLC. A typical MALDI mass spectrum of a peptide + aminolinker-dT_{10} reaction mixture is shown in Figure 22. Signals corresponding to unreacted peptide (*m/z* 1,241.4) and unreacted aminolinker-dT_{10} (*m/z* 3,159.0) are detected along with a signal corresponding to the hybrid at *m/z* 4,381.3. The identity of the complex as a covalent hybrid is confirmed by the fact that the mass of this species is

the sum of the mass of the peptide and the mass of the nucleotide less the mass of a water molecule lost during the reaction. There is no signal that could correspond to a non-covalent complex. Thus, MALDI mass spectrometry, although not quantitative, provides a quick, sensitive, and straightforward method to monitor for the presence of the covalent peptide-linker-oligonucleotide product.

In order to obtain quantitative data microbore RP-HPLC was used to monitor the coupling reactions. Oligonucleotides and hybrids were detected by UV at 256 nm wavelength. The HPLC chromatogram obtained from a peptide + aminolinker-dT_6 reaction mixture (Figure 23) exhibits several peaks that were identified by comparison of the elution times of the initial reagents and, in some cases, by off-line MALDI mass spectrometric analysis of the collected fraction. A new species eluting between the oligonucleotide and the peptide, was collected, lyophilized, and analyzed by MALDI mass spectrometry (Figure 24). The molecular weight (3,162.8) of this species matches the calculated molecular weight of the expected peptide-linker-dT_6 hybrid to within one mass unit (Table VII). A shoulder on the high mass side of the peak corresponds to the potassium adducted molecular ion species.

The isolated peptide-linker-dT_{10} hybrid produced a similar spectrum dominated by a signal that corresponds to the predicted structure (Table VII).

The reaction, although not going to completion, seemed to take place within a very short time (<10 minutes) as has been reported previously for PyBOP catalyzed formation of peptide amide bonds (176). The relative intensities of nucleotide and hybrid peaks in the HPLC chromatogram did not change significantly over several days even after addition of fresh pH-buffer and catalyst.

Based on the absorption at 256 nm of the HPLC eluents we estimate that yields of hybrids were approximately 5% relative to the initial amount of oligonucleotide. The low yield could be due to the secondary structure of the peptide preventing efficient alignment with the linker-region of the oligonucleotide. Judging from the response from the microbore HPLC UV-detector, we estimate that the MALDI mass spectrum shown in Figure 24 was obtained by using only 10-20 pmoles of hybrid. Using normal bore HPLC, larger amounts of the peptide-oligonucleotide hybrids were purified and used to optimize the conditions for ESI mass spectrometric analysis of these species. We have not yet attempted further experiments to increase the yields of hybrids since the reaction produced more than enough sample for this study.

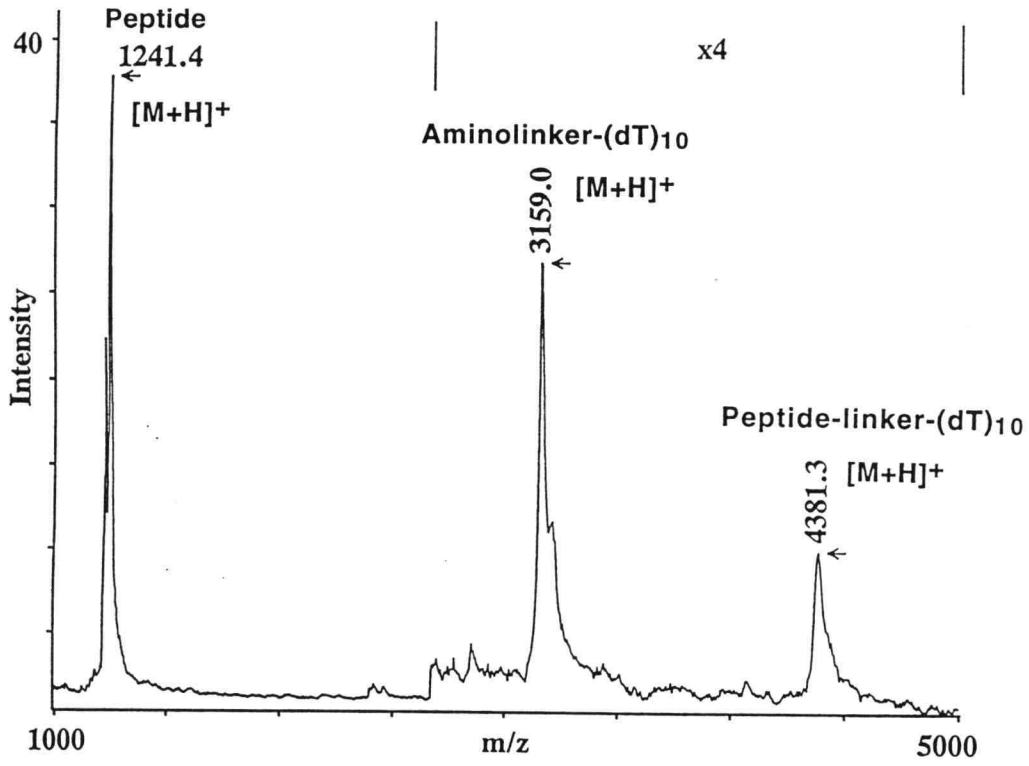


Figure 22. MALDI mass spectrum of peptide + aminolinker-dT$_{10}$ reaction mixture.

Signals corresponding to peptide (m/z 1,241.4), aminolinker-dT$_{10}$ (m/z 3,159.0), and the peptide-linker-dT$_{10}$ hybrid (m/z 4,381.6) are detected. Note the expanded Y-axis.

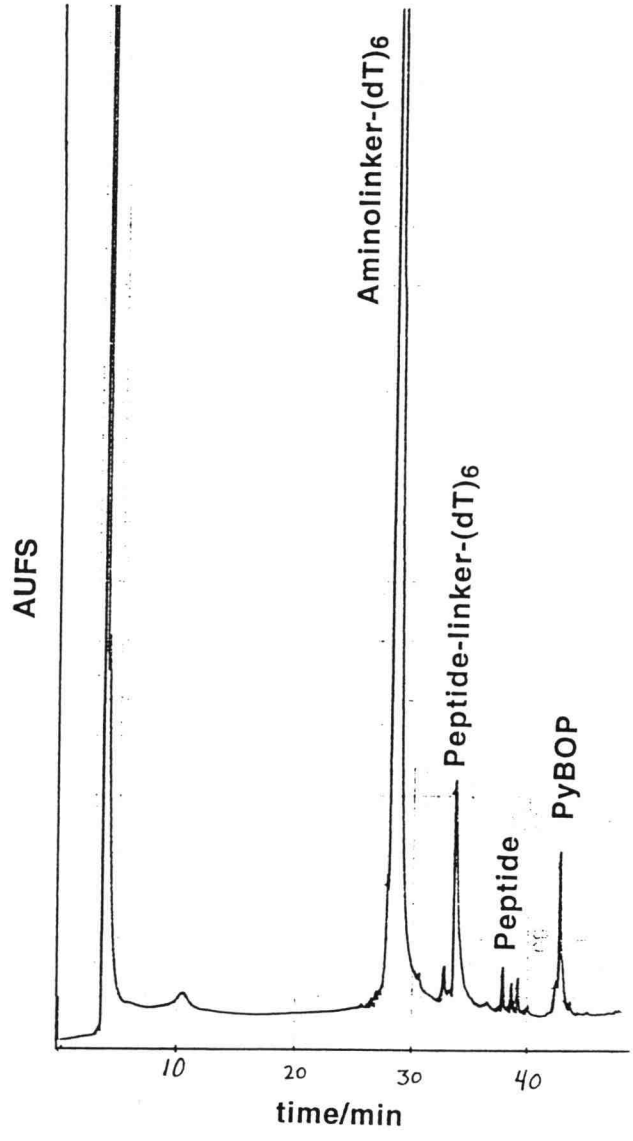


Figure 23. Reverse-phase HPLC UV chromatogram of a peptide + aminolinker-dT$_6$ reaction mixture.
Monitored by UV-absorbance at 256 nm. Peaks were identified by their elution time or by off line MALDI MS analysis. The mass spectrum of the peak indicated with an arrow is shown in the next figure.

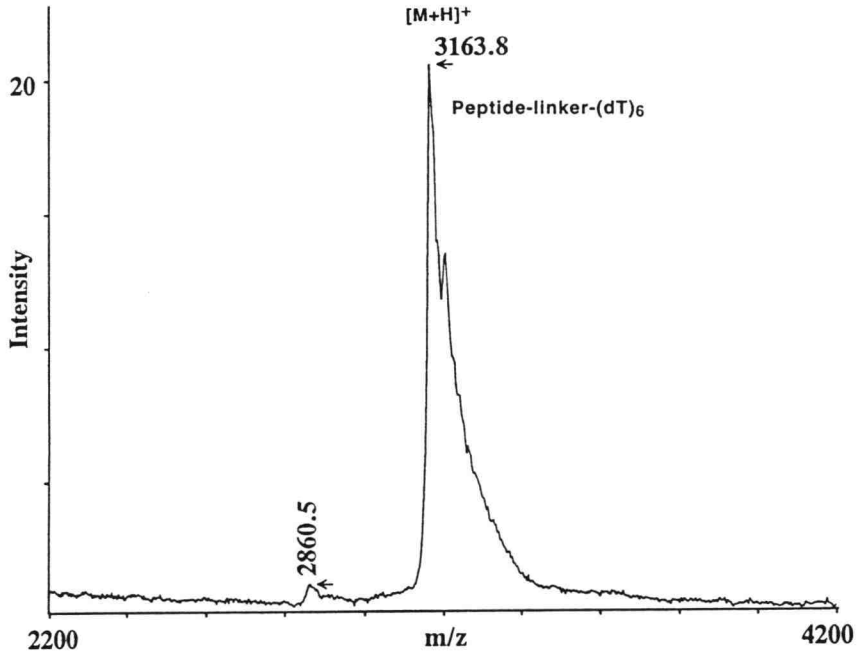


Figure 24. MALDI mass spectrum of an HPLC fraction containing peptide-linker-dT_6 hybrid.

The matrix was a saturated solution of 2,4,6-trihydroxyacetophenone in 50 mM ammonium citrate/50% acetonitrile.

Table VII. Molecular weight determinations of peptide, aminolinker-oligonucleotides, and peptide-linker-oligonucleotide hybrids by MALDI and ESI mass spectrometry.

	Molecular weight		
Analyte	**Calculated**	**MALDI**	**ESI**
Peptide	1240.4	1240.9	1240.4
Aminolinker-dT_6	1942.2	1942.4	1942.0
Aminolinker-dT_{10}	3159.2	3156.6	3158.0
Peptide-linker-dT_6	3164.6	3164.4	3164.3
Peptide-linker-dT_{10}	4381.6	4379.6	4382.0

ESI-MS OF HYBRIDS

Using negative and positive ion electrospray ionization we obtained mass spectra of the aminolinker-oligo(dT) species and the HPLC purified peptide-linker-oligo(dT) hybrid species. The masses obtained for aminolinker-dT_6 and aminolinker-dT_{10} given in Table VII are in excellent agreement with the calculated molecular weight and the data obtained by MALDI MS analysis.

Subsequently, the purified peptide-linker-oligo(dT) hybrids were analyzed by electrospray ionization MS. A negative ion ESI mass spectrum of a purified hybrid is shown in Figure 25a. The major species in this fraction is the peptide-linker-dT_{10} hybrid of MW 4,382 giving rise to doubly $[M - 2H]^{2-}$ and triply $[M - 3H]^{3-}$ charged ions at *m/z* 2,189.7 and 1,459.8, respectively. Sodium adducted ions, $[M + nNa - 2H]^{2-}$ and $[M + nNa - 3H]^{3-}$ (n=0,.., 5), are also detected. A similar spectrum was acquired in the positive ion mode (Figure 25b). Sodium adduction was more pronounced in the positive ion mode than in the negative ion mode, resulting in slightly better sensitivity in the negative ion mode. A minor species at MW 4,077 (*m/z* 1,358.2 and *m/z* 2,037.6), tentatively assigned to be the [peptide-linker-dT_9] species, was also detected by MALDI MS analysis (data not shown). The aminolinker-dT_9 species was also detected in the original aminolinker-dT_{10} sample by MALDI mass spectrometry.

ESI mass spectra of peptide-linker-dT6 in negative (Figure 26a) and positive (Figure 26b) ion mode were also acquired. The doubly and triply charged ions are detected in both polarities, with sodium adduction being more pronounced in the positive ion mode. However, the peptide-linker-dT_6 signals detected in the positive ion mode are almost three times more intense as those

detected in the negative ion mode, indicating that the peptide influences the ionization more in this hybrid than in the peptide-linker-dT_{10} hybrid. The relative contributions (peptide : oligo nucleotide) to the total mass of the hybrids are approximately 40:60 for the peptide-linker-dT_6 and 30:70 for the peptide-linker-dT_{10}.

Injection of aliquots of the crude reaction mixture revealed a more complex spectrum that exhibited ion signals corresponding to peptide and aminolinker-oligo(dT) (data not shown). In the positive ion mode only peptide ions were detected. In either case, no signals that could correspond to the hybrid were detected. This is probably due to the low amount of hybrid relative to the initial reagents.

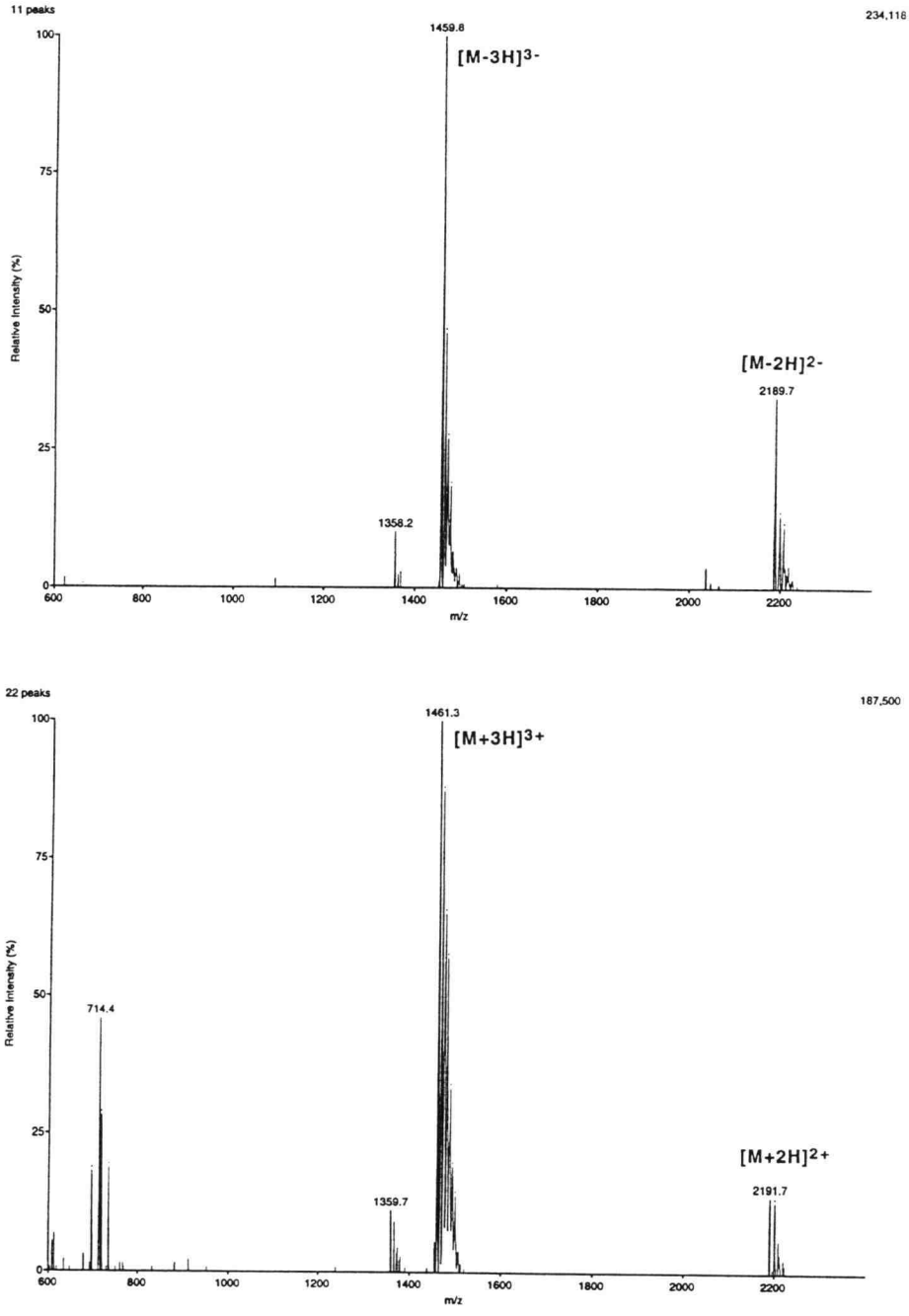


Figure 25. Negative (a) and positive (b) ion electrospray ionization mass spectra of HPLC purified peptide-linker-dT$_{10}$ hybrid.

The doubly and triply charged peptide-linker-dT$_{10}$ ions are detected in both cases. The detected ion current is similar for both polarities, but sodium adduction is more pronounced in the positive ion mode.

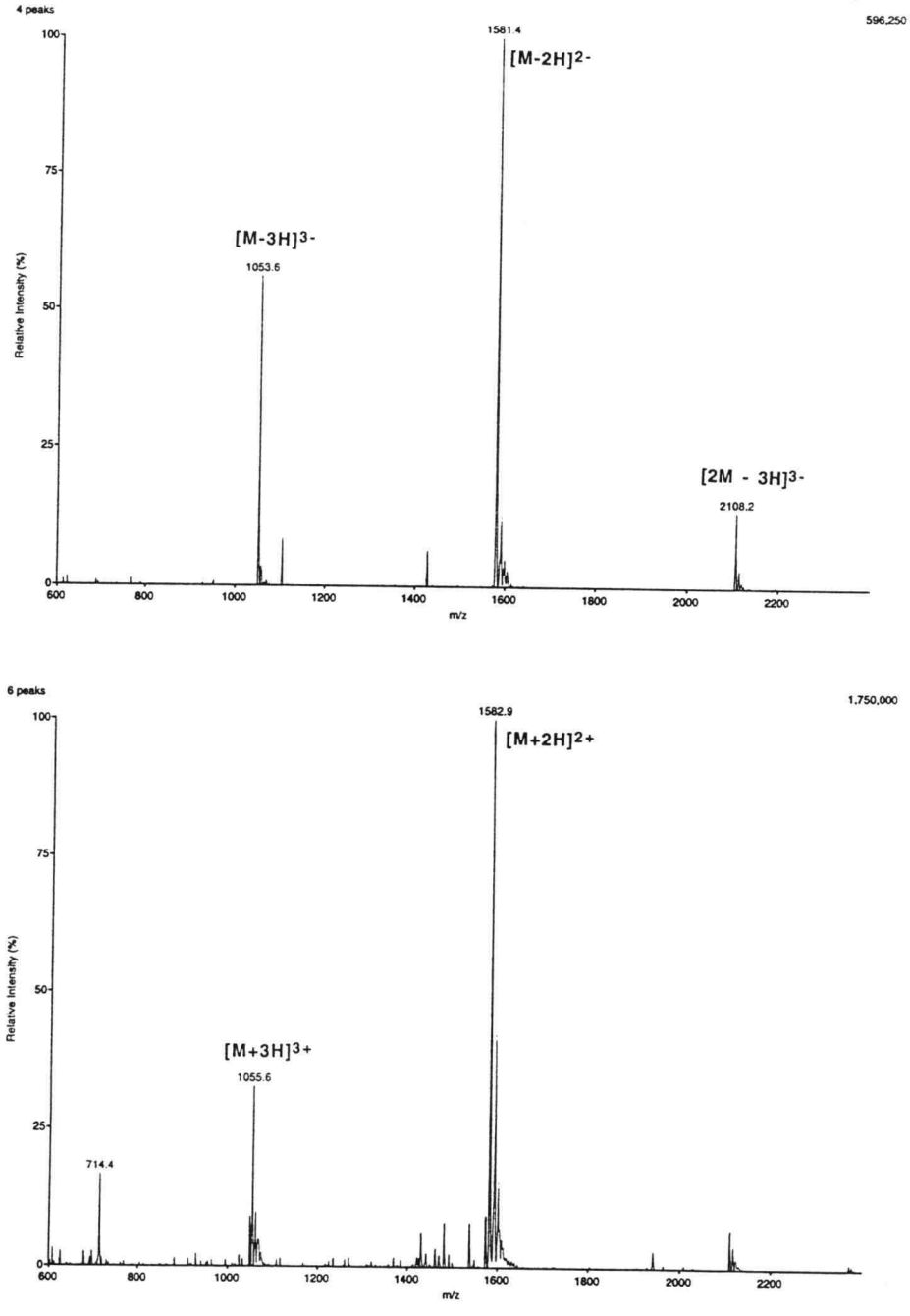


Figure 26. Negative (a) and positive (b) ion electrospray ionization mass spectra of HPLC purified peptide-linker-dT_6 hybrid.

Doubly and triply charged ions of peptide-linker-dT_6 are detected along with a triply charged dimer ion (m/z 2,108). The ion current in the positive mode is 3 fold stronger than for the negative mode.

The molecular weights obtained by ESI-MS of linker-oligo(dT)'s and peptide-linker-oligo(dT) hybrids are summarized in Table VII. In all cases, the obtained masses agrees with the calculated masses to within 0.05%. The sensitivity of ESI MS for these species is at least 8-10 times lower than the sensitivity of MALDI MS as estimated from the amounts of sample consumed in these studies. However, the resolution obtained by ESI is better than by MALDI.

Having investigated the experimental conditions necessary for ESI MS analysis of peptide-oligonucleotide hybrids, we tested to see if online monitoring of HPLC eluents by ESI MS was possible. In this experiment, the UV-detector was substituted with the ESI mass spectrometer. A total ion current (TIC) chromatogram of a negative ion LC-MS analysis of the peptide + aminolinker-dT_6 reaction mixture is shown in Figure 27. The TIC chromatogram appears similar to the UV chromatogram (Figure 23). The main difference between the two chromatograms is in the relative intensities of the chromatographic peaks.

In this case peptide detection by ESI MS is more efficient than UV-detection. This is due to the low absorbance of peptides at 256 nm wavelength normally used for sensitive UV-detection of oligonucleotides. Conversely, oligonucleotide UV-detection at 256 nm is more sensitive than ESI MS detection of these species, especially as the length of oligonucleotide increases. Using both detection methods allows both sensitivity and specificity (identification) to be obtained.

Interestingly, the intensities of the amino-linker-dT_6 signal and the peptide-linker-dT_6 hybrid signal are similar in the TIC chromatogram, while the former is 20 fold more intense in the UV chromatogram. Assuming that the UV response is quantitative this indicates that the ionization efficiency (and thereby the sensitivity of the mass spectrometer) of the hybrid is higher than for pure oligonucleotides. This feature is probably facilitated by the conjugated peptide. However, this effect is not observed for the peptide-dT_{10} hybrid, indicating that this effect is dependent on the relative contributions of peptide and oligonucleotide to the total mass of the hybrid.

The mass spectrum of the compound eluting at 26 minutes is shown in Figure 28. The mass (3,165.0) calculated from the doubly and triply charged ions corresponds to the expected molecular weight of the peptide-linker-dT_6 hybrid (Table VII). A signal from a triply charged ion of the non-covalent dimer appears at *m/z* 2,108.7. A minor contaminant of unknown origin produced a signal at *m/z* 1,106.6. The spectra obtained from the other eluting compounds were similar in quality.

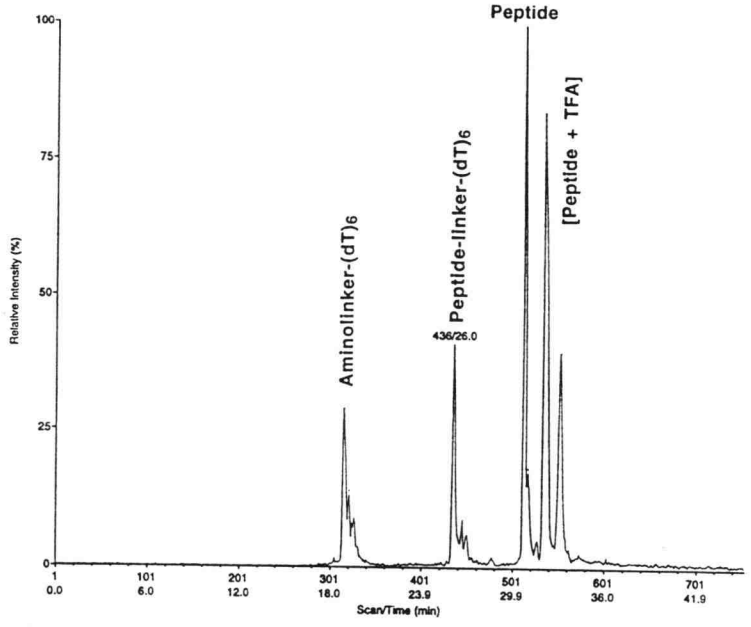


Figure 27. LC-MS chromatogram of a reaction mixture containing peptide and aminolinker-dT$_6$.
The elution profile looks similar to the UV-trace (Figure 28), but the relative response of each species is different.

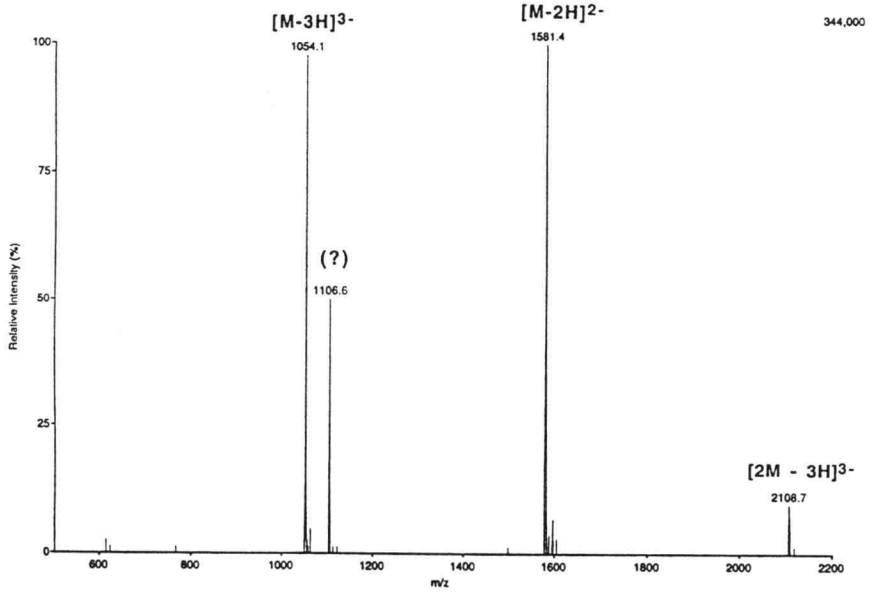


Figure 28. Extracted spectrum of peptide-linker-dT$_6$ peak from LC-MS analysis.
The negative ion spectrum identifies this species to be peptide-linker-dT$_6$.

The LC-MS chromatogram of the reaction mixture containing peptide-linker-dT_{10} hybrid did not produce satisfactory ion signals under the conditions that was used for the peptide-linker-dT_6 hybrid. Attempts to improve the sensitivity for this species by optimizing the electrospray conditions were not succesful. Apparently, the length of the oligonucleotide influences the ionization efficiency in microbore HPLC-MS analysis. This sensitivity problem in LC-MS analysis of longer oligonucleotides has also been observed for pure dT_{20}.

TANDEM MASS SPECTROMETRY OF PEPTIDE-LINKER-dT_6 HYBRID

Having a well characterized peptide-oligonucleotide hybrid we wanted to investigate the behaviour of this species in tandem mass spectrometry experiments. As mentioned above, the peptide itself was found to fragment easily under MS/MS conditions. The aminolinker-dT6 oligonucleotide was analyzed by positive and negative ion MS/MS using the doubly charged parent ion. The resulting negative ion spectrum is shown in Figure 29. The spectrum is dominated by the parent ion (*m/z* 969.5) and mononucleotide ions $[TMP\text{-}H_2O\text{-}H]^-$ and $[TMP\text{-}H]^-$ at *m/z* 303.2 and *m/z* 321.2, respectively. Signals from PO_3^- (*m/z* 79), thymine ion $[T]^-$ (*m/z* 125), $[TMP - T - H_2O]^-$ (*m/z* 177.1) and $[TMP - T]^-$ (*m/z* 195.1) are also present. A weak series of peaks corresponding to $[dT_n]^-$ and $[dT_n - H_2O]^-$ (n = 2,.., 5) are also detected.

Both positive and negative ion tandem mass spectra of the peptide-linker-dT_6 hybrid were acquired. The positive ion tandem mass spectrum of the doubly charged peptide-linker-dT_6 hybrid (*m/z* 1,583) is more complex (Figure 30) than the tandem mass spectrum of the aminolinker-oligonucleotide.

The spectrum shares some features with the tandem mass spectrum of aminolinker-dT_6 which aids the interpretation. Many peaks are 304 mass units apart which corresponds to one unit of the mononucleotide dT. This allows us to sequence the oligonucleotide that has been adducted to the peptide. An example is the $[dT_n - H_2O + H]^+$ nucleotide ion series at *m/z* 305.2, 608.8, and 912.5 (n = 1, 2, 3), which could correspond to fragmentation from the 3' end of the oligonucleotide. The $[dT_n + H]^+$ ion series takes over for n = 3, 4, 5, 6 at *m/z* 930.5, 1,234.7, 1,538.3, and 1842.0. The ions corresponding to peptide-linker-5'-oligo(dT) fragments can also be found. A $[(\text{peptide-linker-}dT_n) + H]^+$ ion series (n = 1, 2, 3) appear at *m/z* 1,625.4, 1,930.1, and 2,234.2. The loss of thymine base is evident from the ion series $[(\text{peptide-linker-}dT_n) - T + H]^+$, (n = 1, 2, 3) at *m/z* 1,499.8, 1,804.0, and 2,108.2. Loss of both the base and water is demonstrated by the ion series $[(\text{peptide-linker-}dT_n) - T - H_2O + H]^+$ (n = 1, 2, 3) at *m/z* 1,481.8, 1,787.0, and 2,091.0.

The negative ion tandem mass spectrum of the doubly charged parent ion (m/z 1,581) exhibited many of the species that were detected in the positive ion spectrum (data not shown).

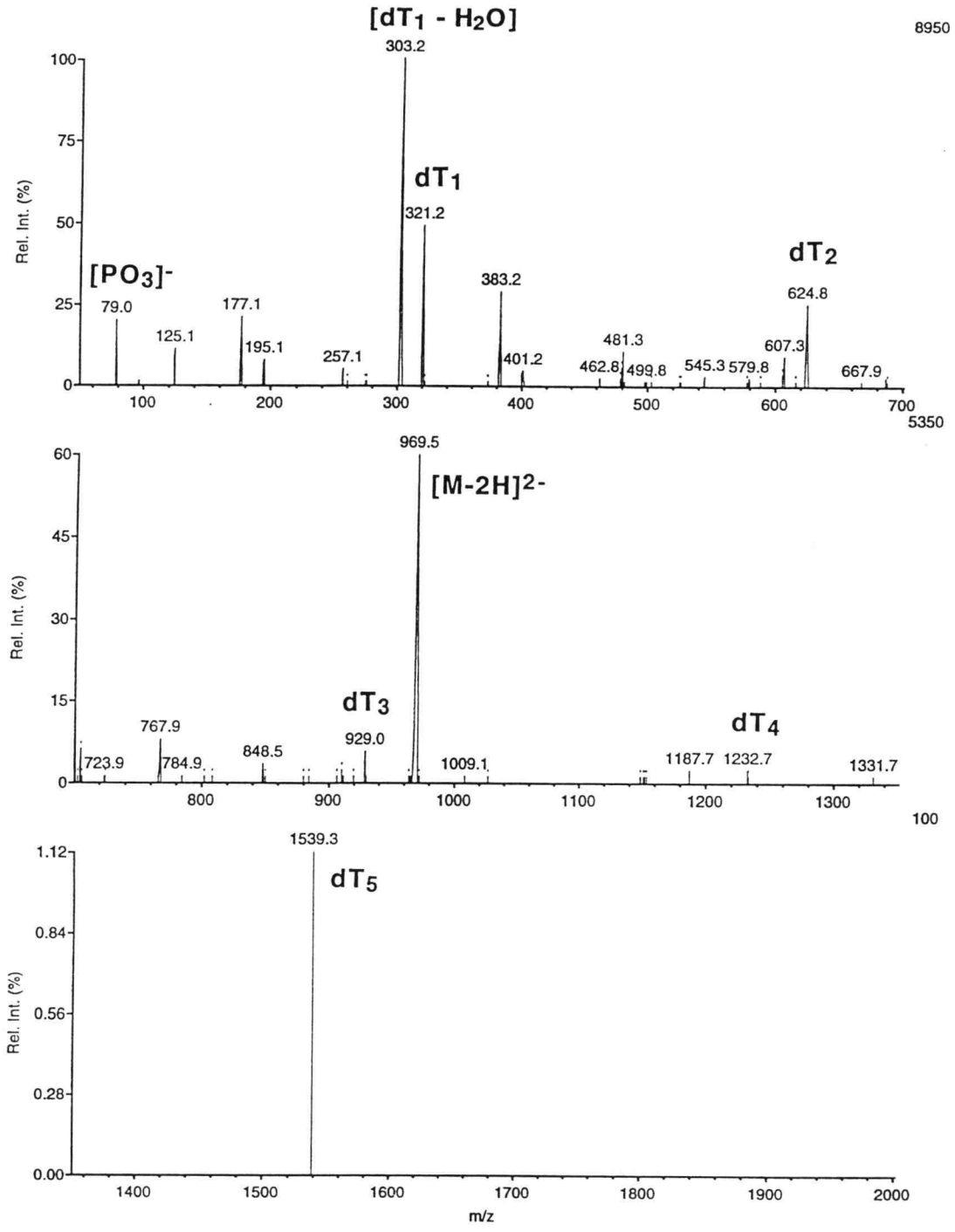


Figure 29. Negative ion ESI tandem mass spectrum of aminolinker-dT$_6$. *Ion series corresponding to dT$_n$ (n=1,..,6) are indicated.*

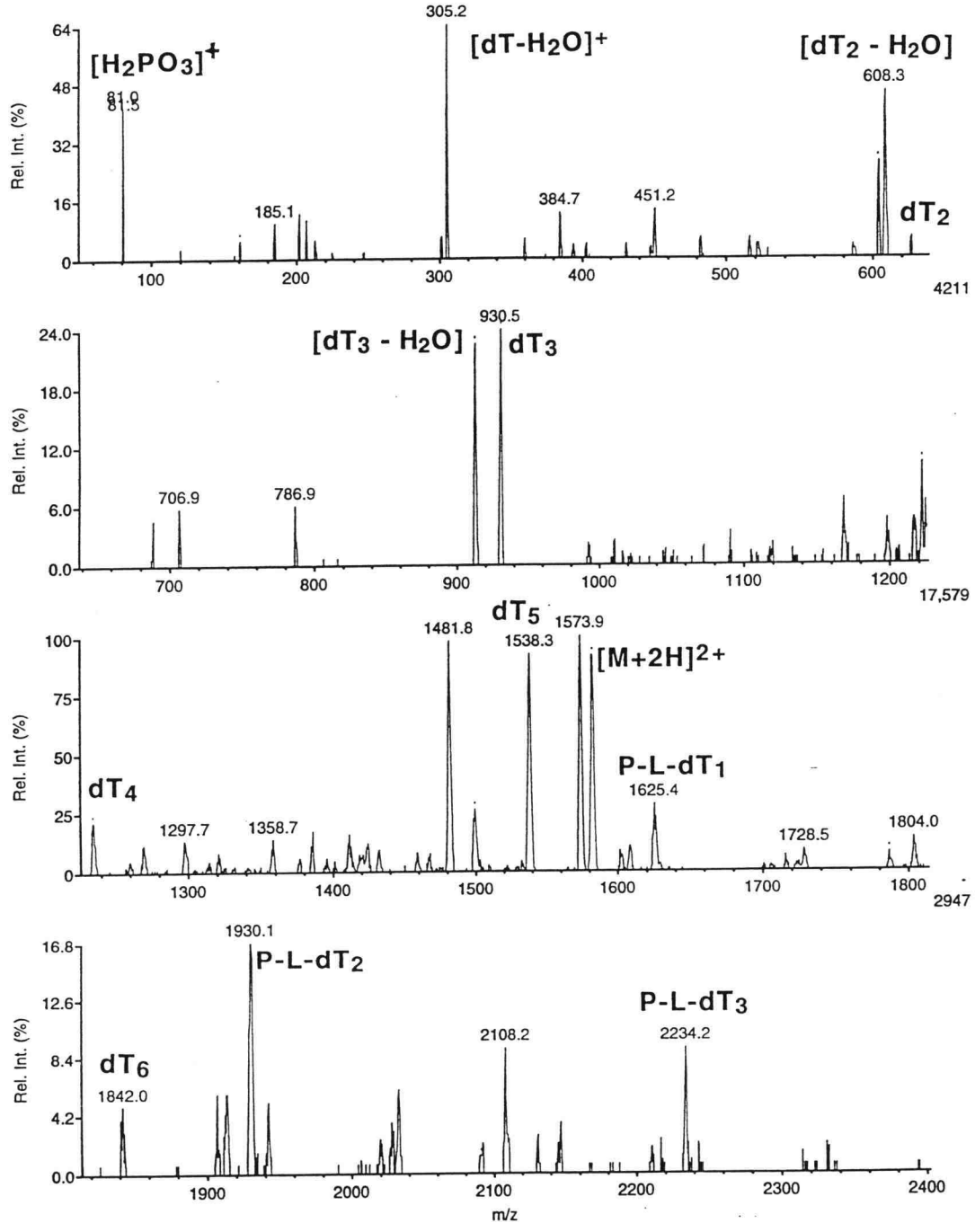


Figure 30. Positive ion ESI tandem mass spectrum of peptide-linker-dT$_6$ hybrid.

Ions corresponding to dT$_n$ and peptide-linker-dT$_n$ are labeled.

Mass spectrometric analysis of purified peptide × dT_{20} complexes derived from Ung × dT_{20}

An immediate application of the above investigations were to analyze by mass spectrometry a mixture of peptide × dT_{20} hybrids that were isolated by anion-exchange chromatography after tryptic digestion of UV-crosslinked Uracil-DNA glycosylase × dT_{20} (3). Representative fractions, as judged from an autoradiogram after tricine-gel electrophoresis, were desalted and subjected to mass spectrometric analysis. Due to limited amounts of sample and the size of the oligonucleotide dT_{20}, MALDI MS was used to characterize these hybrids. Several matrix/solvent combinations were surveyed to optimize the signal from peptide × dT_{20} hybrids (Table VIII). Matrices that have been reported to work well for oligonucleotides, especially oligo(dT)'s, produced the best signals. This is to be expected since the oligonucleotide dT_{20} contributes 70-90% of the mass of these hybrids.

MALDI MS analysis of UV-irradiated dT_{20} did not detect any photoinduced damage that resulted in molecular weight change or fragmentation. MALDI mass analyses of tryptic peptide × dT_{20} complexes were performed using a variety of matrices. The best results were obtained using sample preparation conditions that tend to favor MALDI analysis of dT_{20}, that is, by using either ferulic acid (Dr. Klaus Schneider, Rockefeller University, personal communication) or 2,4,6-trihydroxyacetophenone (100) as matrix and diammonium-hydrogen citrate and acetonitrile as solvent.

A positive ion MALDI mass spectrum of a mixture of peptide × dT_{20} hybrids is shown in Figure 31. Quantitation by radioactivity, using [^{32}P]-labeled dT_{20}, indicate that approximately 5 pmoles of sample – about one fifth of the total amount of sample available – was used to obtain the spectrum. The spectrum exhibits signals from dT_{20} at *m/z* 6,025 and some peptide × dT_{20} hybrids at higher molecular weight. Although the quality (resolution and intensity) of the spectra obtained from these UV-crosslinked samples is not as good as that achieved for the purified synthetic hybrids, shown above, the molecular weight information obtained this way is very useful (Table IX). The reduced quality might be due to the longer oligonucleotide and to the presence of multiple components in the sample. Purification of a peptide × dT_{20} hybrid by RP-HPLC followed by MALDI MS analysis did indeed provide a spectrum that was superior in quality to those obtained by direct analysis of peptide × dT_{20} mixtures (Figure 32). However, sample loss on the HPLC column was substantial so we were not able to use this approach for all peptide × dT_{20} hybrids isolated from Ung × dT_{20}.

Negative ion MALDI mass spectrometry was not found to improve the sensitivity or the quality of the spectra produced by our MALDI instrument (data not shown).

By subtracting the mass of dT_{20} (MW 6,025) from the experimentally obtained masses of the crosslinked peptide $\times$ dT_{20} hybrids, the molecular weight of the conjugated peptide can be calculated. Because the sequence of *E. coli* Ung is known, we are able to predict the sequences and masses of peptide fragments that will result upon proteolytic cleavage of Ung with trypsin. By matching the predicted peptide masses with the experimentally obtained masses of conjugated peptides, the latter can be identified. A list of the molecular weights of peptide $\times$ dT_{20} complexes identified this way are presented in Table IX.

An example of this is the hybrid detected at *m/z* 7,193 $\pm$ 3. Subtracting the mass of dT_{20} (6,025 Da) results in a peptide MW of 1,168 $\pm$ 3. Matching this number to the list of masses calculated for Ung tryptic peptides identifies the conjugated peptide as T18 (Calc. MW 1,169).

Other hybrids were assigned to consists of dT_{20} crosslinked to T6, T18/19 and T19 (Table IX). Automated Edman sequencing of the hybrid mixtures (performed by Bennett and Mosbaugh) was used to confirm these assignments. Sequencing identified the tryptic peptides T1, T6, T18, and T19 as being present in the samples (3). Combining the results obtained by MALDI mass spectrometry and Edman sequencing we assigned the peptides T6, T18, T18/19, and T19 as being conjugated to the oligonucleotide.

Table VIII. Survey of matrices for MALDI analysis of anion-exchange purified, desalted peptide × dT_{20} complexes.

The sample used in this survey contained a mixture of 2-3 different peptide × dT_{20} complexes and also contained some non-crosslinked dT_{20}. 0.5 μL sample (estimated concentration 5-10 μM) + 0.7 μL matrix was mixed on the mass spectrometric probe. The sample was allowed to crystallize by spontaneous evaporation of the solvent. The quality of positive ion mass spectra of peptide × dT_{20} hybrids produced by each matrix is indicated with - (not good) , + (weak), ++ (fair), +++ (good).

Matrix	Solvent	Quality
2,4,6-trihydroxy-acetophenone	50 mM citrate, 50% acetonitrile	+++
	0.1% TFA, 50% acetonitrile	+
	50% acetonitrile (aq)	-
Ferulic acid	50 mM citrate, 50% acetonitrile	++
Caffeic acid	50 mM citrate, 50% acetonitrile	+
3-hydroxypicolinic acid	50 mM citrate, 50% acetonitrile	-
2-amino-5-nitropyridine	50 mM citrate, 50% acetonitrile	-
	50% acetonitrile (aq)	-
Sinapinic acid	50 mM citrate, 50% acetonitrile	-
	50% acetonitrile (aq)	-

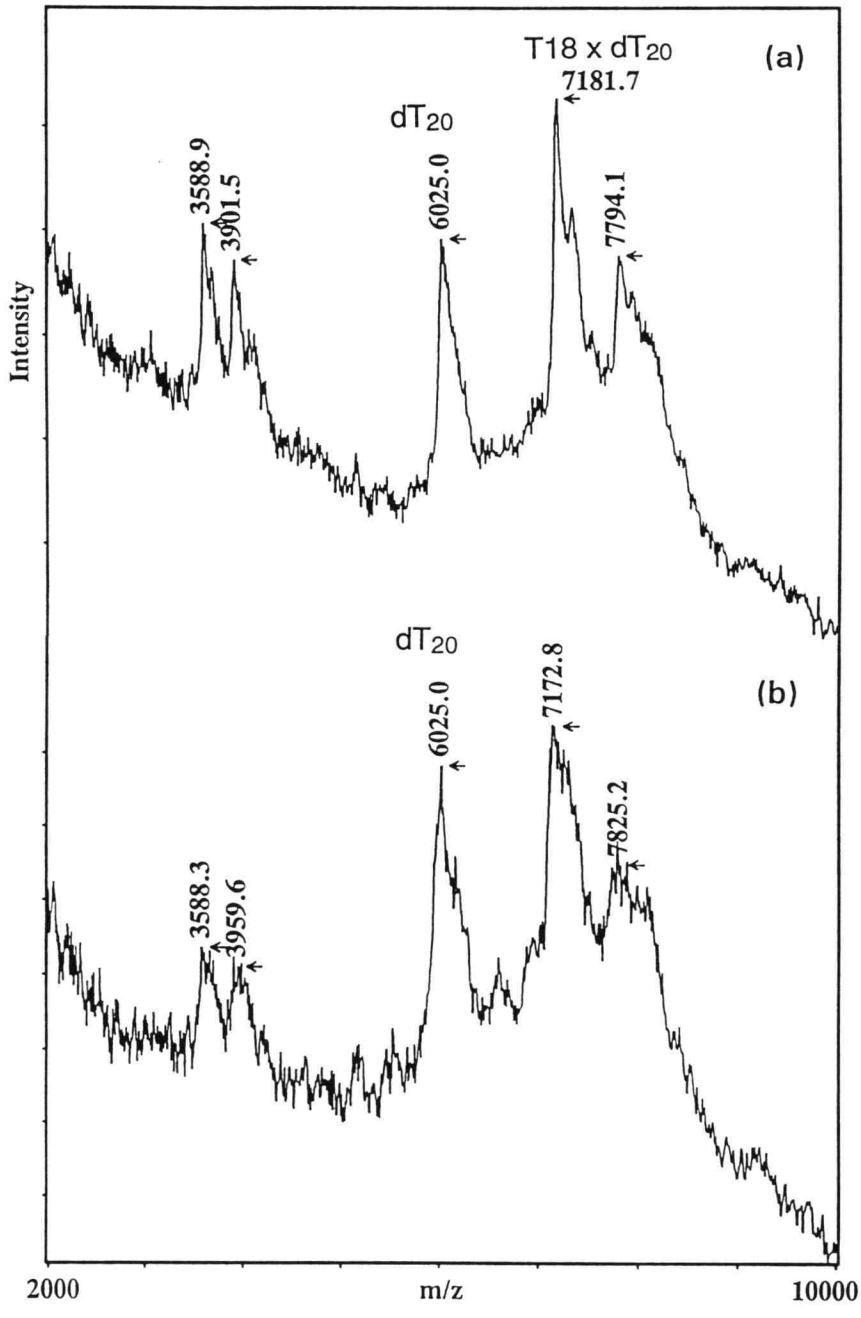


Figure 31. MALDI mass spectrum of a Nensorb desalted fraction corresponding to the major peak eluting off the anion-exchange column. *The matrix used in this experiment was (a) 2,4,6-trihydroxy-acetophenone and (b) caffeic acid. Both matrices were dissolved in 50 mM diammonium hydrogen citrate and 50% acetonitrile. Signals corresponding to dT_{20} (m/z 6,025) and several peptide x dT_{20} complexes (m/z 7,170-8,000) are observed.*

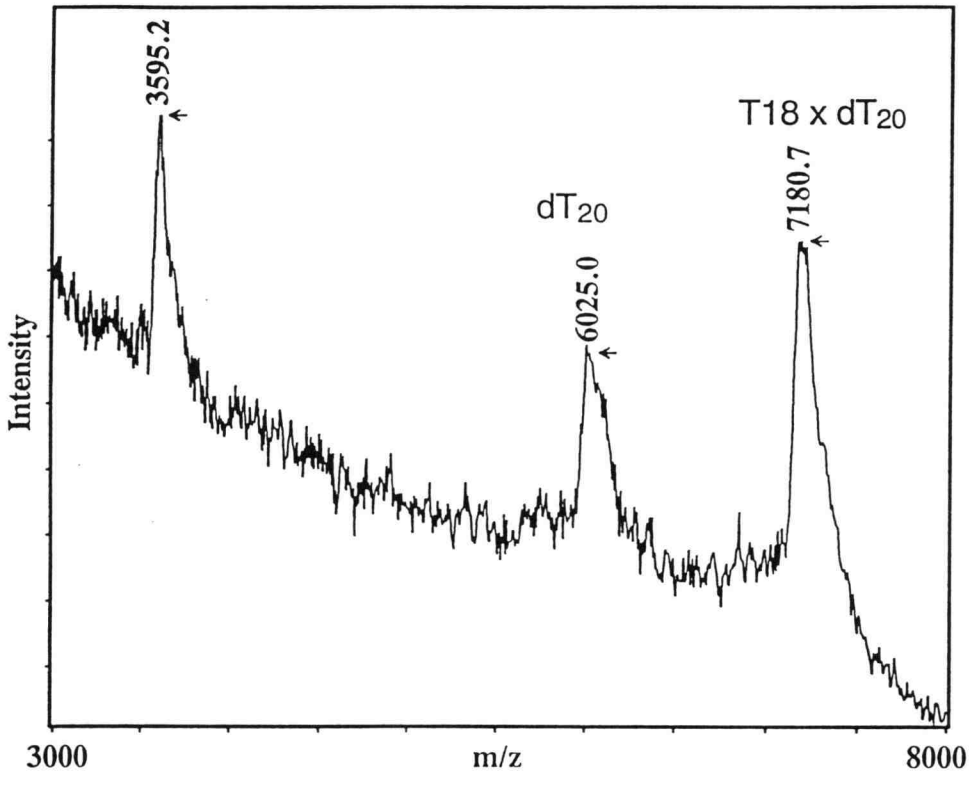


Figure 32. MALDI mass spectrum of an HPLC purified hybrid.
Ion signals corresponding to dT_{20} (m/z 6,025.0) and to [T18] × dT_{20} (m/z 7180.7) are observed.

Table IX. Molecular weight determinations of Ung × dT_{20} hybrids by MALDI mass spectrometry and correlation to Ung tryptic peptides.
Refer to Figure 19 for alignment to the Ung amino acid sequence.

Ung peptide	Calculated mass	Experimental [Peptide × dT_{20}]	Calculated [Peptide × dT_{20}] - [dT_{20}]
T6	2,419	8,448 ± 3	2,424 ± 3
T18	1,169	7,193 ± 3	1,168 ± 3
T18/19	3,376	9,417 ± 21	3,391 ± 21
T19	2,223a	8,145 ± 8	2,121 ± 8
dT_{20}	6,024	6,026 ± 8	2 ± 8

a) S-alkylated Cys. Non-alkylated Cys results in MW 2,165. Implies that Cys is involved in forming crosslink.

Peptide mapping of Rho and Rho × 4sUDP

Peptide mapping by MALDI was used to optimize conditions for proteolytic digestion and to verify the sequence of Rho. Microbore and capillary HPLC-ESI MS peptide mapping was employed in order to identify the crosslinked peptide × 4sUDP complexes by changes in elution times and/or molecular weight.

Several proteases and cyanogen bromide (CNBr) were employed to fragment Rho. In this section we present the results obtained by endoprotease Lys-C digestion of Rho. Digestion by trypsin, endoprotease Asp-N, and CNBr produced similar results, but with less coverage of the Rho sequence. These results can be found in *Appendix A.*

MALDI MS PEPTIDE MAPPING OF RHO

The Rho samples were incubated with endoprotease Lys-C for 18-22 hours. Aliquots were removed and analyzed by MALDI mass spectrometry. A MALDI peptide map of Lys-C digested UV-irradiated Rho is shown in Figure 33. The spectrum is rather complex due to the presence of at least 30 peptides. There are no signals that could correspond to intact Rho or large fragments of Rho, indicating that enzymatic digestion went to completion. A signal detected at approximately *m/z* 14,400 is a contaminant found in all Rho samples.

Based on our knowledge of the Rho amino acid sequence it is possible to predict the sequences and molecular weights of all the peptide fragments that will arise after Lys-C digestion (cleavage C-terminal to Lys residues). The predicted masses can then be compared to the experimental masses, or vice versa, to tentatively assign a peptide fragment to an ion signal. By this approach each ion signal was assigned to correspond to a Lys-C peptide as indicated in Figure 33. The background signals due to the enzyme are indicated with a B. The data obtained by going through this process, listed in Table X, maps 75% of the Rho amino acid sequence.

By comparing MALDI peptide maps from all the Rho samples and controls one should be able to detect major differences in peptide masses that could be caused by irradiation, incubation with 4sUDP, or both. All the Rho containing samples produced very similar MALDI peptide maps. There were no obvious differences between the Rho + UV (Figure 33), Rho + 4sUDP (Figure 34) and the Rho + 4sUDP + UV (Figure 35) samples. The latter sample contained signals, indicated with an asterisk (*) that can be assigned to oxidized peptides L27 (MW 1,971 + 16) and L2 or L18 (MW 2,686.4 + 16 or MW 2,685.5 + 16, respectively) and also a peptide at m/z 4,751.1 that could

originate from the peptide L13 (MW 4,653.4), since an ion signal that could correspond to the latter is absent. The origin of the mass difference of 98 Da between these species is unknown, but the fact that the mass of H_2PO_4 is 97 could be used as a starting point to speculate on the origin of this phenomenon, since phosphate may somehow originate from 4sUDP. These putative oxidized peptides were also detected in a tryptic MALDI peptide map.

Figure 33. MALDI peptide map of Lys-C digested UV-irradiated Rho.
The peaks are labeled with the Rho peptide that matches the experimental mass.

Figure 34. MALDI peptide map of Lys-C digested mixture of Rho and 4sUDP.
The spectrum is very similar to the one obtained from a Lys-C digest of UV-irradiated Rho.

Figure 35. MALDI peptide map of a Lys-C digest of a UV-irradiated mixture of Rho and 4sUDP.
Peptide ions that could be assigned by assuming oxidation of Rho Lys-C peptides are indicated with an asterisk (*).

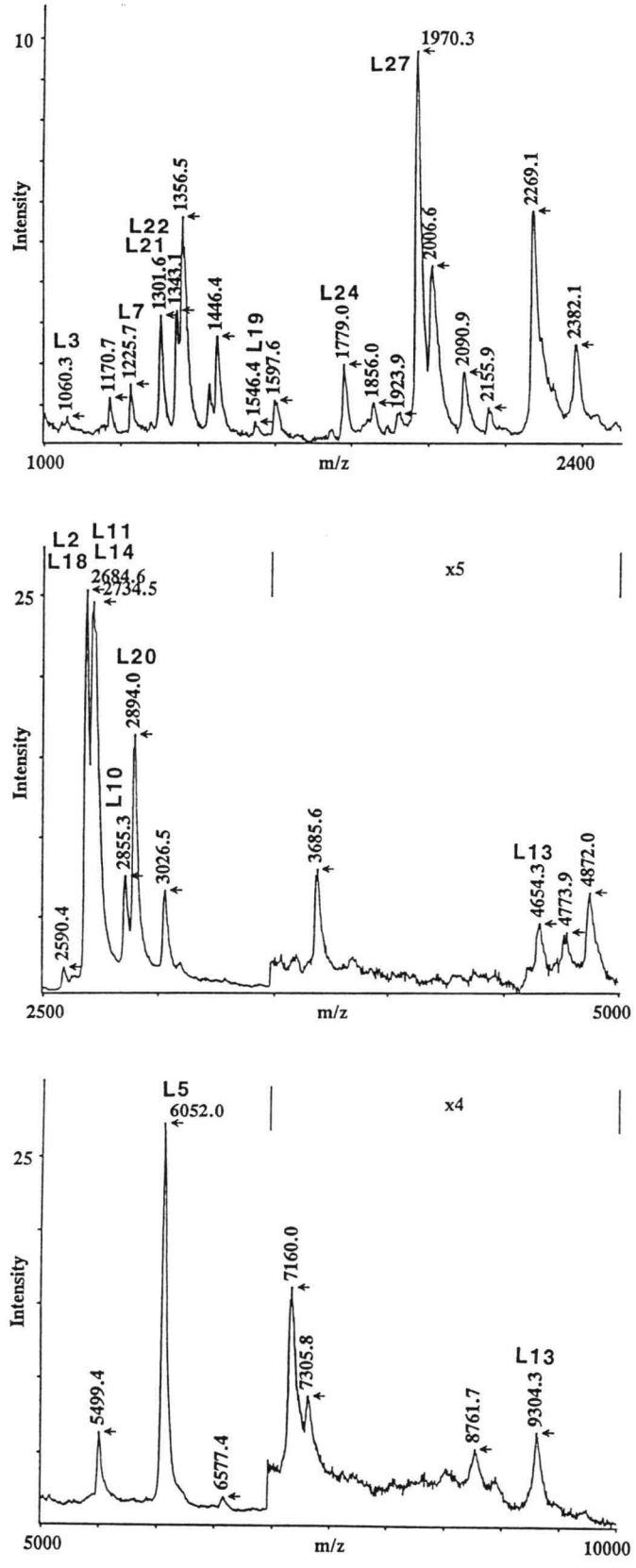

10
Intensity
1000
m/z
2400
L3
1060.3
1170.7
L7
1225.7
L22
L21
1301.6
1343.1
1356.5
1446.4
1546.4
L19
1597.6
L24
1779.0
1856.0
1923.9
1970.3
L27
2006.6
2090.9
2155.9
2269.1
2382.1
25
Intensity
2500
m/z
5000
L2
L18
L11
L14
2684.6
2734.5
x5
L20
2894.0
L10
2855.3
3026.5
2590.4
3685.6
L13
4654.3
4773.9
4872.0
25
Intensity
5000
m/z
10000
L5
6052.0
x4
5499.4
6577.4
7160.0
7305.8
8761.7
L13
9304.3

Figure 33

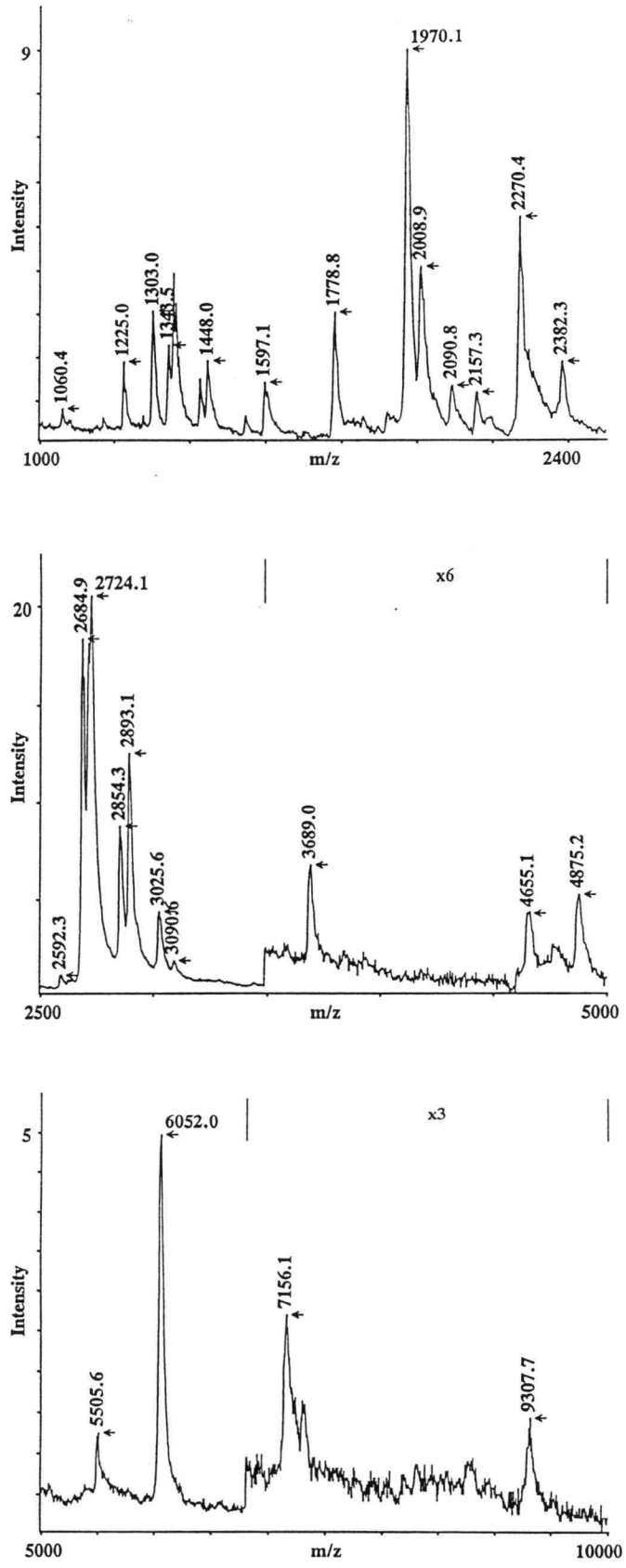

9
Intensity
1970.1
2270.4
2008.9
1303.0
1225.0
1448.0
1778.8
2382.3
1060.4
1597.1
2090.8
2157.3
1000
m/z
2400
20
Intensity
2724.1
2684.9
x6
2893.1
2854.3
3689.0
3025.6
4655.1
4875.2
2592.3
2500
m/z
5000
5
Intensity
6052.0
x3
7156.1
5505.6
9307.7
5000
m/z
10000

Figure 34

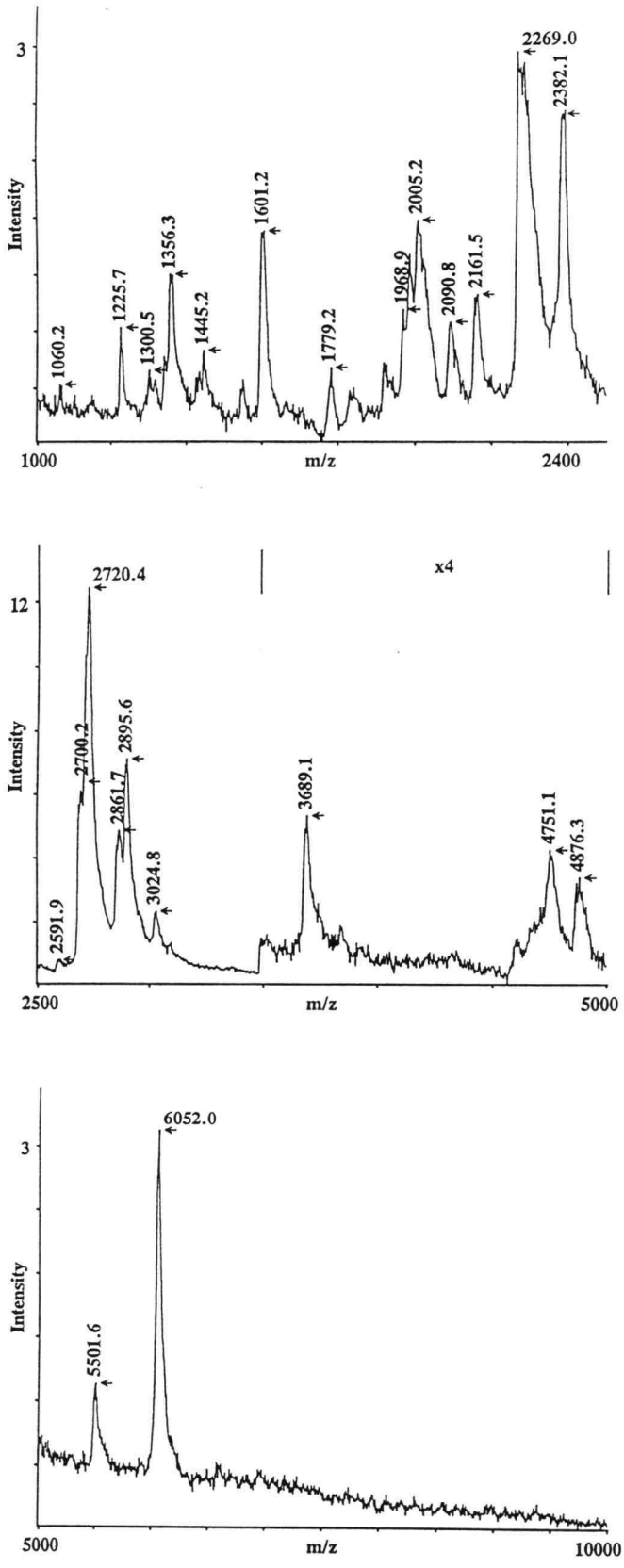

3
Intensity
1060.2
1225.7
1300.5
1356.3
1445.2
1601.2
1779.2
1968.9
2005.2
2090.8
2161.5
2269.0
2382.1
1000
m/z
2400
12
Intensity
x4
2720.4
2700.2
2895.6
2861.7
3024.8
2591.9
3689.1
4751.1
4876.3
2500
m/z
5000
3
Intensity
6052.0
5501.6
5000
m/z
10000

Figure 35

LC-MS PEPTIDE MAPPING

Separation peptides after Lys-C digestion of Rho and Rho × 4sUDP was accomplished by microbore or capillary HPLC with 214 nm UV-detection and/or mass spectrometric detection (LC-MS). LC-MS provide both high separation efficiency and mass specific detection of peptides. The LC-MS peptide map of UV-irradiated Rho is shown in Figure 36. The TIC chromatogram demonstrates the separation efficiency of a custom-packed 320 µm × 110 mm reverse phase column, although in this case the column was slightly overloaded. Molecular weight information could be extracted from the data and was used to tentatively assign peptides to the HPLC peaks.

An example of this is shown in Figure 37. This spectrum was acquired during scan 467-470 of the LC-MS run shown in figure 36. The molecular weight of the peptide, 2,728.0, matches the mass of Rho Lys-C peptide number 14 (MW 2,728.4) (Table X).

Almost all expected Rho Lys-C peptides above MW 400 were detected by this method, ensuring that almost the whole Rho sequence is covered by this peptide map (Table X and Figure 41). The masses obtained were in agreement with the calculated masses indicating that the Rho amino acid sequence is correct.

The Cys-containing peptide L13 was detected as the disulphide linked dimer at MW 9,307. This dimer was also detected in the MALDI MS peptide map. The LC-MS Lys-C peptide maps covers most of the Rho sequence (385 of 419 amino acids; 92%) as illustrated in Figure 41.

LC-MS chromatograms of Lys-C digested Rho (not shown), Lys-C digested UV-irradiated Rho (Figure 36), and Lys-C digested Rho + 4sUDP (Figure 38) were very similar, if not identical. However, the LC-MS chromatogram of Lys-C digested, UV-irradiated Rho + 4sUDP revealed several new peaks, as characterized by their elution time and molecular weight (Figure 39). An example of this is given in the extracted mass spectrum of a MW 2702 peptide that was not detected in the other samples (Figure 40). The molecular weights of several "new" peptides are given in Table XI.

None of the new peptides have a molecular weight that could correspond to peptide × 4sUDP adduct, assuming a zero-lenght photochemical bond and stability of the adduct. However, the molecular weights obtained for these peaks indicate that they might have arisen from oxidation of Rho Lys-C peptide fragments as mentioned previously. Again the peptides L27 and L2 or L18 were the best candidates for a molecular weight match assuming that one oxygen was added to the peptide (oxidation, adds 16 Da to peptide mass). The new peptide with a mass of 4,750 previously observed in the MALDI peptide map of the same sample was also detected by LC-MS.

Microbore HPLC peptide mapping of the above samples with both online UV and MS detection produced UV chromatograms that were consistent with the LC-MS chromatograms, indicating that the mass spectrometer did not fail to detect any eluting peptide.

Figure 36. LC-MS TIC chromatogram (peptide map) of UV-irradiated and Lys-C digested Rho.
The peaks are labeled with the corresponding Rho peptide as identified by mass determination.

Figure 37. Mass spectrum extracted from the LC-MS TIC chromatogram.
The spectrum is dominated by the doubly (m/z 1364.8) and triply (m/z 910.5) charged ions corresponding to the peptide L14 (MW 2728.4). Average of scan 467-470.

Figure 38. LC-MS peptide map of a Lys-C digest of a mixture of Rho and 4sUDP.
The TIC chromatogram appears very imilar to the one obtained by Lys-C digests of Rho and UV-irradiated Rho.

Figure 39. LC-MS peptide map of a Lys-C digest of UV-irradiated mixture of Rho and 4sUDP.
Some new peaks, as characterized by their altered elution time and molecular weight are indicated with asterisks.

Figure 40. Extracted mass spectrum of a new peptide from the LC-MS peptide map.
The molecular weight of this peptide is 2,702.2. It can be assigned to be an oxidized form of either L2 or L18.

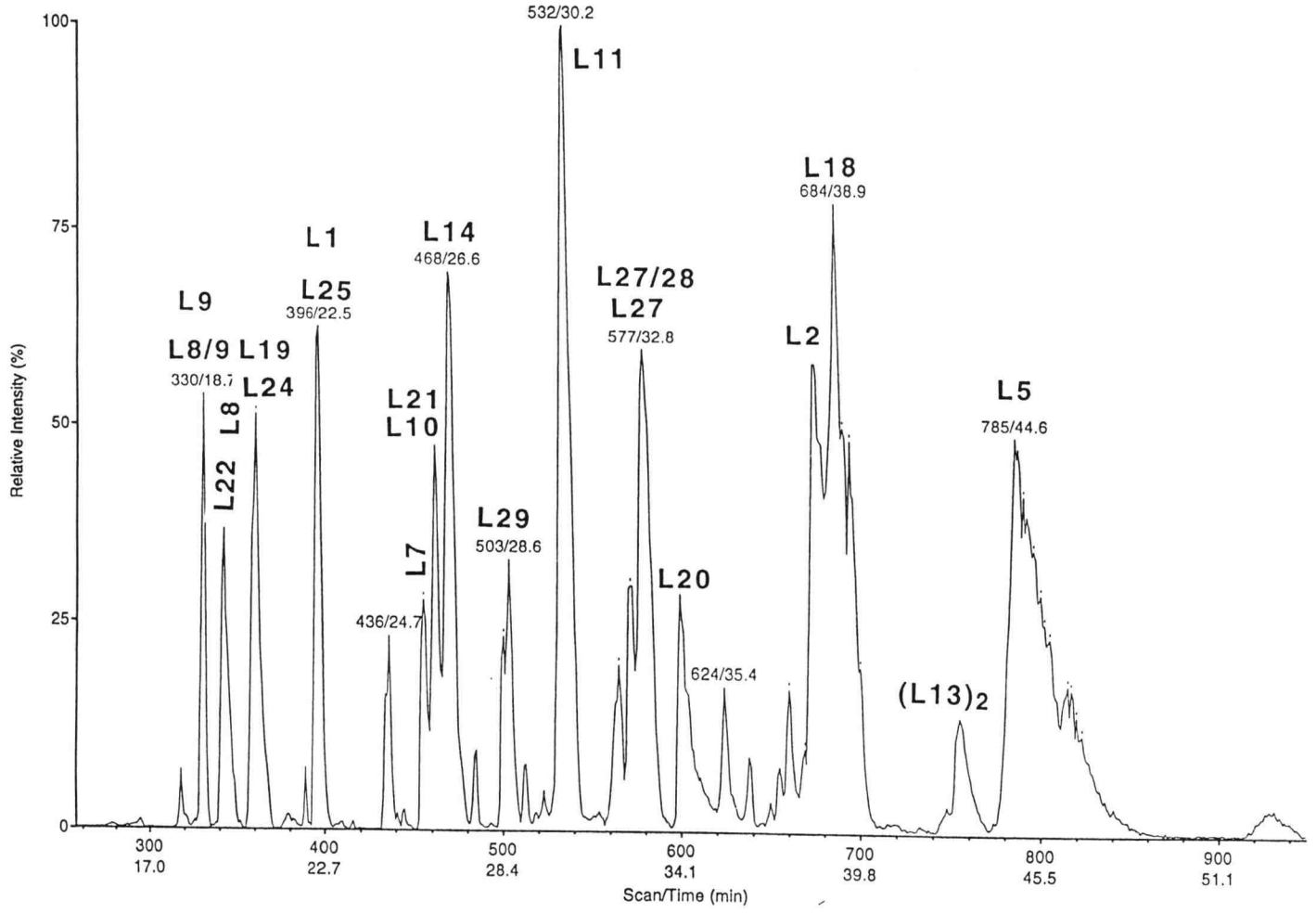

100
75
50
25
0
Relative Intensity (%)
532/30.2
L11
L18
684/38.9
L14
468/26.6
L1
L25
396/22.5
L9
L27/28
L27
577/32.8
L8/9 L19
330/18.7
L24
L8
L22
L21
L10
L2
L5
785/44.6
L7
L29
503/28.6
436/24.7
L20
624/35.4
(L13)2
300
17.0
400
22.7
500
28.4
600
34.1
700
39.8
800
45.5
900
51.1
Scan/Time (min)

Figure 36.

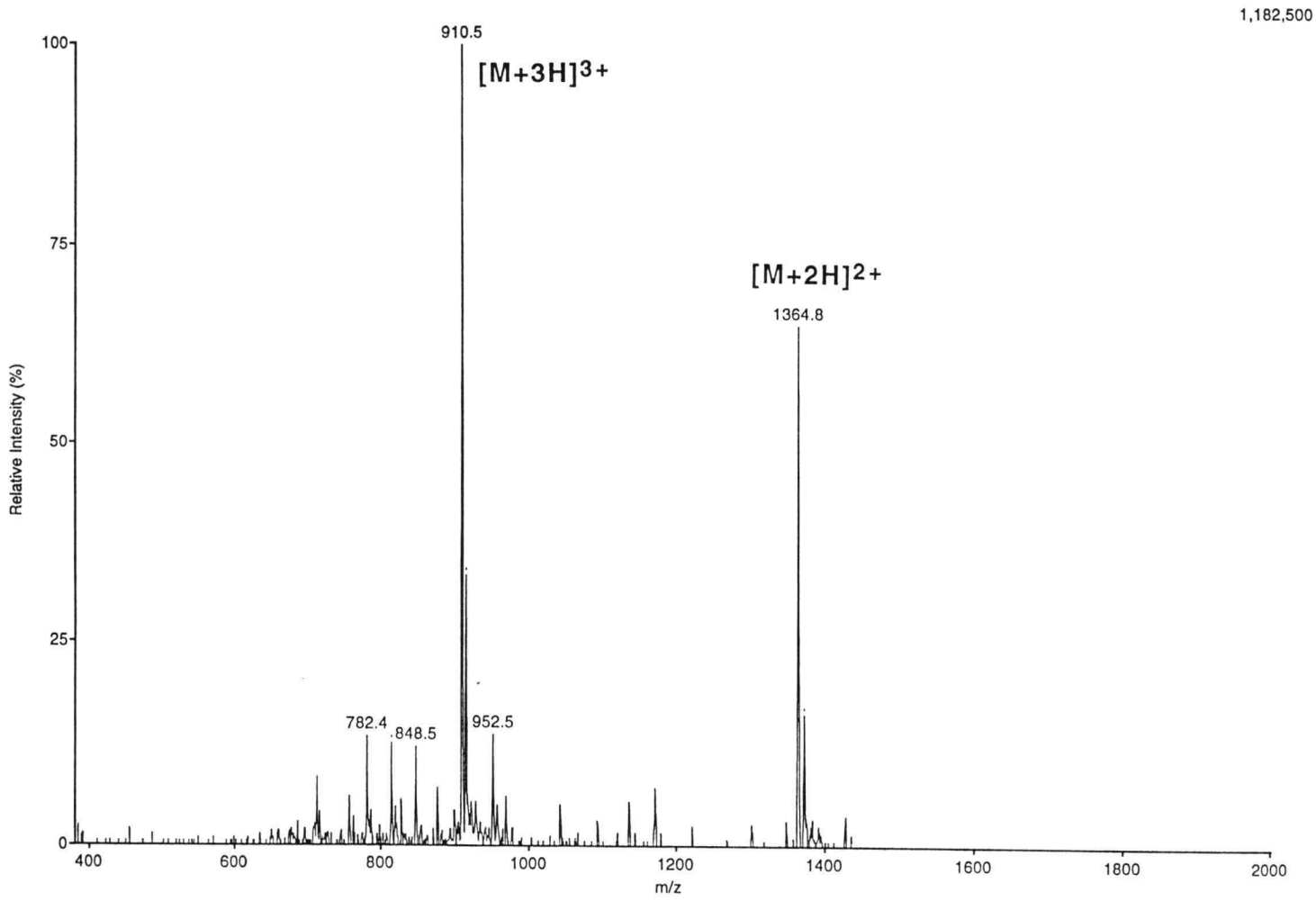

1,182,500
910.5
[M+3H]3+
[M+2H]2+
1364.8
782.4
848.5
952.5
Relative Intensity (%)
100
75
50
25
0
400
600
800
1000
1200
1400
1600
1800
2000
m/z

Figure 37.

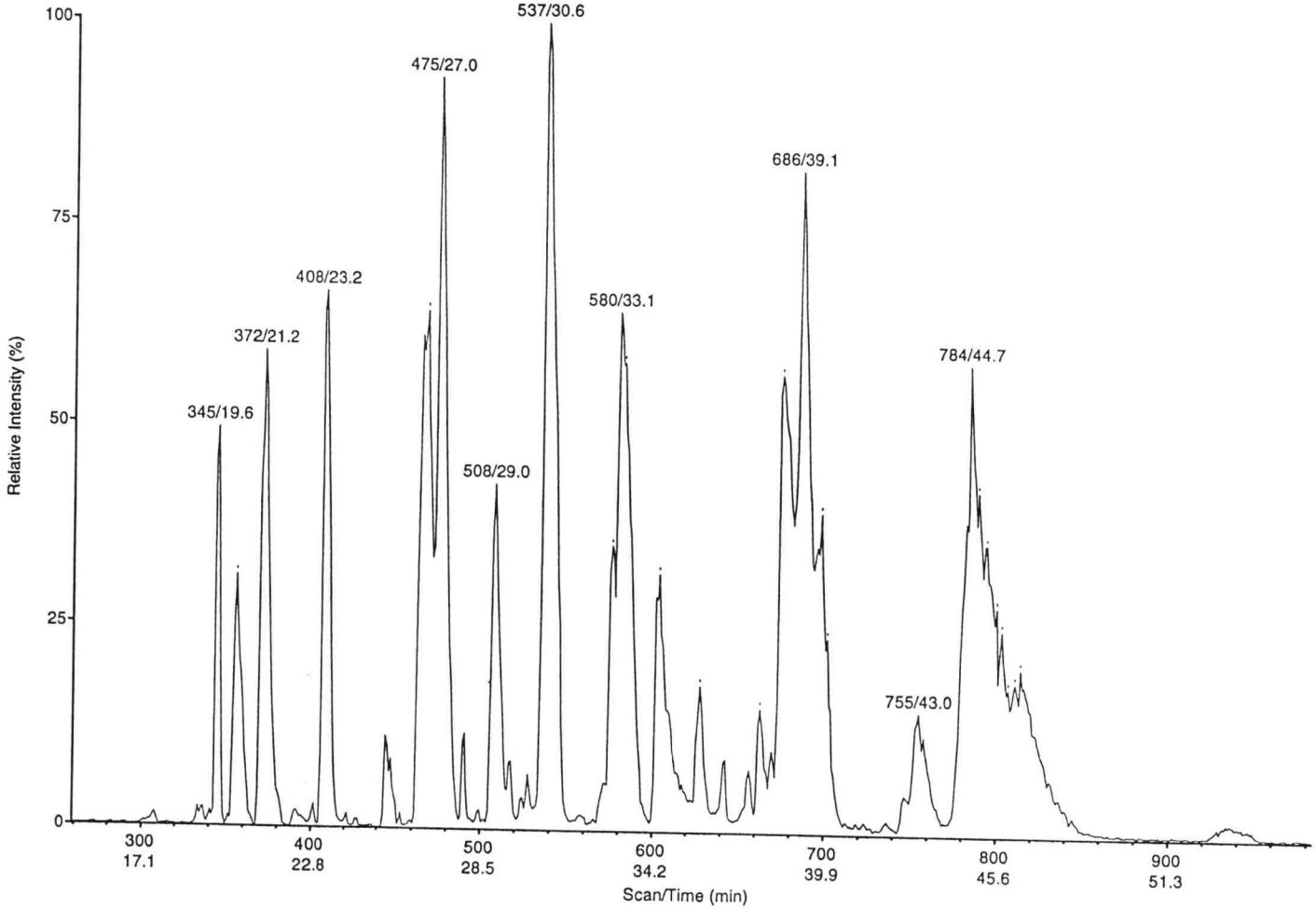

100
75
50
25
0
Relative Intensity (%)
537/30.6
475/27.0
686/39.1
408/23.2
580/33.1
372/21.2
784/44.7
345/19.6
508/29.0
755/43.0
300
17.1
400
22.8
500
28.5
600
34.2
700
39.9
800
45.6
900
51.3
Scan/Time (min)

Figure 38.

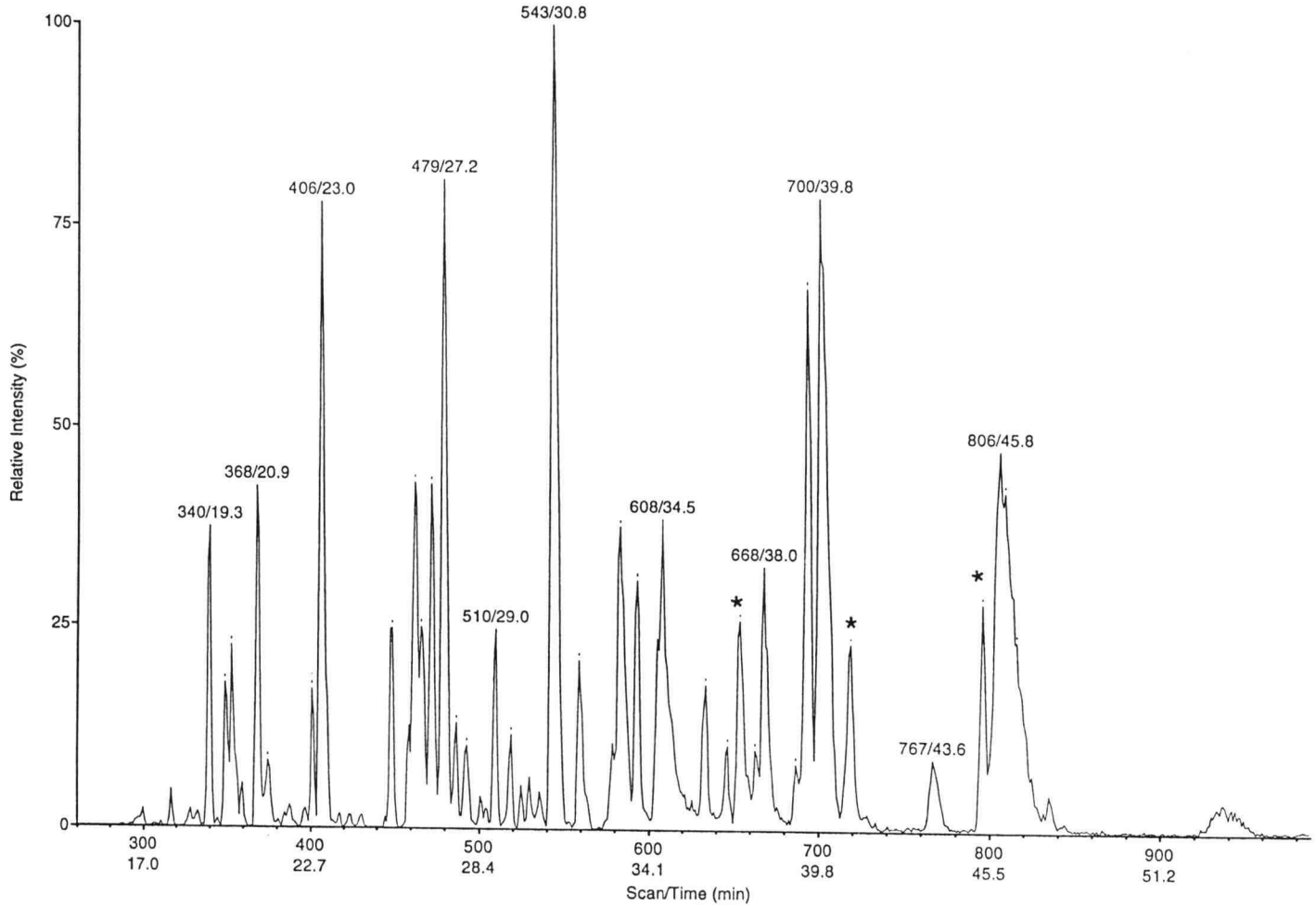

543/30.8
479/27.2
406/23.0
700/39.8
806/45.8
368/20.9
340/19.3
608/34.5
668/38.0
510/29.0
767/43.6
100
75
50
25
0
Relative Intensity (%)
300
17.0
400
22.7
500
28.4
600
34.1
700
39.8
800
45.5
900
51.2
Scan/Time (min)

Figure 39.

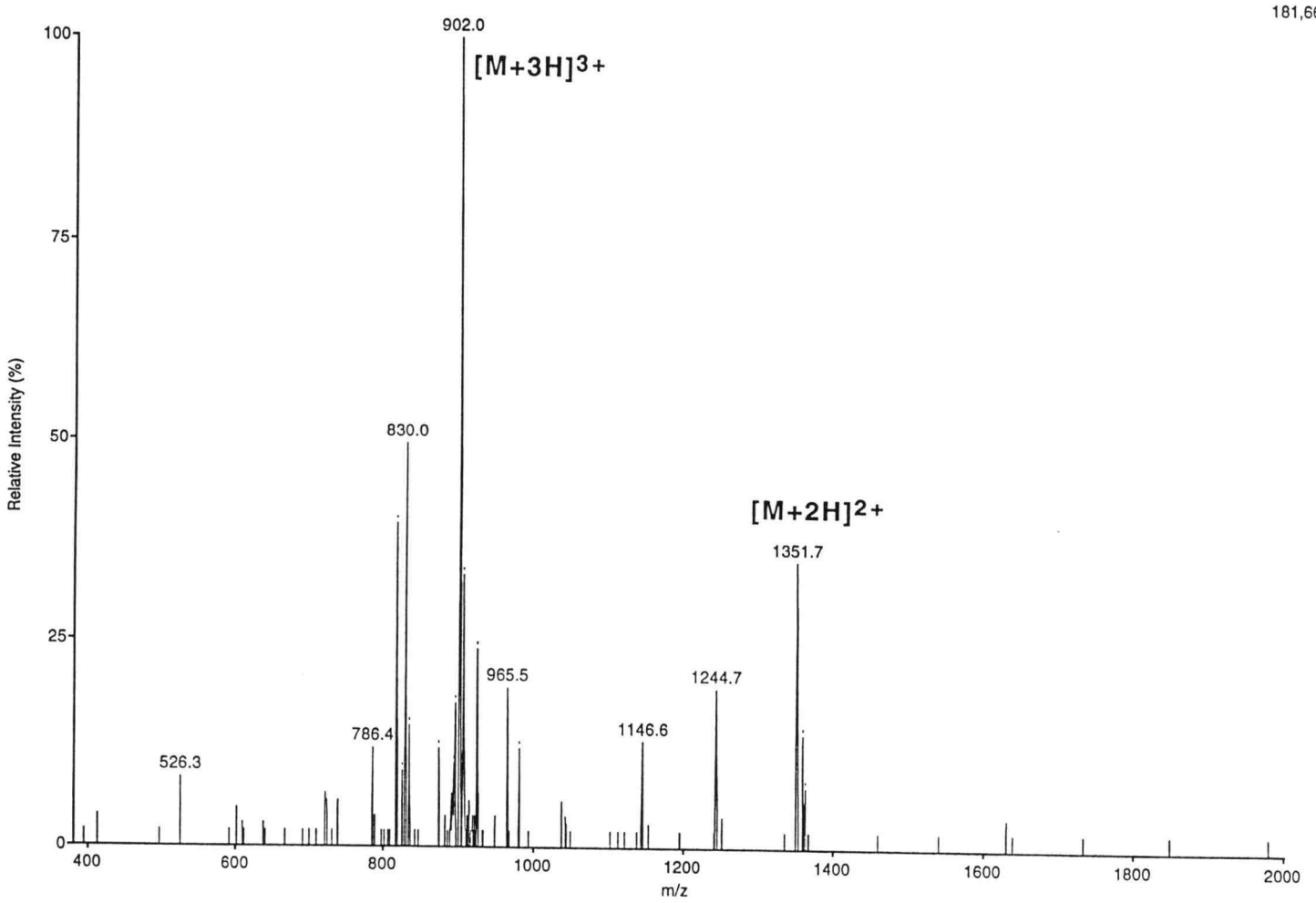

181,667
902.0
[M+3H]3+
100
75
50
25
0
Relative Intensity (%)
830.0
[M+2H]2+
1351.7
965.5
1244.7
786.4
1146.6
526.3
400
600
800
1000
1200
1400
1600
1800
2000
m/z

Figure 40.

Table X. Molecular weight data obtained by mass spectrometric peptide mapping of a Lys-C digest of Rho by MALDI and HPLC-ESI MS.

Num.	Position	Calc. (MH+)	MALDI (MH+)	ESI (MH+)
1	1 - 7	848.46	-	848.5
2	8 - 31	2686.40	2684.6	2686.5
3	32 - 40	1060.64	1060.4	1061.1
4	41 - 44	483.27	-	-
5	45 - 100	6052.00	6052.0	6053.6
6	101 - 105	610.40	-	-
7	106 - 115	1225.66	1225.00	1225.7
8	116 - 123	964.47	-	964.5
8,9	116 - 130	1773.89	-	1774.5
9	124 - 130	828.43	-	828.5
10	131-155	2854.51	2854.3	2854.9
11	156-181	2713.58	2712.0	2714.0
12	182 - 184	275.17	-	-
13	185 - 224	4653.39	4655.10	
			9307.7[a]	9307.8[a]
14	225 - 249	2728.36	2724.10	2729.1
15	250 - 251	218.15	-	-
16	252 - 257	781.47	-	-
17	258 - 258	147.11	-	-
18	259 - 283	2685.53	2684.9 ?	2685.2
19	284 - 298	1547.87	1547.1	1548.4
20	299 - 326	2894.54	2893.1	2894.8
21	327 - 336	1302.59	1303.0	1302.7
21,22	327 - 348	2626.26	-	2626.3
22	337 - 348	1342.69	1343.5	1342.7
23	349 - 352	460.28	-	-
24	353 - 367	1779.96	1778.8	1780.0
25	368 - 379	1460.75	-	1460.8
26	380 - 385	846.50	-	n.d.
27	386 - 402	1971.01	1970.1	1971.5
28	403 - 407	563.32	-	-
29	408 - 417	1277.52	-	1277.7
30	418 - 419	262.15	-	-

a) Disulphide bridged dimer

MNLTELK▾NTP VSELITLGEN MGLENLARMR K▾QDIIFAILK▾ 40
L1 (848.5) **L2 (2686.4)** **L3 (1060.6)**

QHAK▾SGEDIF GDGVLEILQD GFGFLRSADS SYLAGPDDIY 80
L4 **L5 (6052.0)**

VSPSQIRRFN LRTGDTISGK▾ IRPPK▾EGERY FALLK▾VNEVN 120
L6 **L7 (1225.7)** **L8 (964.6)**

FDK▾PENARNK▾ ILFENLTPLH ANSRLRMERG NGSTK▾DLTAR 160
L9 (828.4) **L10 (2854.5)**

VLDLASPIGR GQRGLIVAPP K▾AGK▾TMLLQN IAQSIAYNHP 200
L11 (2713.6) L12

DCVLMVLLID ERPEEVTEMQ RLVK▾GEVVAS TFDEPASRHV 240
L13 (4653.4) **L14 (2728.4)**

QVAEMVIEK▾A K▾RLVEHK▾K▾DV IILLDSITRL ARAYNTVVPA 280
L15 L16 L17 **L18 (2685.5)**

SGK▾VLTGGVD ANALHRPK▾RF FGAARNVEEG GSLTIIATAL 320
L19 (1547.9) **L20 (2894.5)**

IDTGSK▾MDEV IYEEFK▾GTGN MELHLSRK▾IA EK▾RVFPAIDY 360
L21 (1302.6) **L22 (1342.7)** L23 **L24 (1780.0)**

NRSGTRK▾EEL LTTQEELQK▾M WILRK▾IIHPM GEIDAMEFLI 400
L25 (1460.8) L26 **L27 (1971.0)**

NK▾LAMTK▾TND DFFEMMK▾RS 419
L28 **L29 (1277.5)** L30

Figure 41. Mapping of Lys-C peptides to the Rho amino acid sequence based on LC-MS and MALDI MS data.
Cleavage points are indicated with ***v.*** *Rho Lys-C peptides are numbered from the N-terminus. Peptides detected by LC-MS peptide mapping are in boldface with their molecular weight in parentheses.*

Table XI. Molecular weights of "new" Rho Lys-C peptides present in the UV-irradiated Rho + 4sUDP sample.

Observed *new* molecular weight	Nearest Rho peptide match	Mass	Comment
1986.3	L27	1971.0 (386-402)	Oxidation
2702.2	L2 or	2685.4 (8-31) or	Oxidation
	L18	2684.5 (259-283)	Oxidation
4750.3	L13	4653.4 (185-224)	97 Da mass difference

TANDEM MASS SPECTROMETRY OF 4SUDP

We have used ESI tandem MS to investigate the mass spectrometric behaviour of 4sUDP (MW 420). Negative ion mass spectra were collected after collision induced dissociation of the [4sUDP - H]$^-$ ion at *m/z* 419. The daughter spectrum shown in Figure 42 exhibits signals corresponding to [PO_3]$^-$ (*m/z* 79), [H_2PO_4]$^-$ (*m/z* 96.7), [4sU]$^-$ (*m/z* 126.7), and [HP_2O_6]$^-$ (*m/z* 158.9). The complementary ions [4sUDP - PO3 - H]$^-$ (*m/z* 338), [4sUDP - HPO_4 - H]$^-$ (*m/z* 320.6), and [4sUDP - 4sU - H]$^-$ (*m/z* 291) are also observed. Several signals arising from loss of water are detected at *m/z* 400.6, 320.6, 272.7, and 193.1.

The most facile fragmentation pathways for 4sUDP are therefore loss of water, loss of base, and loss of one or two phosphate groups.

The specific loss of phosphates from 4sUDP can be used to selectively detect peptide × 4sUDP complexes in mixtures. This is accomplished by using either a "parent-of-*m/z* 79" scanning mode of tandem MS, or by using collisional excitation in the source region by ramping the orifice potential while scanning the low mass region (116). In the "parent-of" scan mode, MS 1 is scanning the mass range of interest, while MS2 is scanning at the same rate but offset by a m/z corresponding to the fragmentation product of interest, for example, *m/z* 79 for PO_3^-. In the collisional excitation approach the orifice potential is synchronized with the MS1 *m/z* scan by computer control. When MS 1 is scanning the low mass range (*m/z* 60-100), the orifice potential is set to a value that induces fragmentation of all ions, in turn producing PO_3^- ions if a phosphate containing analyte is present. When MS1 is scanning the rest of the mass range (*m/z* 100-2000), the orifice potential is set to normal allowing intact analyte ions to be detected. Note that this experiment can be done on a single-quadropole instrument.

The latter experiment was attempted by LC-MS analysis of peptide mixtures derived from Rho × 4sUDP. However, our current instrumental setup is not optimal for this type of experiment: The manufacturers limit for the orifice potential is -250 V, whereas -350 V has been reported to produce best results (116). This might explain why the experiments failed. Alternately, the expected peptide × 4sUDP adducts were not present as anticipated. Regardless, this approach should, implemented properly, be extremely useful for identifying peptide × nucleic acid adducts in peptide mixtures.

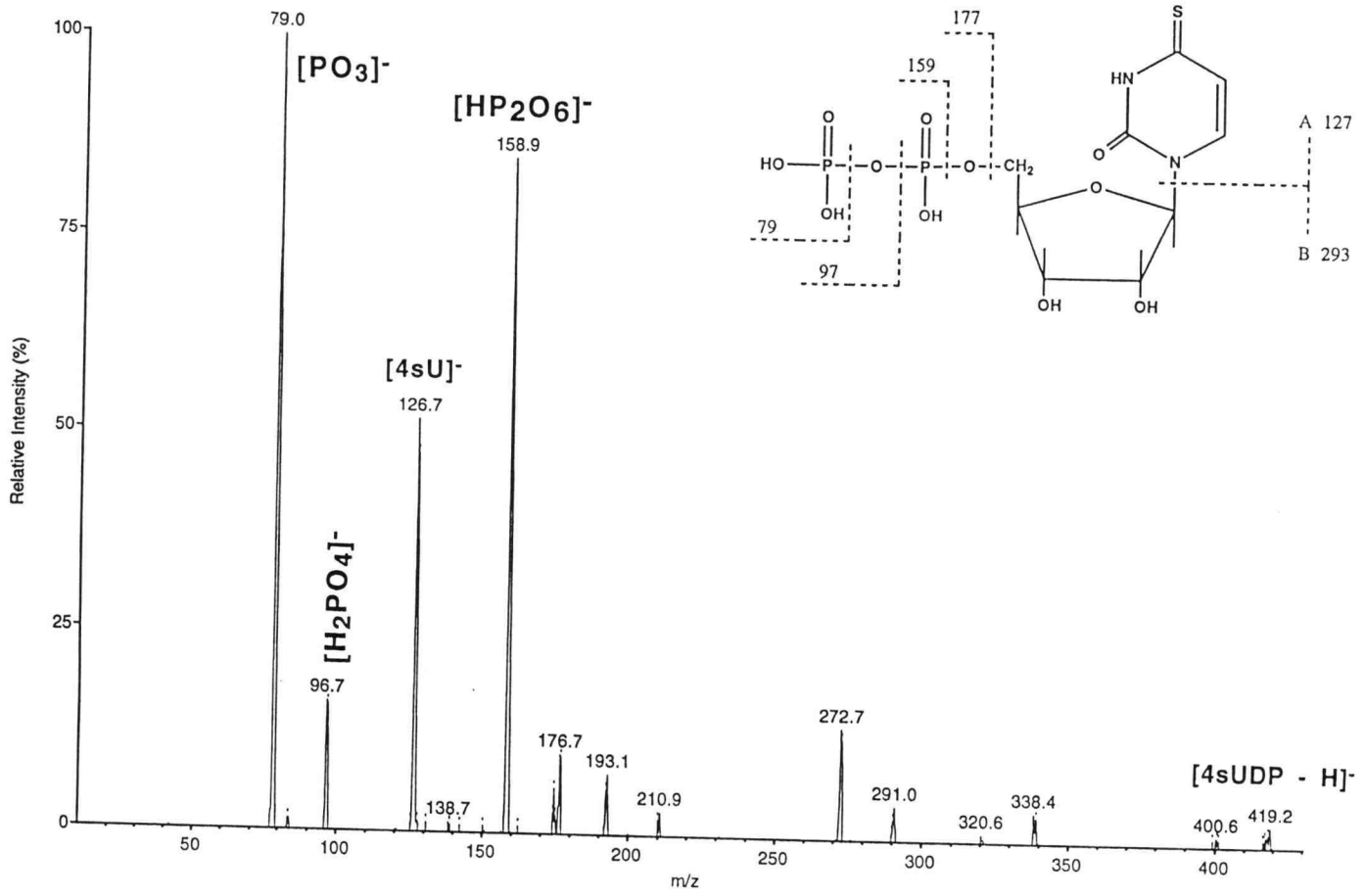


Figure 42. Negative ion tandem mass spectrum of 4sUDP.
The fragments ions are indicated.

Chapter 7: Discussion

We have presented a novel strategy for the identification of nucleic acid binding domains in proteins. The method employes mass spectrometric techniques for the analysis of photochemically crosslinked protein-nucleic acid complexes. The experimental data described in the previous chapter documents our efforts to establish and implement the analytical protocol. Using several different protein-nucleic acid systems and different crosslinking techniques, we have demonstrated the feasibility of a mass spectrometric based approach to investigate the character of UV-crosslinked protein-nucleic acid complexes. Matrix-assisted laser desorption/ionization mass spectrometry was used for detection and molecular weight determination of intact, UV-crosslinked protein-nucleic acid complexes, for peptide mapping, and for molecular weight determination of synthetic and UV-crosslinked peptide-nucleic acid hybrids. Electrospray ionization mass spectrometry and tandem mass spectrometry was used for analysis of synthetic peptide-nucleic acid hybrids and, in conjunction with HPLC, for peptide mapping of a protein.

MOLECULAR WEIGHT DETERMINATION OF INTACT UV-CROSSLINKED PROTEIN-NUCLEIC ACID COMPLEXES BY MALDI MASS SPECTROMETRY

At the onset of this study, it was not known how effective the analysis of crude, UV-irradiated protein samples by MALDI mass spectrometry would be. This was particularly so in the case of the pulsed laser UV-crosslinking experiment because the small quantity of crosslinked analyte produced by a single UV laser pulse (0.01 to 3 μM) is close to MALDI mass spectrometry's practical detection limit for pure proteins. It is now clear that MALDI mass spectra of a nucleoprotein complex can be produced directly from the crosslinking reaction mixture without any cleanup even when the radiation is in the form of a single laser pulse of a few nanoseconds duration. To the best of our knowledge, these spectra, of which those shown in the Results section are representative, provide the first direct, mass spectrometric confirmation of the formation of UV-induced, covalent protein-DNA and protein-RNA complexes.

Our experiments with gp32 and dT_{20} clearly indicate that the success of MALDI mass analysis with covalent protein-nucleic acid complexes is highly dependent on the combination of matrix material and solvent system employed. Indeed, it was found that different matrices are necessary to produce good spectra of pure gp32, dT_{20}, and gp32 $\times$ dT_{20} (Table IV). Selection

of matrix compound is of primary importance in producing MALDI mass spectra; however, it is evident from our experiments that choice of solvent is also of considerable consequence. The correct solvent is not only necessary for maintaining the analyte in solution but it is apparently essential for getting the analyte of interest incorporated into the matrix crystals. This matrix-solvent dependency could conceivably be advantageous in discriminating against all the components in a complex sample except the analyte of interest. Our present sample preparation procedure for gp32 $\times$ dT_{20} fails to produce a spectrum about 25% of the time. Perhaps the matrix/solvent combination is not optimum, or maybe there is variability in the components of the reaction mixture.

An interesting observation is that the presence of 1% (v/v) glycerol in the sinapinic acid/formic acid/acetonitrile matrix solution seems to be important for proper incorporation of gp32 $\times$ dT_{20} into the matrix crystals and, thus, for obtaining spectra of the gp32 $\times$ dT_{20} complex. The glycerol must, however, be removed prior to MALDI mass analysis by washing the sample/matrix deposit in cold water. We became aware of glycerol's role in the preparation of these samples when we attempted to improve the quality of our spectra by analyzing glycerol-free instead of glycerol-containing gp32 $\times$ dT_{20} samples.

The Ung $\times$ dT_{20} complex was analyzed under conditions similar to those used for mass determination of the gp32 $\times$ dT_{20} complex, indicating that the use of protein solubilizing agents (formic acid) is important for succesful analysis of UV-crosslinked protein-oligonucleotide complexes. This was also reflected in the analysis of Rho $\times$ 4sUDP where either 1,1,1,3,3,3-hexafluoroisopropanol or formic acid had to be employed during sample preparation in order to obtain MALDI mass spectra of this species.

In all cases, i.e. gp32 $\times$ dT_{20} , Ung $\times$ dT_{20} , and Rho $\times$ 4sUDP, the molecular weight determinations by MALDI MS indicate that crosslinked complexes are present. However, there is a small discrepancy between the experimentally determined average molecular weight and the calculated molecular weight of the complex (Table V). In the case of gp32 $\times$ dT_{20} and Rho $\times$ 4sUDP, the experimental value is larger than would be expected from attaching the nucleic acid to the protein with a zero-length covalent bond. For gp32 $\times$ dT_{20} , the discrepancy could be due to adduction of alkali metal ions to the oligonucleotide. In the case of Rho $\times$ 4sUDP, the difference between the experimental and theoretical values for the added 4sUDP moiety is about 1 part in 5 (20%, 420 vs. 560 Da). This relatively large discrepancy could be due to the presence of an appreciable, but unresolved matrix adduct, or it could be due in large part to the fact that the experimental value was obtained by subtracting two, large, nearly identical, independently determined numbers one of which has an appreciable uncertainty associated with it. If the latter possibility is the case, a MALDI mass analysis of a mixture of Rho and Rho $\times$ 4sUDP should yield a more accurate and more precise estimate of the nucleic acid moiety mass; however, we have not

yet found a matrix/solvent combination that allows simultaneous analysis of both species. Given the mostly unknown nature of the photochemically induced covalent bond, it is also reasonable to think in either case that the discrepancy and the associated uncertainty might be due to the existence of different types of photochemical crosslinks between protein and nucleic acid or, in the worst case, non-specific labeling. As discussed below, LC-MS peptide mapping detected several oxidized peptides in the Lys-C digest of Rho. Oxidized residues could also be part of the mass deviation that is observed for intact Rho × 4sUDP.

The molecular weight obtained for Ung × dT_{20} is about 100 mass units below the theoretical molecular weight calculated by assuming that a 'zero-length' covalent bond has been formed. This deviation could be due to non-linear calibration brought about by using low molecular weight calibrants (matrix and dT_{20} signals) in combination with the use of above-normal laser fluence to obtain the spectrum. The high laser fluence may induce decomposition of the complexes which could give rise to peak broadening and decreased precision.

The relevance of these minimal mass deviations of experimental *versus* calculated molecular weights of the intact, crosslinked protein-nucleic acid complexes is uncertain. The accuracy of the molecular weight determinations (0.1-0.2%) is an order of magnitude better than what can be obtained by any other technique, and in all three cases it provides direct evidence that UV-crosslinked protein-nucleic acid complexes have been formed.

MALDI PEPTIDE MAPPING OF PROTEIN × NUCLEIC ACID COMPLEXES

We have used peptide mapping by MALDI mass spectrometry to monitor the time-course of proteolytic digestions of proteins and protein × nucleic acid complexes. The speed and sensitivity of this technique allows digestion reactions to be monitored and optimized using only small amounts of sample. The MALDI peptide maps detected 75-90% of the expected peptides from the proteins studied and provided a quick way of confirming large parts of amino acid sequence of the polypeptides.

MALDI peptide maps of Ung × dT_{20} and Rho × 4sUDP were very similar to peptide maps obtained from Ung and Rho, respectively. A single site of crosslinking should, if the sample is homogenous, result in the disappearance of a signal from the MALDI peptide map of the protein × nucleic acid complex, as compared to the non-crosslinked protein peptide map. The signal that disappeared would correspond to the crosslinked peptide. Since this is obviously not observed, it would seem that there is no crosslinked peptide present or that there is more than one crosslinked peptide, i.e. a given crosslinked peptide is also present in the non-crosslinked native

form. Since the latter apparently is the case, at least for Ung, the peptide × nucleic acid species must have been suppressed under the conditions used (see discussion below).

In the case of Ung, most of the polypeptide sequence was mapped by MALDI mass analysis of a tryptic peptide mixture (Table VI and Figure 19). Attempts to produce signals from peptide × dT_{20} hybrids, known to be present in the peptide mixture, by using a variety of matrices, were not successful. The presence of many non-crosslinked peptides may suppress the peptide × dT_{20} species. Presumably, the peptides are preferentially incorporated into the matrix crystals, in effect excluding the crosslinked hybrids from analysis. The presence of 1 M urea in the digest mixture could play a role, since this denaturant might interferes with crystallization of the matrix. Further work is necessary to identify matrix/solvent combinations that will allow selective desorption/ionization of peptide × nucleic acid complexes.

Similar observations were made for MALDI peptide mapping of Lys-C digests of Rho and Rho × 4sUDP, respectively. We did not observe any signals that could be attributed to peptide × 4sUDP complexes. Since we have not been able to detect any crosslinked peptide-4sUDP species by any of the mapping methods employed, we assume that the photoaffinitylabel is unstable and has degraded, as discussed below.

SYNTHETIC PEPTIDE-LINKER-OLIGONUCLEOTIDE HYBRID

In order to investigate the applicability of MALDI and ESI mass spectrometry to the analysis of peptide-oligonucleotide conjugates, we have produced peptide-oligodeoxythymidylic acid hybrids by chemical synthesis and by photochemical crosslinking.

A novel synthetic approach employing PyBOP as a catalyst to form an amide bond between a peptide and oligothymidylic acids was used. The yields of hybrids, approximately 5% as judged from UV-absorption of oligonucleotides, was sufficient for mass spectrometric studies. MALDI mass spectrometry was used in combination with HPLC to monitor the time course of the reaction. Off-line analysis of HPLC fractions by MALDI MS and ESI MS identified fractions that contained covalent hybrids.

Molecular weight information on purified peptide-linker-oligonucleotide hybrids can be easily obtained by either MALDI mass spectrometry or by ESI mass spectrometry. MALDI MS is the most sensitive of the two ionization techniques used in this work and could detect covalent hybrids directly in the reaction mixture. Low (5-10) picomole amounts of hybrids could be analyzed by MALDI, whereas ESI required larger amounts (> 50 pmoles/μL) for flow injection analysis. LC-MS analysis was only possible when the oligonucleotide was similar in size to the peptide. The resolution of ESI MS is superior to what can be obtained by MALDI MS. For both

MALDI and ESI MS, the accuracy of molecular weight determinations where within 0.05% of the calculated masses for peptide, aminolinker-oligo(dT)'s, and peptide-linker-oligo(dT) hybrids.

MALDI analysis of hybrids in the positive ion mode was superior to analysis in the negative ion mode on our instrument in that better resolution and more intense signals were observed. However, our instrument is not optimized for negative ion analysis.

ESI analysis of hybrids in the negative ion mode produced less sodium adducted ions than the positive ion mode causing negative ion analysis of the peptide-linker-dT_{10} hybrid to be slightly more sensititve than positive ion analysis. However, the hybrid with the shorter oligonucleotide (peptide-linker-dT_6) produced the most intense signals in the positive ion mode. This suggests that short oligonucleotides are preferable in order to increase sensitivity of ESI MS analysis, including LC-MS peptide mapping. However, in order to get biologically meaningful results by our protocol, the length of the oligonucleotide used in an experiment should be dictated by the DNA binding protein of interest. For example, the shortest oligo(dT) molecule that binds to gp32 is dT_8. However, the length of the oligonucleotide can be trimmed *after* UV-crosslinking has taken place which should allow both maximum sensitivity and meaningful results to be obtained.

The experimental conditions used for MALDI analysis and ESI analysis of peptide-nucleic acid hybrids are similar to those found to be applicable to oligonucleotide analysis by MALDI (100) and ESI (172). This was anticipated because the oligonucleotide makes up a majority (70-90%) of the mass of the hybrids studied here. More systematical studies are needed to learn more about the relative contributions of peptide and oligonucleotide to ionization of covalent hybrids. ESI MS and LC-MS are useful methods for studying peptide-oligonucleotide hybrids in cases where the oligonucleotide chain is relatively short (<10 residues) or where the contribution of peptide and oligonucleotide to the total mass is close to unity. It is possible that hybrids containing longer oligonucleotides can be analyzed by ESI MS by further optimization of the conditions in the ESI ion source or by increasing the amount of sample introduced into the mass spectrometer.

The use of positive ion ESI MS/MS for analysis of the peptide-linker-dT_6 hybrid revealed that the oligonucleotide part of the molecule fragments more easily than the peptide-part, allowing sequencing of the nucleotide portion of the hybrid. This fact, if true in general, should be advantageous for the analysis of UV-crosslinked peptide-oligonucleotide complexes in that it is possible to obtain sequence information on the target-DNA sequence of a DNA binding protein.

Covalent peptide-oligonucleotide hybrids have recently been synthesized and used in antisense research, under the assumption that the peptide portion of the hybrid might help stabilize, transport, or target antisense oligonucleotides (177, 178, 179, 180, 181). Mass spectrometric analysis of antisense peptides (182) and antisense oligonucleotides (183, 184) have previously been demonstrated. Our results for the peptide-linker-oligonucleotide hybrids

illustrates additional uses of mass spectrometry as a tool in antisense research, namely as a means to monitor the synthesis of peptide-oligonucleotide hybrids, to confirm molecular weights of the hybrids, and to some structural information.

MASS SPECTROMETRIC ANALYSIS OF PURIFIED UNG PEPTIDE × NUCLEIC ACID COMPLEXES

Crosslinking of Ung to dT_{20} was accomplished by using continous UV irradiation and crosslinked peptide × dT_{20} were purified after tryptic digestion of the Ung × dT_{20} complex. The pools of crosslinked peptide-dT_{20} hybrids were then identified by using MALDI mass spectrometry and Edman degradation. Using positive ion MALDI mass spectrometry, we obtained molecular weight information for 5 peptide × dT_{20} complexes. By subtracting the mass of the oligonucleotide dT_{20} from these experimentally obtained masses we calculated the molecular weights of the tryptic peptide part of the complexes. Correlation of the mass of a tryptic fragment to a stretch of the Ung amino acid sequence served to map the DNA binding region of the enzyme.

The molecular weights determined by MALDI mass spectrometry are consistent with the idea that the tryptic peptides [T18], [T19], [T18/19] and [T6] are covalently attached to the oligonucleotide. The [T18] × dT_{20} complex consistently showed up in all the chromatographic fractions analyzed. Edman degradation identified all these peptides and additionally the peptide T1. The T18 peptide was the most abundant as judged from the automated sequencer data.

The stretches of amino acids in the protein that are covered by the conjugated peptides must define part of the DNA binding site of the protein because UV-crosslinking takes place only between nucleic acids and amino acid residues that are in close contact (48). We therefore conclude that the DNA binding domain of Ung is defined by amino acids 58-80 (T6) and 185-213 (T18 and T19) (3) (Figure 43). These results demonstrate that photochemical crosslinking in combination with proteolytic digestion and mass spectrometric analysis can be used to map nucleic acid binding sites in proteins.

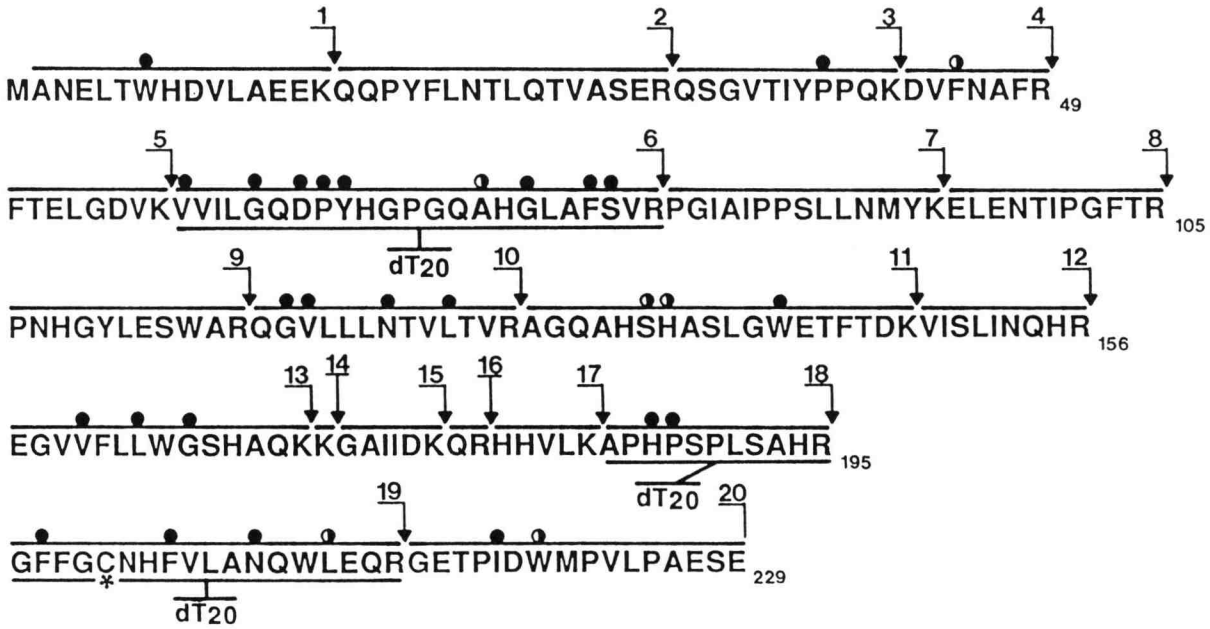


Figure 43. Mapping of regions that constitute the DNA binding site of Uracil-DNA glycosylase.

Tryptic cleavage sites are indicated with arrows and numbers. Conserved residues are indicated with circles. The regions that were found to crosslink to dT_{20} are underlined.

SOME GENERAL COMMENTS ON MASS SPECTROMETRIC ANALYSIS OF COVALENT PEPTIDE- AND PROTEIN-OLIGO(dT) HYBRIDS

In this study we have used MALDI and ESI mass spectrometry to analyze peptide-linker-dT_6, peptide-linker-dT_{10},and peptide × dT_{20} hybrids (the latter only by MALDI MS), and we have used MALDI MS to analyze intact protein × dT_{20} complexes. Some overall principles have emerged that I will discuss in the following.

In MALDI MS of covalent peptide-oligo(dT) hybrids, the quality of the spectra seems to degrade as the size of the oligonucleotide portion of the complex increases, that is, the resolution and the mass accuracy decreases. This is especially evident for the Ung peptide × dT_{20} complexes. However, in the latter case the complexes were always present in mixtures containing at least 3 components. The spectra obtained from purified peptide-linker-dT_6 and -dT_{10}, and also for the HPLC purified Ung [T18] × dT_{20} complex indicates that sample purity is a major factor in the attainable spectrum quality. This is a well known fact for peptide and protein analysis, but it seem to be even more important for the analysis of covalent peptide-nucleic acid complexes by MALDI. Discovery of other matrix/solvent combinations might help improve the sensitivity and thereby spectrum quality.

The dependency of sensitivity on the relative contribution of the oligonucleotide to the total mass of peptide-linker-oligo(dT) hybrids was also evident in ESI MS. The sensitivity of ESI analysis of the peptide-linker-dT_6 hybrid in the positive ion mode was better than in the negative ion mode, whereas the response from the peptide-linker-dT_{10} hybrid was similar in positive vs. negative mode, probably due to increased length of the nucleotide and to increased alkali ion adduction.

The relative contribution of polypeptide and oligonucleotide to the total mass of the complex can be used as a guideline as to which matrix should be used for MALDI analysis. In the case of protein × dT_{20} complexes the protein is the major component of the molecule and therefore a "protein-matrix", such as sinapinic acid, should be used in combination with a good solubilizing agent, such as formic acid. In the case of peptide × dT_{20}, the oligonucleotide is the major component of the molecule and a "oligonucleotide-matrix", such as THAP, should be employed.

Our results also suggest that ESI MS can be used most efficiently in cases where the peptide and oligonucleotide parts of a hybrid are similar in size or where the latter is smaller than the former. For both MALDI and ESI one should try to use as short an oligonucleotide as the protein-nucleic acid complex under study allows, in order to obtain the best mass spectrometric data, as judged by resolution, accuracy, and sensitivity.

PEPTIDE MAPPING OF TRANSCRIPTION ERMINATION FACTOR RHO

Transcription termination factor Rho was digested with a variety of proteases (Lys-C, trypsin, Asp-N) and with a chemical reagent (CNBr). MALDI MS peptide mapping and microbore or capillary HPLC-ESI MS peptide mapping were used to monitor the time course of the digestions. The discussion here is restricted to the data obtained from Lys-C digestions of Rho. However, it also applies in general to the other peptide maps, presented in Appendix A. The Lys-C digest of Rho produced the best and most complete peptide maps of Rho since the detected peptide fragments covered the largest proportion of the Rho polypeptide sequence.

MALDI mass analysis of the peptide mixture after Lys-C digestion of Rho indicated that the cleavage had gone to completion since no high molecular weight (> MW 10,000) species were detected under the conditions used. Assigning the experimentally obtained *m/z* values to Lys-C peptides was straightforward (Table X) and shows that the MALDI peptide map covers approximately 75% of the Rho sequence; only low molecular weight species and a few higher MW peptides are missing. MALDI MS is very useful because it can provide a peptide map from a digest mixture in a matter of minutes using only very little sample (5-10 pmoles in 0.5 μL solution).

Capillary LC-ESI MS could detect more Rho Lys-C peptides than MALDI MS due to the separation achieved by HPLC prior to MS analysis. The LC-MS peptide map covered over 90% of the Rho sequence, making it the method of choice for detailed investigations of modified proteins. In our current setup, LC-MS is more time and sample consuming than MALDI MS due to the one hour acetonitrile gradient used for peptide elution. However, the use of smaller diameter columns should improve sensitivity, and the use of perfusion particles as packing material should allow steeper gradients to be used for peptide elution without degrading the separation efficiency, and, thus, should decrease the time needed to run an LC-MS analysis.

Alkylation of the single Cys residue in Rho by iodoacetic acid or iodoacetamide improved the recovery of the Cys-containing peptide but did not affect the overall cleavage efficiency of Rho by Lys-C protease as compared to the non-alkylated sample. In order to avoid artifacts introduced by the alkylation agent on the putative peptide × 4sUDP complex we omitted the alkylation step.

Although microbore HPLC-UV, MALDI MS, and capillary LC-ESI MS all provided excellent Lys-C peptide maps of Rho, UV-irradiated Rho, Rho + 4sUDP, and UV-irradiated Rho + 4sUDP samples, we were not able to detect any peptide × 4sUDP complexes. In those cases where a change in elution time and a change in mass of a peptide was evident in the UV-irradiated Rho + 4sUDP sample, we could assign the new mass to an oxidized form of a Lys-C peptide. Oxidized peptides were not detected to any significant extent in the control samples. We are not certain about the origin of this oxidation phenomenon. It seems to be present only in the Rho sample

that contains 4sUDP and that has been UV-irradiated. Another new peptide with a mass of 4,750 was also present in the UV-irradiated Rho + 4sUDP sample. The nearest match to this peptide is the Cys containing peptide L13 (MW 4,653). A mass increase of 97 can maybe explained by a phosphate (H_2PO_4), although this remains very speculative.

In the hope that peptide × 4sUDP complexes, if present, would form negative ions, we attempted to use negative ion MALDI MS and LC-MS peptide mapping without any success. We observed many of the peptide ions that were detected in the positive ion mode, but with a substantial reduction overall in sensitivity.

The absence of an ion-signal in MALDI that can be attributted to a peptide × 4sUDP complex could be due to several factors: First, the complexity of the peptide mixture may cause the signal from peptide × 4sUDP to be hidden under another peptide ion signal or in the low mass background. Second, the peptide × 4sUDP signal may be weak or absent due to the changed ionization properties of the adducted peptide or due to factors related to sample preparation for MALDI. Third, the peptide × 4sUDP adduct might not have been formed. Fourth, the peptide × 4sUDP photoproduct might be unstable under the analysis conditions used and would be present, therefore, in a rearranged, fragmented, or degraded form.

The molecular weight obtained for intact Rho × 4sUDP is slightly above the expected mass, which might lend support to the hypothesis that non-specific labeling or oxidation of Rho in the presence of 4sUDP is substantial. However, care must be taken in interpreting this set of data because the mass deviation could arise from matrix adduction or alkali metal adduction as discussed above.

Although we have attempted in a variety of ways to detect adducted peptides, no obvious identification of Rho peptide-nucleic acid hybrids have been attained. We speculate that the intact Rho × 4sUDP covalent complex is stable due to protection of the bound photoaffinity-label by the polypeptide backbone. Proteolytic degradation of Rho removes the protective environment of the photolabel and causes it to degrade. The difficulties that we have encountered in our attempts to analyze Rho × 4sUDP can therefore seem most attributable to the unknown nature and stability of the photochemical adduct. We therefore need to study simpler model systems in order to learn more about the photochemical reactivity of 4sU towards proteins and peptides.

In order to learn more about the behaviour of 4sUDP under mass spectrometric analysis, we employed ESI tandem mass spectrometry to study the fragmentation pattern of this molecule. Mass spectrometry of 4sUDP in the negative ion mode was straightforward. Tandem mass spectrometry of 4sUDP revealed that loss of phosphate and base are the most facile fragmentation pathways. This feature should be explored for specific and sensitive LC-MS detection of peptide nuclceic acid complexes using collisional excitation in the electrospray

source/inlet region of the mass spectrometer and monitoring of masses of phosphate and nucleobase ions as discussed previously.

Chapter 8: Conclusion

We have explored a novel approach based on photochemical crosslinking and mass spectrometric mapping techniques for identifying amino acid-base contacts in protein-nucleic acid complexes. The results presented in this thesis, the first ever of this sort to ever be obtained, are encouraging because they establish the feasibility of the proposed protocol and provide a good foundation for future studies. Using three different protein-nucleic acid complexes and synthetic peptide-oligonucleotide hybrids, we have demonstrated that intact nucleoprotein and nucleopeptide complexes are amendable to mass spectrometric analysis in the form of matrix-assisted laser desorption/ionization and electrospray ionization.

We have not applied the strategy exhaustively to any one model system although the Rho protein was anticipated to be a good candidate. However, taken as a whole, the combination of the results we have obtained, indicate that all steps of the proposed protocol are feasible, although we have not yet demonstrated the use of tandem mass spectrometry to localize the crosslinked amino acid residue in a peptide.

The application of MALDI mass spectrometric techniques to studies of intact UV-crosslinked nucleoprotein complexes is feasible over a broad range of conditions. Molecular weight information could be obtained using only low picomole amounts of protein × nucleic acid complexes. In the case of the gp32 × dT_{20} complex the photocrosslinked species was detected by direct MALDI MS analysis of the irradiated solution. This was also the case for the Rho × 4sUDP complex. Purified Ung × dT_{20} complex was analyzed by MALDI and the molecular weight information used to ascertain the purity and identity of the complex and stoichiometry of the constituents.

The accuracies of molecular weight determination of protein × nucleic acid complexes were in the range of 0.1-0.2%, which is only slightly worse than the accuracy that can be obtained in analyses of pure proteins but is still an order of magnitude better than what can be obtained by any other currently practiced method. The reduced accuracy may be attributable to the added complexity of the analyte, to adduction of alkali ions or matrix molecules to the crosslinked complexes, or to nonlinear mass calibration due to the relatively high laser fluences used to produce protein × nucleic acid ions. The mass spectra obtained from samples containing protein × nucleic acid complexes can be used to establish the presence, purity, and stoichiometry of these species.

Successful MALDI mass spectrometry of UV-crosslinked protein-nucleic acid complexes is as critically dependent on the choice of solvents and additives as it is on the primary matrix compound; this observation may have general import for MALDI mass analysis of biomolecules. The use of good protein solubilizing agents, such as formic acid, was found to be necessary in order to obtain MALDI mass spectra of protein × nucleic acid complexes. Not surprisingly, the adduction of a nucleic acid to a protein changes the solubility profile of the latter, which to some extent might reflect the conformational changes brought about by binding of nucleic acid to protein.

MALDI mass spectrometry is a good tool for monitoring proteolytic digestions of proteins. It provides a mass specific peptide map that reflects the degree of protein digestion at given time points. However, peptide × nucleic acid complexes were not detected by direct peptide mixture analysis by MALDI MS after proteolytic digestion of protein × nucleic acid complexes; this was illustrated by the tryptic digest of Ung × dT_{20} and the endoproteinase Lys-C digest of Rho × 4sUDP. Sample preparation is probably the critical step. Finding matrix/solvent combinations that allow peptide × nucleic acid hybrids to co-crystallize with the matrix while excluding non-liganded peptides might circumvent this problem and allow MALDI MS to be used for rapid identification of crosslinked peptide-nucleic acid species.

The use of mass spectrometry for the characterization of purified peptide-nucleic acid complexes has been demonstrated for the first time. This was illustrated by MALDI and ESI mass spectrometric analysis of synthetic, covalent peptide-linker-oligodeoxythymidylic acid hybrids. The mass spectra provided molecular weight information that confirmed the structure of the covalent peptide-nucleic acid hybrids. Tandem mass spectrometry was used to obtain more detailed molecular weight and structural information, mainly on the nucleic acid portion of the molecule.

The application of our strategy to a biological system was demonstrated by identifying the DNA binding domains of *E. coli* uracil-DNA glycosylase. By using a combination of MALDI mass spectrometric analysis and Edman degradation, we have provided direct evidence that the Ung tryptic peptides T6 and T18/19 form part of the nucleic acid binding domain of Ung. The results presented here demonstrate for the first time that low picomole amounts of UV-crosslinked hybrids derived from a protein can be analyzed by MALDI mass spectrometry and that the information obtained this way can be used to identify nucleic acid binding domains in proteins. The MALDI data was obtained using approximately 100 fold less sample than was consumed by a parallel Edman sequencing experiment. That these regions encompasses some of the most conserved residues in Ung species from different sources adds the weight of biological relevance to our findings. It has recently been demonstrated that Ung inhibitor protein (Ugi) does not bind to a UV-crosslinked Ung × dT_{20} complex and that the presence of Ugi prevents UV-crosslinking

of Ung to dT_{20} (3). This indicates that the DNA binding site of Ung is overlaps the Ugi binding site.

We have illustrated the use of LC-ESI-MS for peptide mapping of Rho after digestion with Lys-C protease. The sensitivity and accuracy of this method for peptide detection and identification should prove it a valuable tool for future studies of protein × nucleic acid complexes. We detected almost all of the expected Rho peptide fragments, i.e., covering 92% of the Rho sequence. Unfortunately, no species were detected that could be assigned to peptide × 4sUDP hybrids. This might be due to one of several reasons: The peptide adduct is not amendable to analysis under the conditions used; the photolabel is unstable under the conditions used; or the crosslinked species was not formed in the first place. Interestingly, in the Rho + 4sUDP + UV sample many ion signals were detected that could be tentatively assigned to Rho Lys-C peptides assuming that oxidation of Met residues that would lead to a mass increase of 16 Da per Met residue had taken place. These signals were not detected in Rho + UV or Rho + 4sUDP control samples. At present the reason for this apparent oxidation is unknown, although it seems to be dependent on UV-irradiation of Rho in the presence of 4sUDP. In any case, the observation of the latter phenomenon illustrates the usefulness of mass spectrometric detection of HPLC effluents.

MALDI MS provided good sensitivity and fairly good mass accuracy for analysis of protein × nucleic acid complexes. The matrix solvent combination used for MALDI analysis of intact protein × nucleic acid complexes reflected the protein-like nature of the complexes, whereas the matrix/solvent combination used for MALDI analysis of peptide × dT_{20} complexes reflected the oligonucleotide-like nature of these complexes. The quality of the MALDI spectra, as illustrated by spectra obtained from peptide-linker-dT_6, peptide-linker-dT_{10}, and Ung peptide × dT_{20}, decreased as the length of the oligonucleotide increases. This effect was also observed in ESI MS analysis of the synthetic peptide-oligonucleotide hybrids.

FUTURE DIRECTIONS

Overall, this study has provided very encouraging results that establish the feasibility of the proposed protocol for characterization of protein-nucleic acid interactions. Of course, much more work is required in order to optimize each step of the analytical strategy. The nature of this future research is taken up briefly in this final section.

This work has of necessity focused on the analytical task, i.e. the mass spectrometric analysis, involved in characterizing photochemically crosslinked protein × nucleic acid complexes. However, our understanding of the photochemical crosslinking process, that is, the types of

bonds formed, stabilities, etc., in the protein × nucleic acid complexes is sadly lacking. In order to be able to take full advantage of photochemical crosslinking techniques, we need to know more about the photochemistry involved. There have not yet been any comparisons made between the photoproducts generated by continous versus pulsed UV irradiation. The gene 32 protein or the uracil-DNA glycosylase might be good systems to study because we already have some experience working with these DNA binding proteins.

A major improvement of the overall strategy would be brought about by increasing the crosslinking yields. This would allow less protein and nucleic acid to be consumed in the experiments. The combined use of photoaffinity-labeling and laser generated pulsed UV-crosslinking has been demonstrated to provide high yields (65) and should also minimize side-reactions.

A large fraction of the experimental effort put into this investigation has been to find the best sample preparation conditions for MALDI mass analysis of protein × nucleic acid complexes and peptide × nucleic acid complexes. At present, there are no guidelines as to how to systematically design the optimal matrix/solvent combination for a given class of compound, although solubility seems to be a major factor to consider for protein analysis. Finding a means to predictably prepare samples for MALDI, especially for direct detection of peptide × nucleic acid complexes by MALDI peptide mapping, should be a high priority for future research on the protocol.

We have not yet attempted to use ESI MS for mass analysis of intact protein × nucleic acid complexes. ESI MS provides better mass accuracy and higher resolution than MALDI, but is more sensitive to involatile contaminants such as buffers, alkali ions, and denaturing agents which necessitates desalting prior to MS analysis. The sensitivity of ESI MS for oligonucleotide analysis is lower than MALDI which might be a drawback when limited sample amounts are available. However, ESI analysis of protein × mononucleotide complexes should be possible using LC-MS because the nucleotide will be much smaller than the protein and therefore cause very little perturbation of the protein-like character of the complex. Methods to quantitavely remove "excess" oligonucleotide after UV crosslinking that would facilitate improved mass spectrometric detection of intact crosslinked complexes and also improve the chances of getting sequence information by tandem MS after isolation of the peptide × nucleic acid complex should be investigated.

Successful peptide mapping by LC-MS needs to be demonstrated for protein × nucleic acid complexes. Developing nucleic acid specific detection methods, such as LC-MS with collisional excitation and mass specific monitoring for loss of phosphate or nucleobase might be the most straightforward and sensitive approach to identify peptide × nucleic acid hybrids in complex mixtures.

Despite the shortcomings of the current state-of-the-art, as discussed above, the successful identification of the DNA-binding domain of Ung by our protocol should encourage immediate application of the protocol to other protein-nucleic acid systems.

We believe that our first results on mass spectrometric characterization of protein-nucleic acid interactions provide a good incentive to pursue the above issues.

Slid, men vid
Ting ta'r tid

Piet Hein

THE END

Bibliography

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

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

3. S. E. Bennett, O. N. Jensen, D. F. Barofsky, and D. W. Mosbaugh. "UV catalyzed crosslinking of Escherichia coli Uracil-DNA Glycosylase to DNA." *J. Biol. Chem.* **269**, 21870-21879 (1994).

4. O. N. Jensen, S. Kulkarni, J. Aldrich, S. E. Bennett, D. W. Mosbaugh, and D. F. Barofsky. "Characterization of covalent peptide-oligonucleotide hybrids by mass spectrometry." *Manuscript in preparation* , (199x).

5. N. C. Seeman, J. M. Rosenberg, and A. Rich. "Sequence-specific recognition of double-helical nucleic acids by proteins." *Proc. Natl. Acad. Sci. USA* **73**, 804-808 (1976).

6. T. A. Steitz. "Structural studies of protein-nucleic acid interaction: The sources of sequence specific binding." *Q. Rev. Biophys.* **23**, 205-280 (1990).

7. R. A. Anderson and J. E. Coleman. "Physicochemical properties of DNA binding proteins: Gene 32 protein of T4 and *Escherichia coli* unwinding protein." *Biochemistry* **14**, 5485-5491 (1975).

8. J. E. Coleman and I. M. Armitage. "Tyrosyl-base-phenylalanyl interchelation in gene 5 protein-DNA complexes: Proton nuclear magnetic resonance of selectively deuterated gene 5 protein." *Biochemistry* **17**, 5038-5045 (1978).

9. P. H. von Hippel. "Protein-DNA recognition: New perspectives and underlying themes." *Science* **263**, 769-770 (1994).

10. R. S. Spolar and T. Record Jr. "Coupling of local folding to site-specific binding of proteins to DNA." *Science* **263**, 777-784 (1994).

11. S. K. Burley. "DNA-binding motifs from eucaryotic transcription factors." *Curr. Opinion Struct. Biol.* **4**, 3-11 (1994).

12. S. C. Harrison. "A structural taxonomy of DNA-binding proteins." *Nature* **353**, 715-719 (1991).

13. C. O. Pabo and R. T. Sauer. "Transcription factors: Structural families and principles of DNA recognition." *Ann. Rev. Biochem.* **61**, 1053-1095 (1992).

14. P. S. Freemont, A. N. Lane, and M. R. Sanderson. "Structural aspects of protein-DNA recognition." *Biochem. J.* **278**, 1-23 (1991).

15. R. G. Brennan and B. W. Matthews. "The helix-turn-helix DNA binding motif." *J. Biol. Chem.* **264**, 1903-1906 (1989).

16. R. G. Brennan and B. W. Matthews. "Structural basis of DNA-protein recognition." *Trends Biochem. Sci.* **14**, ? (1989).

17. C. Branden and J. Tooze. "Introduction to protein structure." (Garland Publishing, Inc., New York, 1991).

18. S. C. Harrison and A. K. Aggarwal. "DNA recognition by proteins with the helix-turn-helix motif." *Annu. Rev. Biochem.* **59**, 933-969 (1990).

19. B. E. Raumann, B. M. Brown, and R. T. Sauer. "Major groove DNA recognition by β-sheets: the ribbon-helix-helix family of gene regulatory proteins." *Curr. Opinion Struct. Biol.* **4**, 36-43 (1994).

20. T. B. Kornberg. "Understanding the homeodomain." *J. Biol. Chem.* **268**, 26813-26816 (1993).

21. P. E. Wright. "POU domains and homeodomains." *Curr. Opinion Struct. Biol.* **4**, 22-27 (1994).

22. A. Klug and D. Rhodes. "Zinc fingers: A novel protein motif for nucleic acid recognition." *Trends Biochem. Sci.* **12**, 464-469 (1987).

23. M. Schmiedeskamp and R. E. Klevit. "Zinc finger diversity." *Curr. Opinion Struct. Biol.* **4**, 28-35 (1994).

24. N. P. Pavletich and C. O. Pabo. "Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 A." *Science* **252**, 809-817 (1991).

25. J. M. Berg. "Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules." *J. Biol. Chem.* **265**, 6513-6516 (1990).

26. N. P. Pavletich and C. O. Pabo. "Crystal structure of a five-finger GLI-DNA complex: New perspectives on Zinc fingers." *Science* **261**, 1701-1707 (1993).

27. T. Ellenberger. "Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains." *Curr. Opinion Struct. Biol.* **4**, 12-21 (1994).

28. E. K. O'Shea, J. D. Klemm, P. S. Kim, and T. Alber. "X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil." *Science* **254**, 539-544 (1991).

29. J. W. R. Schwabe and D. Rhodes. "Beyond zinc fingers: Steroid hormone receptors have a novel structural motif for DNA recognition." *Trends Biochem. Sci.* **16**, 291-296 (1991).

30. R. Marmorstein, M. Carey, M. Ptashne, and S. C. Harrison. "DNA recognition by GAL4: Structure of a protein-DNA complex." *Nature* **356**, 408-414 (1992).

31. R. K. Wierenga, P. Terpstra, and W. G. J. Hol. "Prediction of the occurence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint." *J. Mol. Biol.* **187**, 101-107 (1986).

32. M. Saraste, P. S. Sibbald, and A. Wittinghofer. "The P-loop – a common motif in ATP and GTP binding proteins." *Trends Biochem. Sci.* **11**, 430-434 (1990).

33. J. Carey. "Gel retardation." in *Methods Enzymol.* R. T. Sauer, Ed. (Academic Press, San Diego, 1991), vol. 208, pp. 103.

34. R. H. Ebright. "Identification of amino acid-base pair contacts by genetic methods." in *Methods Enzymol.* R. T. Sauer, Ed. (Academic Press, San Diego, 1991), vol. 208, pp. 620.

35. H. Echols. "Nucleoprotein structures initiating DNA replication, transcription, and site-specific recombination." *J. Biol. Chem.* **265**, 14697-14700 (1990).

36. M. Dodson and H. Echols. "Electron Microscopy of protein-DNA complexes." in *Methods Enzymol.* R. T. Sauer, Ed. (Academic Press, San Diego, 1991), vol. 208, pp. 168.

37. S. B. Zimmerman. "Macromolecular crowding: Biochemical, biophysical, and physiological consequences." *Annu. Rev. Biophys. Biomol. Struct.* **22**, 27-65 (1993).

38. K. C. Smith. "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**, 157-163 (1962).

39. P. Alexander and H. Moroson. "Cross-linking of deoxyribonucleic acid to protein following ultraviolet irradiation of different cells." *Nature* **194**, 882-883 (1962).

40. N. A. Simukova and E. I. Budowsky. "Conversion of non-covalent interactions in nucleoproteins into covalent bonds: UV-induced formation of polynucleotide-protein crosslinks in bacteriophage Sd virions." *FEBS Letters* **38**, 299-303 (1974).

41. A. Havron and J. Sperling. "Specificity of photochemical crosslinking in protein-nucleic acid complexes: Identification of the interacting residues in RNase-pyrimidine nucleotide complex." *Biochemistry* **16**, 5631-5635 (1977).

42. S. Y. Wang, Ed., *Photochemistry and photobiology of nucleic acids. I. Chemistry*, vol. 1 (Academic Press, New York, 1976).

43. E. I. Budowsky and G. G. Abdurashidova. "Polynucleotide-protein cross-links induced by ultraviolet light and their use for structural investigation of nucleoproteins." *Prog. Nucl. Acid Res. Molec. Biol.* **37**, 1-65 (1989).

44. M. D. Shetlar. "Cross-linking of proteins to nucleic acids by ultraviolet light." in *Photochem. Photobiol. Rev.* K. C. Smith, Ed. (Plenum Press, New York, 1980), vol. 5, pp. 105-198.

45. I. Saito and H. Sugiyama. "Photoreactions of nucleic acids and their constituents with amino acids and related compounds." in *Bioorganic photochemistry* H. Morrison, Ed. (John Wiley & Sons, New York, 1990), vol. 1, pp. 317-340.

46. M. D. Shetlar, J. Christensen, and K. Hom. "Photochemical addition of amino acids and peptides to DNA." *Photochem. Photobiol.* **39**, 125-133 (1984).

47. A. A. Shaw, A. M. Falick, and M. D. Shetlar. "Photoreactions of thymine and thymidine with N-acetyltyrosine." *Biochemistry* **31**, 10976-10983 (1992).

48. K. R. Williams and W. H. Konigsberg. "Identification of amino acid residues at interface of protein-nucleic acid complexes by photochemical cross-linking." in *Methods Enzymol.* R. T. Sauer, Ed. 1991), vol. 208, pp. 516-539.

49. J. W. Hockensmith, W. L. Kubasek, W. R. Vorachek, E. M. Evertsz, and P. H. von Hippel. "Laser cross-linking of protein-nucleic acid complexes." in *Methods Enzymol.* R. T. Sauer, Ed. (Academic Press, San Diego, 1991), vol. 208, pp. 211-236.

50. I. G. Pashev, S. I. Dimitrov, and D. Angelov. "Cross-linking proteins to nucleic acids by ultraviolet laser irradiation." *Trends Biochem. Sci.* **16**, 323-326 (1991).

51. C. E. Harrison, D. H. Turner, and D. C. Hinkle. "Laser cross-linking of E. Coli RNA polymerase and T7 DNA." *Nucleic Acids Res.* **10**, 2399-2414 (1982).

52. D. Angelov, V. Y. Stefanovsky, S. I. Dimitrov, V. R. Russanova, E. Keskinova, and I. G. Panchev. "Protein-DNA cross-linking in reconstituted nucleohistone, nuclei and whole cells by picosecond UV laser irradiation." *Nucleic Acids Res.* **16**, 4525-4538 (1988).

53. V. Y. Stefanovsky, S. I. Dimitrov, V. R. Russanova, D. Angelov, and I. G. Pashev. "Laser-induced cross-linking of histones to DNA in chromatin and core particles: Implications in studying histone-DNA interactions." *Nucleic Acids Res.* **17**, 10069-10081 (1989).

54. V. Y. Stefanovsky, S. I. Dimitrov, D. Angelov, and I. G. Pashev. "Interactions of acetylated histones with DNA as revealed by UV laser induced histone-DNA cross-linking." *Biochem. Biophys. Res. Commun.* **164**, 304-310 (1989).

55. S. I. Dimitrov, V. Y. Stefanovsky, L. Karagyozov, D. Angelov, and I. G. Pashev. "The enhancers and promoters of the Xenopus laevis ribosomal spacer are associated with histones upon active transcription of the ribosomal genes." *Nucleic Acids Res.* **18**, 6393-6397 (1990).

56. E. I. Budowsky, M. S. Axentyeva, G. G. Abdurashidova, N. A. Simukova, and L. B. Rubin. "Induction of polynucleotide-protein cross-linkages by ultraviolet irradiation. Pecularities of the high-intensity laser pulse irradiation." *Eur. J. Biochem.* **159**, 95-101 (1986).

57. L. D. Mesner and J. W. Hockensmith. "Probing the energetics of oligo(dT)-poly(dA) by laser crosslinking." *Proc. Natl. Acad.Sci. USA* **89**, 2521-2525 (1992).

58. J. W. Hockensmith, W. L. Kubasek, W. R. Vorachek, and P. H. von Hippel. "Laser cross-linking of nucleic acids to proteins." *J. Biol. Chem.* **261**, 3512-3518 (1986).

59. J. W. Hockensmith, W. L. Kubasek, W. R. Vorachek, and P. H. von Hippel. "Laser cross-linking of proteins to nucleic acids. I. Examining physical parameters of protein-nucleic acid complexes." *J. Biol. Chem.* **268**, 15712-15720 (1993).

60. J. W. Hockensmith, W. L. Kubasek, E. M. Evertsz, L. D. Mesner, and P. H. von Hippel. "Laser cross-linking of proteins to nucleic acids. II. Interactions of the bacteriophage T4 DNA replication polymerase accessory proteins complex with DNA." *J. Biol. Chem.* **268**, 15721-15730 (1993).

61. A. Favre. "4-thiouridine as an intrinsic photoaffinity probe of nucleic acid structure and interaction." in *Bioorganic Photochemistry* H. Morrison, Ed. (John Wiley & Sons, New York, 1990) pp. 379-425.

62. B. Bartholomew, B. R. Braun, G. R. Kassavetis, and E. P. Geidusek. "Probing close DNA contacts of RNA polymerase III transcription complexes with the photoactive nucleoside 4-thiodeoxythymidine." *J. Biol. Chem.* **269**, 18090-18095 (1994).

63. E. E. Blatter, Y. W. Ebright, and R. H. Ebright. "Identification of an amino acid base contact in the GCN4-DNA complex by bromouracil-mediated photocrosslinking." *Nature* **359**, 650-652 (1992).

64. Q. Dong, E. E. Blatter, Y. W. Ebright, K. Bister, and R. H. Ebright. "Identification of amino acid-base contacts in the Myc-DNA complex by site-specific bromouracil mediated photocrosslinking." *EMBO J.* **13**, 200-204 (1994).

65. M. C. Willis, B. J. Hicke, O. C. Uhlenbeck, T. R. Cech, and T. H. Koch. "Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins." *Science* **262**, 1255-1257 (1993).

66. P. R. Paradiso, Y. Nakashima, and W. Konigsberg. "Photochemical cross-linking of protein-nucleic acid complexes. The attachment of the fd gene 5 protein to fd DNA." *J. Biol. Chem.* **254**, 4739-4744 (1979).

67. Y. Shamoo, K. R. Williams, and W. H. Konigsberg. "Photochemical crosslinking of bacteriophage T4 single-stranded DNA-binding protein (gp32) to oligo-p$(dT)_8$: Identification of Phenylalanine-183 as the site of crosslinking." *Proteins: Struct. Funct. Genet.* **4**, 1-6 (1988).

68. B. M. Merrill, K. R. Williams, J. W. Chase, and W. H. Konigsberg. "Photochemical cross-linking of the Escherichia coli single-stranded DNA-binding protein to oligodeoxynucleotides. Identification of Phenylalanine 60 as the site of crosslinking." *J. Biol. Chem.* **259**, 10850-10856 (1984).

69. B. M. Merrill, K. L. Stone, F. Cobianchi, S. H. Wilson, and K. R. Williams. "Phenylalanines that are conserved among several RNA-binding proteins form a part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein." *J. Biol. Chem.* **263**, 3307-3313 (1988).

70. A. Basu, K. A. Ahluwalia, S. Basu, and M. J. Modak. "Identification of the primer binding domain in human immunodeficiency virus reverse transcriptase." *Biochemistry* **31**, 616-623 (1992).

71. T. L. Capson, S. J. Benkovic, and N. G. Nossal. "Protein-DNA crosslinking demonstrates stepwise ATP-dependent assembly of T4 DNA polymerase and its accessory proteins on the primer template." *Cell* **65**, 249-258 (1991).

72. M. Katouzian-Safadi, B. Blazy, J.-Y. Cremet, J.-P. Le Caer, J. Rossier, and M. Charlier. "Photo-crosslinking of CRP to nonspecific DNA in the absence of cAMP. DNA interacts with both the N- and C-terminal parts of the protein." *Biochemistry* **32**, 1770-1773 (1993).

73. R. Prasad, A. Kumar, S. G. Widen, J. R. Casas-Finet, and S. H. Wilson. "Identification of residues in the single-stranded DNA-binding site of the 8-kDa domain of rat DNA polymerase β by UV cross-linking." *J. Biol. Chem.* **268**, 22746-22755 (1993).

74. N. Sheng, E. B. Mougey, S. Kelly, and D. Dennis. "Wheat germ and yeast RNA polymerase II: Photoaffinity labeling by 4-thiouracil 5'-monophoshate positioned uniquely at the 3' end of an enzyme-bound [^{32}P]-containing transcript." *Biochemistry* **32**, 2248-2253 (1993).

75. B. D. Shivanna, M. R. Mejillano, T. D. Williams, and R. H. Himes. "Exchangeable GTP binding site of β-tubulin. Identification of Cys 12 as the major site of crosslinking by direct photoaffinity labeling." *J. Biol. Chem.* **268**, 127-132 (1993).

76. P. R. Schimmel. "Approaches to understanding the mechanisms of specific protein-transfer RNA interactions." *Acc. Chem. Res.* **10**, 411-418 (1977).

77. E. I. Budowsky. "UV-induced formation of polynucleotide-protein cross-linkages as a tool for investigation of the nucleoprotein structure and functions." in *Trends in Photobiology* C. Helene, M. Charlier, T. Montenay-Garestier, G. Laustriat, Eds. (Plenum Press, New York, 1982) pp. 93-108.

78. V. Pandey, N. Kaushik, and M. J. Modak. "Photoaffinity labeling of DNA template-primer binding site in Escherichia coli DNA polymerase I." *J. Biol. Chem.* **269**, 21828-21834 (1994).

79. S. Chen, T. D. Lee, K. Legesse, and J. E. Shively. "Identification of the arylazido-β-alanyl-NAD^+-modified site in rabbit muscle glyceraldehyde-3-phosphate dehydrogenase by microsequencing and fast atom bombardment mass spectrometry." *Biochemistry* **25**, 5391-5395 (1986).

80. T. P. O'Connor and J. E. Coleman. "Proton nuclear magnetic resonance (500 MHz) of mono-, di-, tri-, and tetradeoxynucleotide complexes of gene 5 protein." *Biochemistry* **22**, 3375-3381 (1983).

81. M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, and J. Kraut. "Crystal structure of rat DNA polymerase β: Evidence for a common polymerase mechanism." *Science* **264**, 1930-1935 (1994).

82. H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, and J. Kraut. "Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP." *Science* **264**, 1891-1903 (1994).

83. J. A. McCloskey. "Mass Spectrometry." Methods Enzymol. (Academic Press, San Diego, 1990), vol. 193.

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

85. M. J. Geisow. "Mass measurement at high molecular weight - New tools for biotechnologists." *Trends Biotech.* **10**, 432-441 (1992).

86. K. Biemann. "Mass Spectrometry of Peptides and Proteins." *Ann. Rev. Biochem.* **61**, 977-1010 (1992).

87. M. W. Senko and F. W. McLafferty. "Mass spectrometry of macromolecules: has its time come now?" *Annu. Rev. Biophys. Biomol. Struct.* **23**, 763-785 (1994).

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

89. C. K. Meng, M. Mann, and J. B. Fenn. "Of protons or proteins." *Z. Phys. D.* **10**, 361-368 (1988).

90. M. Karas and F. Hillenkamp. "Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons." *Anal. Chem.* **60**, 2299-2301 (1988).

91. F. Hillenkamp, M. Karas, R. C. Beavis, and B. T. Chait. "Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers." *Anal. Chem.* **63**, 1193A-1202A (1991).

92. T. M. Billeci and J. T. Stults. "Tryptic mapping of recombinant proteins by matrix-assisted laser desorption/ionization mass spectrometry." *Anal. Chem.* **65**, 1709-1716 (1993).

93. B. T. Chait, R. Wang, R. C. Beavis, and S. B. H. Kent. "Protein ladder sequencing." *Science* **262**, 89-82 (1993).

94. K. O. Bornsen, M. Schar, E. Gassman, and V. Steiner. "Analytical applications of matrix-assisted laser desorption and ionization mass spectrometry." *Biological Mass Spectrometry* **20**, 471-478 (1991).

95. R. C. Beavis and B. T. Chait. "Rapid, sensitive analysis of protein mixtures by mass spectrometry." *Proc. Natl. Acad. Sci. USA* **87**, 6873-6877 (1991).

96. T. Huth-Fehre. "Matrix-assisted laser desorption mass spectrometry of oligodeoxythymidylic acids." *Rapid Commun. Mass Spectrom.* **6**, 209-213 (1992).

97. B. Spengler, Y. Pang, R. J. Cotter, and L.-S. Kan. "Molecular weight determination of underivatized oligodeoxyribonucleotides by positive-ion matrix-assisted ultraviolet laser desorption mass spectrometry." *Rapid Commun. Mass Spectrom.* **4**, 99-102 (1990).

98. E. Nordhoff, A. Ingendoh, R. Cramer, A. Overberg, B. Stahl, M. Karas, F. Hillenkamp, and P. F. Crain. "Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared." *Rapid. Commun. Mass Spectrom.* **6**, 771-776 (1992).

99. K. Schneider and B. T. Chait. "Matrix-assisted laser desorption mass spectrometry of homopolymer oligodeoxyribonucleotides. Influence of base composotion on the mass spectrometric response." *Org. Mass Spectrom.* **28**, 1353-1361 (1993).

100. U. Pieles, W. Zurchner, M. Schar, and H. E. Moser. "Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: A powerful tool for the mass and sequence analysis of natural and modified oligonucleotides." *Nucleic Acids Res.* **21**, 3191-3196 (1993).

101. E. Nordhoff, R. Cramer, M. Karas, F. Hillenkamp, F. Kirpekar, K. Kristiansen, and P. Roepstorff. "Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry." *Nucleic Acids Res.* **21**, 3347-3357 (1993).

102. K. J. Wu, A. Steding, and C. H. Becker. "Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix." *Rapid Commun. Mass Spectrom.* **7**, 142-146 (1993).

103. K. J. Wu, T. A. Shaler, and C. H. Becker. "Time-of-flight mass spectrometry of underivatized single-stranded DNA oligomers by matrix assisted laser desorption." *Anal. Chem.* **66**, 1637-1645 (1994).

104. R. C. Beavis and B. T. Chait. "High-accuracy molecular mass determination of proteins using matrix-assisted laser deorption mass spectrometry." *Anal. Chem.* **62**, 1836-1840 (1990).

105. O. Vorm and M. Mann. "Improved mass accuracy in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of peptides." *J. Am. Soc. Mass Spectrom.* **In Press**, (1994).

106. R. C. Beavis and B. T. Chait. "Matrix-assisted laser desorption mass spectrometry using 355 nm radiation." *Rapid Commun. Mass Spectrom.* **3**, 436-439 (1989).

107. R. C. Beavis and B. T. Chait. "Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins." *Rapid Commun. Mass Spectrom.* **3**, 432-435 (1989).

108. R. C. Beavis, T. Chaudhary, and B. T. Chait. "α-Cyano-4-hydroxycinnamic acid as a matrix for matrix assisted laser desorption mass spectrometry." *Org. Mass Spectrom.* **27**, 156-158 (1992).

109. K. Strupat, M. Karas, and F. Hillenkamp. "2,5-dihydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry." *Int. J. Mass Spectrom. Ion Proc.* **111**, 89-102 (1991).

110. M. Karas. "Fundamental aspects of matrix-assisted laser desorption ionization mass spectrometry." in *Fundamental processes in sputtering of atoms and molecules (SPUT92)* P. Sigmund, Ed. (The Royal Danish Academy of Sciences and Letters/Munksgaard, Copenhagen, 1993) pp. 623-641.

111. G. A. Clegg and M. Dole. "Molecular beams of macro-ions. III. Zein and polyvinylpyrrolidone." *Biopolymers* **10**, 821-826 (1971).

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

113. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse. "Electrospray ionization - principles and practice." *Mass Spectrom. Rev.* **9**, 37-70 (1990).

114. T. R. Covey, R. F. Bonner, B. I. Shushan, and J. Henion. "The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry." *Rapid Commun. Mass Spectrom.* **2**, 249-256 (1988).

115. S. A. Carr, M. J. Huddleston, and M. F. Bean. "Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry." *Protein Sci.* **2**, 183-196 (1993).

116. M. J. Huddleston, R. S. Annan, M. F. Bean, and S. A. Carr. "Selective detection of phosphopeptides in complex mixtures by electrospray liquid chromatography/mass spectrometry." *J. Am. Soc. Mass Spectrom.* **4**, 710-717 (1993).

117. P. Thibault, C. Paris, and S. Pleasance. "Analysis of peptides and proteins by capillary electrophoresis/mass spectrometry using acidic buffers and coated capillaries." *Rapid Commun. Mass Spectrom.* **5**, 484-490 (1991).

118. J. A. Loo, H. R. Udseth, and R. D. Smith. "Peptide and protein analysis by electrospray ionization mass spectrometry and capillary electrophoresis-mass spectrometry." *Anal. Biochem.* **179**, 404-412 (1989).

119. E. D. Lee, W. Muck, J. D. Henion, and T. R. Covey. "On-line capillary zone electrophoresis-ion spray tandem mass spectrometry for determination of dynorphins." *J. Chromatogr.* **458**, 313-321 (1988).

120. H. R. Morris, M. Panico, and G. W. Taylor. "FAB-mapping of recombinant-DNA protein products." *Biochem. Biophys. Res. Commun.* **117**, 299 (1983).

121. C. N. McEwen and B. S. Larsen, Eds., *Mass Spectrometry of Biological Materials* (Marcel Dekker, New York, 1990).

122. S. A. Carr, M. E. Hemling, M. F. Bean, and G. D. Roberts. "Integration of mass spectrometry in analytical biotechnology." *Anal. Chem.* **63**, 2802-2824 (1991).

123. V. Ling, A. W. Guzzetta, E. Canova-Davis, J. T. Stults, W. S. Hancock, T. R. Covey, and B. I. Shushan. "Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high performance liquid chromatography-electrospray ionization mass spectrometry." *Anal. Chem.* **63**, 2909-2915 (1991).

124. P. R. Griffin, J. A. Coffman, L. E. Hood, and J. R. Yates III. "Structural analysis of proteins by capillary HPLC electrospray tandem mass spectrometry." *Int. J. Mass Spectrom. Ion Proc.* **111**, 131-149 (1991).

125. E. C. Huang and J. D. Henion. "Packed-capillary liquid chromatography/ionspray tandem mass spectrometry determination of biomolecules." *Anal. Chem.* **63**, 732-739 (1991).

126. D. B. Kassel, M. A. Luther, D. H. Willard, S. P. Fulton, and J.-P. Salzman. "Rapid purification, separation and identification of proteins and enzyme digests using packed capillary perfusion column LC and LC/MS." *Techniques in Protein Chemistry* **IV**, 55-64 (1993).

127. M. T. Davis and T. D. Lee. "Analysis of peptide mixtures by capillary high performance liquid chromatography: A practical guide to small-scale separations." *Protein Sci.* **1**, 935-944 (1992).

128. F. W. McLafferty. "Tandem Mass Spectrometry." (John Wiley & Sons, New York, 1983).

129. D. F. Hunt, J. R. Yates, J. Shabanowitz, S. Winston, and C. R. Hauer. "Peptide sequencing by tandem mass spectrometry." *Proc. Natl. Acad. Sci. USA* **83**, 6233-6237 (1986).

130. K. Biemann and H. A. Scoble. "Characterization by tandem mass spectrometry of structural modifications in proteins." *Science* **237**, 992 (1987).

131. J. A. Loo, C. G. Edmonds, and R. D. Smith. "Primary sequence information form intact proteins by electrospray ionization tandem mass spectrometry." *Science* **248**, 201-204 (1990).

132. J. A. Loo, C. G. Edmonds, and R. D. Smith. "Tandem mass spectrometry of very large molecules: Serum albumin sequence information from multiply charged ions formed by electrospray ionization." *Anal. Chem.* **63**, 2488-2499 (1991).

133. M. W. Senko, S. C. Beu, and F. W. McLafferty. "High-resolution tandem mass spectrometry of carbonic anhydrase." *Anal. Chem.* **66**, 415-417 (1994).

134. P. Roepstorff and J. Fohlmann. "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." *Biomed. Mass Spectrom.* **11**, 601 (1984).

135. D. F. Hunt, R. A. Henderson, J. Shabanowitz, K. Sakaguchi, H. P. Michel, N. Sevilir, A. L. Cox, E. Appella, and V. E. Engelhard. "Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry." *Science* **25**, 1261-1263 (1992).

136. D. F. Hunt, H. Michel, T. A. Dickinson, J. Shabanowitz, A. L. Cox, K. Sakaguchi, E. Appella, H. M. Grey, and A. Sette. "Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-A^d." *Science* **256**, 1817-1820 (1992).

137. R. A. Henderson, H. P. Michel, K. Sakaguchi, J. Shabanowitz, E. Appella, D. F. Hunt, and V. E. Engelhard. "HLA-A2.1-associated peptides from a mutant cell line: A second pathway of antigen presentation." *Science* **25**, 1264-1266 (1992).

138. C. F. Morris, H. Hama-Inaba, D. Mace, N. K. sinha, and B. Alberts. "Purification of gene 43, 44, 45, and 62 proteins of the bacteriophage T4 DNA replication apparatus." *J. Biol. Chem.* **254**, 6787 (1979).

139. N. G. Nossal and B. M. Peterlin. "DNA replication by bacteriophage T4 proteins." *J. Biol. Chem.* **254**, 6032 (1979).

140. M. C. Young, M. K. Reddy, and P. H. von Hippel. "Structure and function of the bacteriophage T4 DNA polymerase holoenzyme." *Biochemistry* **31**, 8675-8690 (1992).

141. J. W. Chase and K. R. Williams. "Single-stranded DNA binding proteins required for DNA replication." *Ann. Rev. Biochem.* **55**, 103-136 (1986).

142. K. R. Williams and W. Konigsberg. "Structural changes in the T4 gene 32 protein induced by DNA and polynucleotides." *J Biol. Chem.* **253**, 2463-2470 (1978).

143. A. Tsugita and J. Hosoda. "DNA binding site of the helix destabilizing protein gp32 from bacteriophage T4." *J. Mol. Biol.* **122**, 255-258 (1978).

144. R. L. Karpel, V. Y. Levin, and B. E. Haley. "Photoaffinity labeling of T4 bacteriophage 32 protein." *J. Biol. Chem.* **262**, 9359-9366 (1987).

145. R. L. Burke, B. M. Alberts, and J. Hosoda. "Proteolytic removal of the COOH terminus of the T4 gene 32 helix-destabilizing protein alters the T4 *in vitro* replication complex." *J. Biol. Chem.* **255**, 11484-11493 (1980).

146. K. B. Krassa, L. S. Green, and L. Gold. "Protein-Protein Interactions with the Acidic COOH Terminus of the Single-Stranded DNA-Binding Protein of the Bacteriophage T4." *Proc. Natl. Acad. Sci., USA* **88**, 4010-4014 (1991).

147. S. G. Nadler, W. J. Roberts, Y. Shamoo, and K. R. Williams. "A novel function for zinc(II) in a nucleic acid-binding protein. Contribution of zinc(II) toward the cooperativity of bacteriophage T4 gene 32 protein binding." *J Biol. Chem.* **265**, 10389-10394 (1990).

148. Y. Shamoo, K. R. Webster, K. R. Williams, and W. H. Konigsberg. "A retrovirus-like zinc domain is essential for translational repression of T4 gene 32." *J. Biol. Chem.* **266**, 7967-7970 (1991).

149. R. V. Prigodich, J. Casas-Finet, K. R. Williams, W. Konigsberg, and J. E. Coleman. "^{1}H NMR (500 MHz) of gene 32 protein-oligonucleotide complexes." *Biochemistry* **23**, 522-529 (1984).

150. R. V. Prigodich, Y. Shamoo, K. R. Williams, J. W. Chase, W. H. Konigsberg, and J. E. Coleman. "1H NMR (500 MHz) identification of aromatic residues of gene 32 protein involved in DNA binding by use of protein containing perdeuterated aromatic residues and by site-directed mutagenesis." *Biochemistry* **25**, 3666-3672 (1986).

151. T. L. Capson, J. A. Peliska, B. F. Kaboord, M. W. Frey, C. Lively, M. Dahlberg, and S. J. Benkovic. "Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4." *Biochemistry* **31**, 10984-10994 (1992).

152. J. W. Hockensmith, W. L. Kubasek, T. A. Cross, M. K. Dolejsi, W. R. Vorachek, and P. H. von Hippel. "Dynamic and static structural studies of the bacteriophage T4-coded DNA replication complex, using UV laser crosslinking." in *DNA Replication and Recombination* R. McMacken, T. J. Kelly, Eds. (Alan R. Liss, New York, 1987), vol. 47, pp. 111.

153. T. D. Yager and P. H. von Hippel. "Transcription elongation and termination in *Escherichia Coli.*" in *The Molecular and Cell Biology of E. coli and S. typhimurium.* F. Neidhardt, Ed. (American Soc. Microbiol., Washington D.C., 1987) pp. 1241-1274.

154. D. G. Bear and D. S. Peabody. "The E. coli Rho protein: an ATPase that terminates transcription." *Trends in Biochem. Sci.* **13**, 343-347 (1988).

155. J. P. Richardson. "Rho-dependent transcription termination." *Biochim. Biophys. Acta* **1048**, 127-138 (1990).

156. T. Platt and J. P. Richardson. "Escherichia coli Rho Factor: Protein and Enzyme of Transcription Termination." in *Transcriptional Regulation* S. L. McKnight, K. R. Yamamoto, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1992), vol. I, pp. 365-388.

157. A. Kornberg and T. A. Baker. "DNA Replication." (W H Freeman & Co., New York, 1991).

158. S. E. Seifried, K. Bjornson, and P. H. von Hippel. "Structure and Assembly of the E. coli Transcription Termination Factor Rho and Its Interactions with RNA. II. Physical Chemical Studies." *J. Mol. Biol.* **221**, 1139-1151 (1991).

159. J. Geiselmann, S. E. Seifried, T. D. Yager, C. Liang, and P. H. von Hippel. "Physical properties of the Escherichia coli transcription termination factor Rho. 2. Quarternary structure of the Rho hexamer." *Biochemistry* **31**, 121-132 (1992).

160. E. P. Gogol, S. E. Seifried, and P. H. von Hippel. "Structure and Assembly of the E. coli Transcription Termination Factor Rho and Its Interactions with RNA. I. Cryo-Electron Microscopy." *J. Mol. Biol.* **221**, 1127-1138 (1991).

161. A. J. Dombroski, C. A. Brennan, P. Spear, and T. Platt. "Site-directed alterations in the ATP-binding domain of Rho protein affects its activities as a termination factor." *J. Biol. Chem.* **263**, 18802-18809 (1988).

162. A. J. Dombroski, J. R. LaDine, R. L. Cross, and T. Platt. "The ATP binding site on Rho protein. Affinity labeling of Lys^{181} by pyridoxal 5'-diphospho-5'-adenosine." *J. Biol. Chem.* **263**, 18810-18815 (1988).

163. I. O and B. L. Stitt. "8-azido-ATP inactivation of Escherichia coli transcription termination factor Rho. Modification of one subunit inactivates the hexamer." *J. Biol. Chem.* **269**, 5009-5015 (1994).

164. J. Geiselmann and P. H. von Hippel. "Functional interactions of ligand cofactors with Escherichia coli transcription termination factor Rho. I. Binding of ATP." *Protein Sci.* **1**, 850-860 (1992).

165. S. E. Seifried, J. B. Easton, and P. H. von Hippel. "ATPase activity of transcription termination factor Rho: The functional dimer model." *Proc. Natl. Acad. Sci. USA* **89**, 10454-10459 (1992).

166. Y. Wang and P. H. von Hippel. "Escherichia coli transcription termination factor Rho. II. Binding of oligonucleotide cofactors." *J. Biol. Chem.* **268**, 13947-13955 (1993).

167. J. Geiselmann, T. D. Yager, and P. H. von Hippel. "Functional interactions of ligand cofactors with Escherichia coli transcription termination factor Rho. II. Binding of RNA." *Protein Sci.* **1**, 861-873 (1992).

168. Y. Wang and P. H. von Hippel. "Escherichia coli transcription termination factor Rho. I. ATPase activation by oligonucleotide cofactors." *J. Biol. Chem.* **268**, 13940-13946 (1993).

169. D. W. Mosbaugh and S. E. Bennett. "Uracil-excision DNA repair." *Progr. Nucl. Acid Res. Mol. Biol.* **48**, 315-370 (1994).

170. R. C. Beavis. "M-over-Z software for MALDI data analysis was provided by Dr. R.C. Beavis who is at The Skirball Institute for Biomolecular Medicine, New York University Medical Center, New York" .

171. W. Ens. "The TOFMA software is developed by Professor W. Ens at the Department of Physics, University of Manitoba, Winnipeg, Canada R3T 2N2" .

172. J. T. Stults and J. C. Marsters. "Improved electrospray ionization of synthetic oligonucleotides." *Rapid Commun. Mass Spectrom.* **5**, 259-263 (1991).

173. N. G. Nossal. "DNA synthesis on a double stranded DNA template by the T4 DNA polymerase and the T4 gene 32 DNA unwinding protein." *J. Biol. Chem.* **249**, 5668-5676 (1974).

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

175. S. E. Seifried, Y. Wang, and P. H. von Hippel. "Fluorescent Modification of the Cysteine 202 Residue in *Escherichia coli* Transcription Termination Factor Rho." *J. Biol. Chem.* **263**, 13511-13514 (1988).

176. J. Coste, D. Le-Nguyen, and B. Castro. "PyBOP: A new peptide coupling reagent devoid of toxic by-products." *Tetrahedron Lett.* **31**, 205-208 (1990).

177. B. Bayard, C. Bisbal, and B. Lebleu. "Activation of ribonuclease L by (2'-5')(A_4)-poly(L-Lysine) conjugates in intact cells." *Biochemistry* **25**, 3730-3736 (1986).

178. J. K. Bashkin, R. J. McBeath, A. S. Modak, K. R. Sample, and W. B. Wise. "Synthesis and characterization of oligonucleotide peptides." *J. Org. Chem.* **56**, 3168-3176 (1991).

179. R. Eritja, A. Pons, M. Escarceller, E. Giralt, and F. Albericio. "Synthesis of defined peptide-oligonucleotide hybrids containing a nuclear transport signal sequence." *Tetrahedron* **47**, 4113-4120 (1991).

180. G. Degols, J.-P. Leonetti, M. Benkirane, C. Devaux, and B. Lebleu. "Poly(L-Lysine)-conjugated oligonucleotides promote sequence specific inhibition of acute HIV-1 infection." *Antisense Res. Dev.* **2**, 293-301 (1992).

181. T. Zhu, C.-S. Tung, K. J. Breslauer, W. A. Dickerhof, and S. Stein. "Preparation and physical properties of conjugates of oligodeoxynucleotides with poly(δ)ornithine peptides." *Antisense Res. Dev.* **3**, 349-356 (1993).

182. J. A. Loo, D. D. Holsworth, and R. S. Root-Bernstein. "Use of electrospray ionization mass spectrometry to probe antisense peptide interactions." *Biol. Mass Spectrom.* **23**, 6-12 (1994).

183. T. R. Baker, T. Keough, R. L. M. Dobson, T. A. Riley, J. A. Hasselfield, and P. E. Hesselberth. "Antisense DNA oligonucleotides I: The use of electrospray ionization tandem mass spectrometry for the sequence verification of methylphosphonate oligodeoxyribonucleotides." *Rapid Commun. Mass Spectrom.* **7**, 190-194 (1993).

184. T. Keough, T. R. Baker, R. L. M. Dobson, M. P. Lacey, T. A. Riley, J. A. Hasselfield, and P. E. Hesselberth. "Antisense DNA oligonucleotides II: The use of matrix-assisted laser desorption/ionization mass spectrometry for the sequence verification of methylphosphonate oligodeoxyribonucleotides." *Rapid Commun. Mass Spectrom.* **7**, 195-200 (1993).

APPENDICES

Appendix A: Mass spectrometric peptide mapping of Rho

DIGESTION OF RHO WITH TRYPSIN

Based on Rho's amino acid sequence, cleavage at Lys and Arg residues by trypsin should produce at least 55 peptides (Figure 44). Many of these tryptic peptides are detected by MALDI peptide mapping (Figure 45) or by microbore LC-MS peptide mapping (Figure 46); these results are summarized in Table XII. The MALDI mass map covers approximately 60% of the Rho amino acid sequence. Many small peptides are not detected. The Cys containing peptide T26 produces a weak signal; however, a stronger signal from this peptide can be obtained after *S*-alkylation with, for example, iodoacetamide.

Comparison of the MALDI mass maps obtained by analysis of UV-irradiated Rho and of a UV-irradiated mixture of Rho and 4sUDP reveals a few changes in peptide mass that could correspond to oxidative damage or alteration as discussed in the Results and Discussion sections. The tryptic peptide T26 exhibits a mass increase of approximately 100 u. This mass increase was not detected in a sample that was S-alkylated suggesting that the modifying reaction involves the Cys residue in this peptide.

Microbore LC-MS analysis of Rho tryptic peptides using formic acid/acetonitrile as the mobile phase produces very good peptide maps that indicate complete digestion (Figure 46 and Table XII). LC-MS of tryptic peptides using ammonium acetate as the aqueous mobile phase, for obtaining better recovery of nucleotide adducted peptides, appeared similar to those obtained with formic acid/acetonitrile, but the chromatogram did not reveal any new information (data not shown).

Rho has many Arg and Lys residues, and therefore, trypsin produces many small fragments that may escape mass spectrometric detection. For this reason, Endoproteinase Lys-C is superior to trypsin for mapping Rho, as described in the Results section.

MNLTELK▾NTP VSELITLGEN MGLENLAR▾MR▾ K▾QDIIFAILK▾ 40
T1 T2 T3 T4 T5

QHAK▾SGEDIF GDGVLEILQD GFGFLR▾SADS SYLAGPDDIY 80
T6 T7 T8

VSPSQIR▾R▾FN LR▾TGDTISGK▾ IR PPK▾EGER▾Y FALLK▾VNEVN 120
T9 T10 T11 T12 T13 T14

FDKPENAR▾NK▾ ILFENLTPLH ANSR▾LR▾MER▾G NGSTK▾DLTAR 160
T15 T16 T17 T18 T19 T20 T21

VLDLASPIGR▾ GQR▾GLIVAPP K▾AGK▾TMLLQN IAQSIAYNHP 200
T22 T23 T24 T25 T26

DCVLMVLLID ERPEEVTEMQ R▾LVK▾GEVVAS TFDEPASR▾HV 240
T27 T28

QVAEMVIEK▾A K▾R▾LVEHK▾K▾DV IILLDSITR▾L AR▾AYNTVVP 280
T29 T30 T31 T32 T33 T34 T35 T36

SGK▾VLTGGVD ANALHRPK▾R▾F FGAAR▾NVEEG GSLTIIATAL 320
T37 T38 T39 T40

IDTGSK▾MDEV IYEEFK▾GTGN MELHLSR▾K▾IA EK▾R▾VFPAIDY 360
T41 T42 T43 T44 T45 T46

NR▾SGTR▾K▾EEL LTTQEELQK▾M WILR▾K▾IIHPM GEIDAMEFLI 400
T47 T48 T49 T50 T51 T52

NK▾LAMTK▾TND DFFEMMK▾R▾S 419
T53 T54 T55 T56

Figure 44. Tryptic cleavage sites in transcription termination factor Rho. *The peptides are named T1-T56, beginning at the N-terminus. Cleavage positions C-terminal to Lys and Arg residues are indicated with a v. Arg-Pro sequences (underlined XRPX) are not cleaved.*

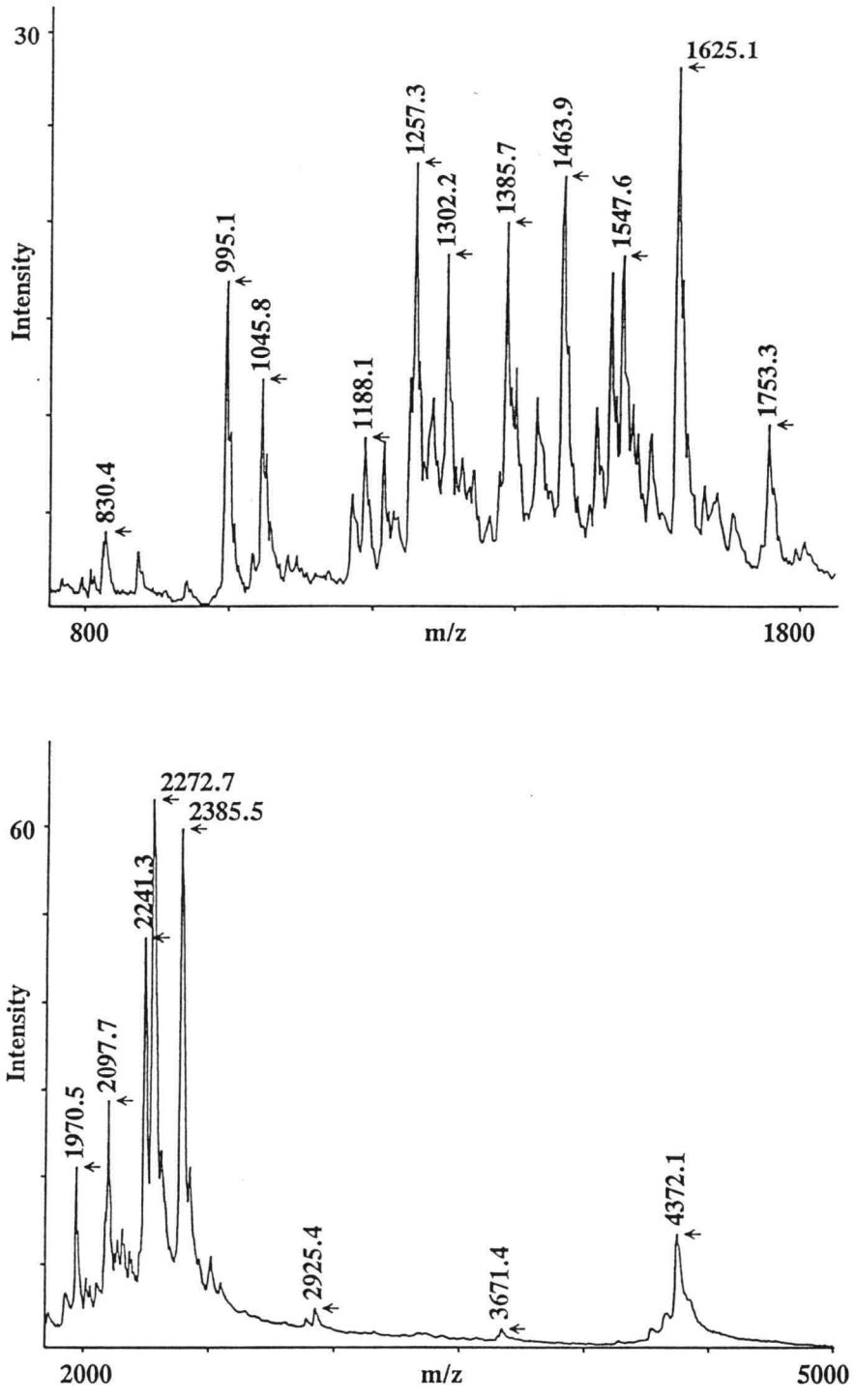


Figure 45. MALDI MS peptide map of a Rho tryptic digest.
The matrix consisted of 4HCCA in 0.1% TFA and acetonitrile (2:1). Sum of 30 laser shots.

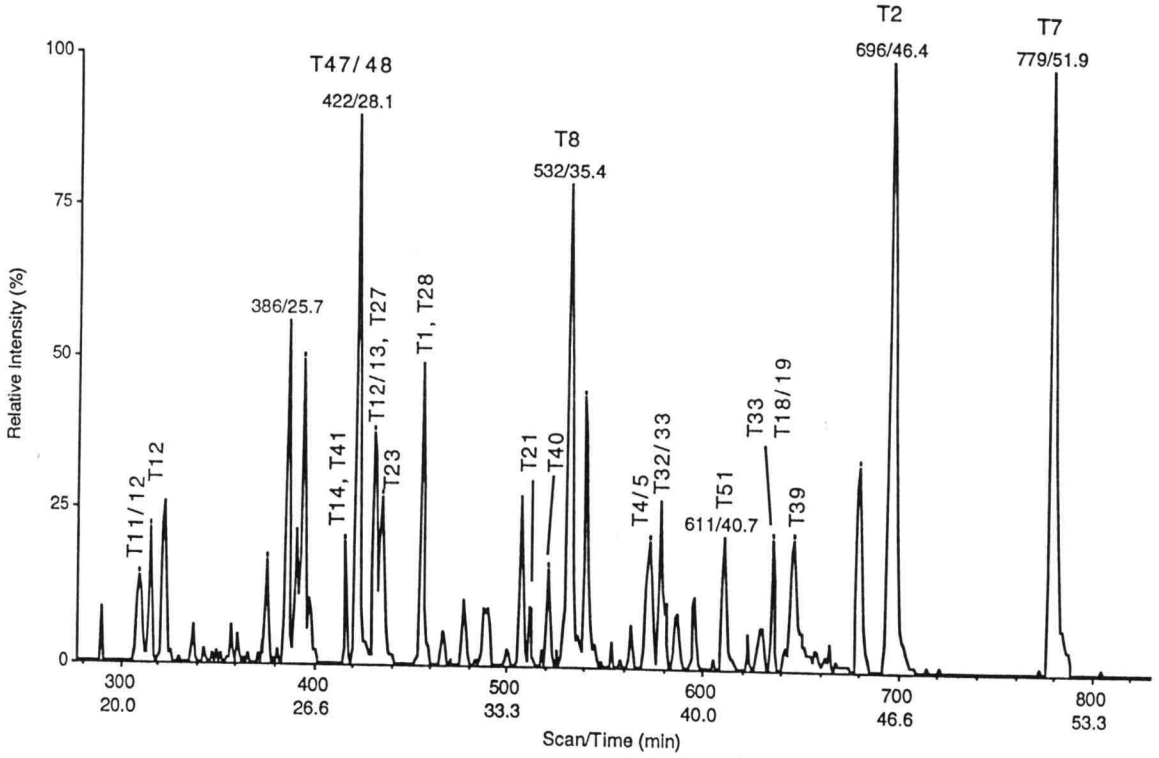


Figure 46. Microbore LC-MS TIC chromatogram of a Rho tryptic digest.

Table XII. MW determination by tryptic mapping of Rho by MALDI MS and LC-ESI-MS.

No.	Peptide position	Fragment MH+	MALDI MH+	ESI MH+
1	1 - 7	848.46	n.d.	849.0
2	8 - 28	2271.17	2272.2	2271.2
3	29 - 30	306.16	-	-
4	31 - 31	147.11	-	-
4,5	31-40	1188.74	1189.0	1189.4
5	32 - 40	1060.64	n.d.	1061.1
6	41 - 44	483.27	-	-
7	45 - 66	2384.18	2386.3	2385.7
8	67 - 87	2241.07	2241.0	2241.2
9	88-88	175.12	-	-
9,10	88- 92	705.42	705.3	705.9
10	89-92	549.31	(550.8)bkgnd	549.8
10,11	89-100	1308.69	n.d.	n.d.
11	93 - 100	778.39	n.d.	778.9
12	101 - 105	610.40	610.4	610.4
12,13	101-109	1081.6	1082.1	1081.1
13	106 - 109	490.23	-	-
14	110 - 115	754.45	n.d.	754.9
15	116 - 128	1531.75	1532.0	(1533.1)
16	129 - 130	261.16	-	-
17	131 - 144	1624.88	1625.0 (Cal)	no!
18	145 - 146	288.20	-	-
18,19	145-149	704.39	(705.3)	704.9
19	147 - 149	435.20	-	-
20	150 - 155		-	-
21	156-160		-	-
22	161 - 170	1040.61	1042.1	1041.1
23	171 - 173	360.20	-	-
24	174 - 181	794.51	n.d.	794.9
24,25	174-184	1050.67	1050.9	1050.6

Table XII (continued)

25	182 - 184	275.17	-	-
26	185 - 221	4313.14	4382.5	4373.1
27	222 - 224	359.27	-	-
27,28	222-238	1804.94	1804.8	n.d.
28	225 - 238	1464.70	1464.7	1465.3
29	239 - 249	1282.68	1282.8	1283.9
29,30	239-251	1481.82	n.d.	1482.3
30	250 - 251	218.15	-	-
31	252 - 252	175.12	-	-
32	253 - 257	625.37	n.d.	626.3
33	258 - 258	147.11	-	-
33,34	258-269	1385.84	1386.6	1386.3
34	259 - 269	1257.74	1258.2	1258.2
35	270 - 272	359.24	-	-
36	273 - 283	1106.58	n.d.	1106.6
37	284 - 298	1547.87	(1548.4)	n.d.
38	299-299	175.1	-	-
38,39	299 - 305	824.45	825.2	825
39	300 - 305	668.35	n.d.	668.4
40	306 - 326	2089.10	n.d.	2090.1
41	327 - 336	1302.59	1302.2	1303.2
42	337 - 347	1214.60	1215.0	1215.4
42,43	337-348	1342.69	(1340.3)	n.d.
43	348 - 348	147.11	-	-
43,44	348-352	588.37	(590.3)	587.8
44	349 - 352	460.28	-	-
44,45	349-353	616.38	(618.9)	615.8
45	353 - 353	175.12	-	-
45,46	353-362	1250.66	1250.3	n.d.
46	354 - 362	1094.56	1094.8	1095.1
47	363 - 366	420.22	-	-
48	367 - 367	147.11	-	-

Table XII (continued)

T48,49	367-379	1588.84	(1586.6)	1589.9
49	368 - 379	1460.75	n.d.	n.d.
50	380 - 384	718.41	n.d.	n.d.
51	385 - 385	147.11	-	-
T51,52	385-402	2099.1	2098.6	2099.1
52	386 - 402	1971.01	1971.4	1971.0
53	403 - 407	563.32	n.d.	n.d.
54	408 - 417	1277.52	(1277.1)	1277.7
54,55	408-418	1433.62	(1429.4)	n.d.
55	418 - 418	175.12	-	-
56	419 - 419	106.05	-	-

Peptides below mass 500 are indicated with a - (hyphen). Peptide masses not detected are indicated with n.d. Masses enclosed in parenthesis are uncertain or from very weak signals.

DIGESTION OF RHO BY ENDOPROTEINASE ASP-N

Digestion of Rho by Endoproteinase Asp-N should produce 24 peptides (Figure 47). MALDI MS peptide maps of Rho after digestion with Endoproteinase Asp-N appear very nice (Figure 48) although digestion did not go to completion. Peptides in the putative ATP binding region appeared in the MALDI map as the combined, partially oxidized peptide D12/13. The presence of an Asp-Pro sequence in this region apparently interferes with the enzymatic cleavage.

Microbore LC-MS analysis of Rho Asp-N peptides also indicates that the reaction did not go to completion (data not shown). The ratio of enzyme to substrate (1:50) might have been too low.

LC-MS analysis of Rho Asp-N peptides using ammonium acetate/acetonitrile as the mobile phase did not produce good peptide maps because of the low intensity of peptide ions. This might be due to the presence of a negatively charged, N-terminal Asp residue in each peptide.

Endoproteinase Asp-N is a potentially good enzyme for peptide mapping of Rho, but optimization of the reaction conditions is necessary. Some of the relatively large peptides formed by cleavage with this enzyme are not suited for tandem MS sequencing. LC-MS analysis employing formic acid/acetonitrile as the mobile phase should be attempted.

MNLTELKNTP VSELITLGEN MGLENLARMR KQ▾DIIFAILK 40
D1 D2

QHAKSGE▾DIF G▾DGVLEILQ▾D GFGFLRSA▾DS SYLAGP▾D▾DIY 80
D3 D4 D5 D6 D7

VSPSQIRRFN LRTG▾DTISGK IRPPKEGERY FALLKVNEVN 120
D8 D9

F▾DKPENARNK ILFENLTPLH ANSRLRMERG NGSTK▾DLTAR 160
D10 D11

VL▾DLASPIGR GQRGLIVAPP KAGKTMLLQN IAQSIAYNHP 200
D12

▾DCVLMVLLI▾D ERPEEVTEMQ RLVKGEVVAS TF▾DEPASRHV 240
D13 D14

QVAEMVIEKA KRLVEHKK▾DV IILL▾DSITRL ARAYNTVVPA 280
D15 D16 D17

SGKVLTGGV▾D ANALHRPKRF FGAARNVEEG GSLTIIATAL 320
D18

I▾DTGSKM▾DEV IYEEFKGTGN MELHLSRKIA EKRVFPAI▾DY 360
D19 D20

NRSGTRKEEL LTTQEELQKM WILRKIIHPM GEI▾DAMEFLI 400
D21

NKLAMTKTN▾D ▾DFFEMMKRS 419
D22 D23 D24

Figure 47. Endoproteinase Asp-N cleavage sites in transcription termination factor Rho.
Cleavage positions N-terminal to Asp residues are indicated with a v. The peptides are named D1-D24, beginning at the N-terminus.

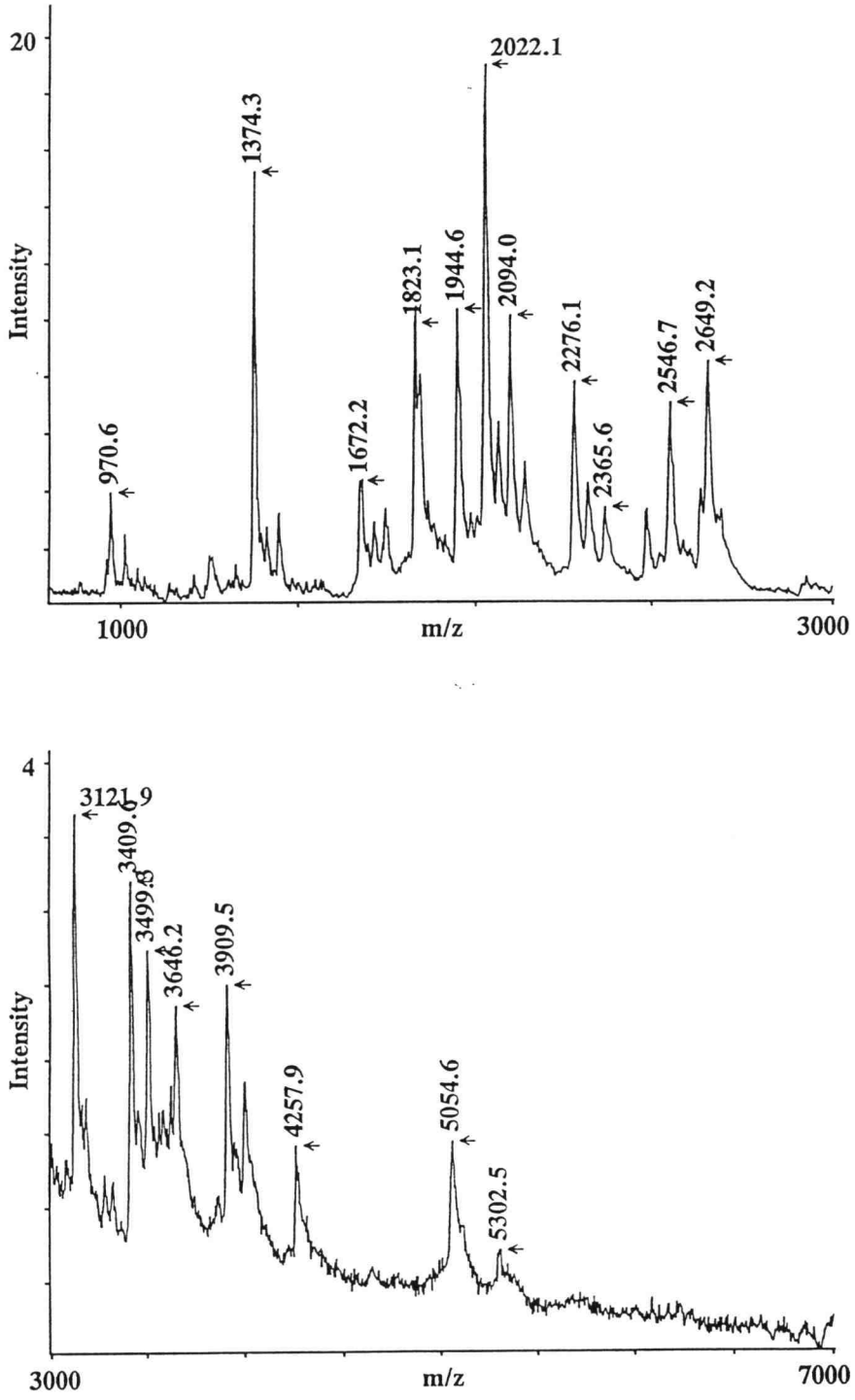


Figure 48. MALDI peptide mass map of an Asp-N digest of Rho.

Table XIII. MALDI mass spectrometric peptide mapping of an Endoproteinase Asp-N digest of Rho.

No.	Peptide position	Calculated MH^+	MALDI MH^+
1	1 - 32	3643.90	(3669.9)
2	33 - 47	1669.93	1670.2
3	48 - 51	451.22	-
4	52 - 59	886.49	-
5	60 - 68	969.48	970.1
6	69 - 76	809.37	-
7	77 - 77	134.05	n.d.
7,8	77-94	2137.1	2136.5
8	78 - 94	2022.09	2023.2
9	95 - 121	3120.68	3121.3
10	122 - 155	3907.99	3909.2
11	156 - 162	787.47	787.8
12	163 - 200	3996.20	-
12,13	163-209	4996.8	5024.0
		5054.8[a]	5054.6[a]
13	201 - 209	1018.57	-
14	210 - 232	2649.32	2649.3
15	233 - 258	3027.65	3025.1
16	259 - 264	685.45	-
17	265 - 289	2545.41	2545.7
18	290 - 321	3408.85	3410.1
19	322 - 327	638.28	-
20	328 - 358	3619.89	3621.1
21	359 - 393	4256.23	4258.9
22	394 - 409	1839.94	1839.2
23	410 - 410	134.05	n.d.
23,24	410-419	1305.6	1306.3
24	411 - 419	1190.53	1191.1

a) S-alkylated (amidomethylated) Cys.

CYANOGEN BROMIDE CLEAVAGE OF RHO

Cleavage of Rho at Met residues by CNBr should produce 17 peptides (Figure 49). Upon cleavage, Met is converted into homoserine lactone which is slowly hydrolyzed into homoserine.

It was relatively difficult to obtain good MALDI MS peptide maps of Rho CNBr fragments, especially from the UV-irradiated Rho sample that contained 4sUDP. The reason for this is not known but might be related to the size and solubility of the peptides and maybe to the oxidative damage by UV-irradiation of Rho, as discussed in a previous section. Oxidation of Met by UV-light in the presence of 4sUDP might explain poor CNBr cleavage of UV-irradiated mixture Rho and 4sUDP relative to the non-irradiated mixture of Rho and 4sUDP. However, using MALDI we were able to detect most of the peptides that contain a C-terminal homoserine lactone residue (Figure 50 and Table XIV).

Peptide mapping by LC-MS produced good data for Rho + 4sUDP (Figure 51) while the Rho+4sUDP+UV sample produced fairly weak signals. In the former case almost all the homoserine lactone peptides were detected (Table XIV).

The acidic conditions required for CNBr cleavage (70% formic acid) might not be desireable since the photochemical bond between Rho and 4sUDP could be labile. More research is needed to address this topic. Furthermore, many of the fragments are too large to be sequenced by tandem MS. Under some circumstances, it might be desireable to produce a less complex peptide mixture for locating the amino acid stretch that has been modified by the photoaffinity-label. In such cases, CNBr cleavage is preferable, assuming that the label is stable under the experimental conditions used for cleavage.

M▾NLTELKNTP VSELITLGEN M▾GLENLARM▾R KQDIIFAILK 40
C1 C2 C3

QHAKSGEDIF GDGVLEILQD GFGFLRSADS SYLAGPDDIY 80
C4

VSPSQIRRFN LRTGDTISGK IRPPKEGERY FALLKVNEVN 120

FDKPENARNK ILFENLTPLH ANSRLRM▾ERG NGSTKDLTAR 160

VLDLASPIGR GQRGLIVAPP KAGKTM▾LLQN IAQSIAYNHP 200
C5 C6

DCVLM▾VLLID ERPEEVTEM▾Q RLVKGEVVAS TFDEPASRHV 240
C7 C8

QVAEM▾VIEKA KRLVEHKKDV IILLDSITRL ARAYNTVVPA 280

SGKVLTGGVD ANALHRPKRF FGAARNVEEG GSLTIIATAL 320
C9

IDTGSKM▾DEV IYEEFKGTGN M▾ELHLSRKIA EKRVFPAIDY 360
C10 C11

NRSGTRKEEL LTTQEELQKM▾ WILRKIIHPM▾ GEIDAM▾EFLI 400
C12 C13

NKLAM▾TKTND DFFEM▾M▾KRS 419
C14 C15 C16 C17

Figure 49. Cyanogen bromide (CNBr) cleavage sites in transcription termination factor Rho.
Cleavage points C-terminal to Met residues are indicated with a v. The peptides are named C1-C17, beginning at the N-terminus. Methionine is converted to homoserine lactone or homoserine by CNBr.

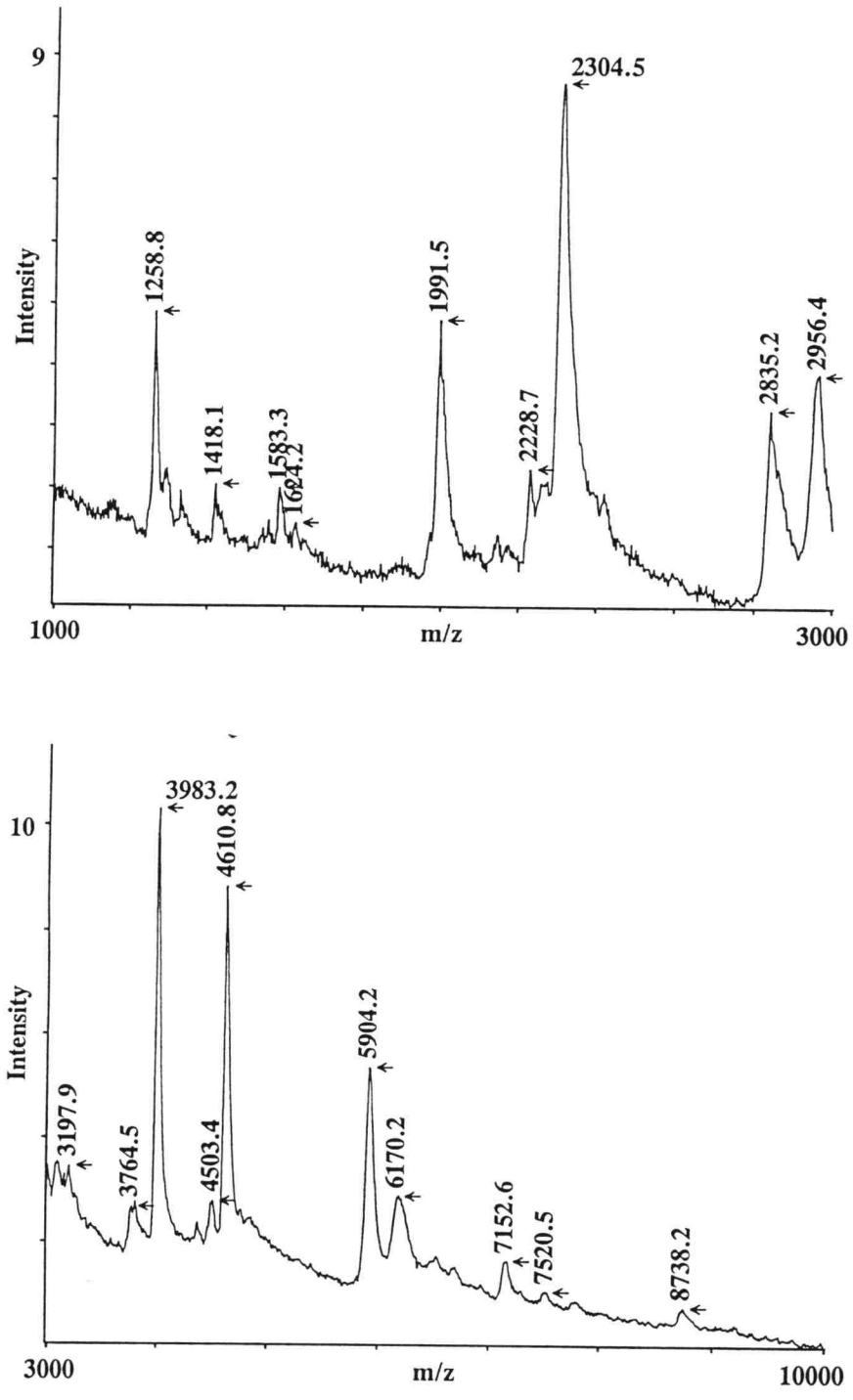


Figure 50. MALDI peptide map of Rho after incubation with CNBr.

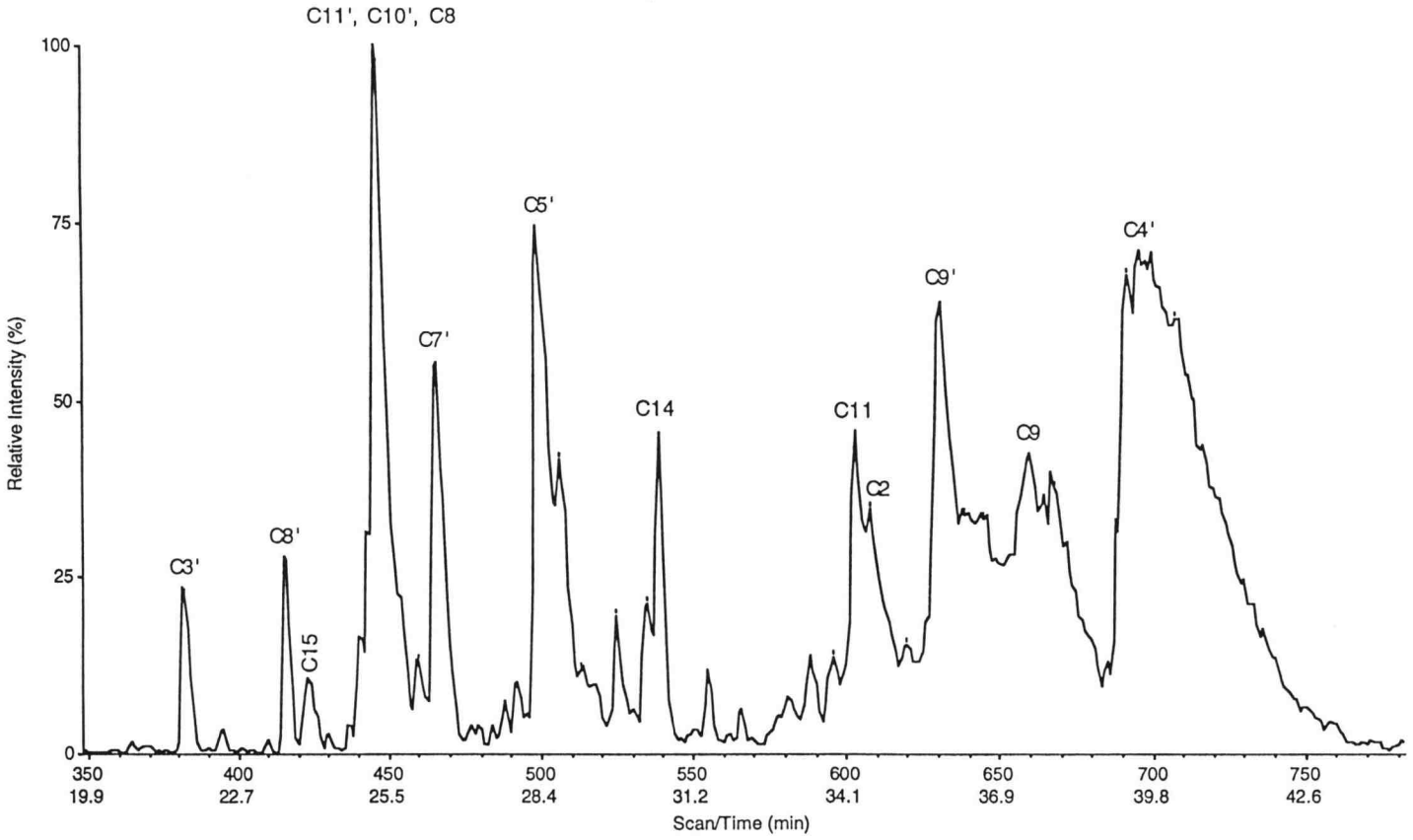


Figure 51. LC-MS peptide map of Rho after cleavage by CNBr.

Table XIV. Molecular weights obtained by MALDI MS and LC-MS peptide mapping of Rho after incubation with cyanogen bromide.

Num.	Peptide position	Calculated MH+	MALDI MH+	ESI MH+
Homoserine Lactone				
1	1 - 1	102.06	-	-
2	2 - 21	2168.15	2169.2	2169.1
3	22 - 29	855.47	-	855.5
4	30 - 147	13353.04	(13340)	13355.7
5	148 - 186	3983.23	3984.1	3984.1
6	187 - 205	2095.06	(2092.8)	?
7	206 - 219	1624.84	(1624.3)	1624.6
7/8	206 - 245	4459.31	-	4460.4
8	220 - 245	2835.48	2835.7	2835.9
9	246 - 327	8731.9	8741.2	8734.7
10	328 - 341	1583.72	1583.3	1582.9
10/11	328 - 380	6192.2	(6172.1)	6192.2
11	342 - 380	4609.48	4612.0	4610
12	381 - 390	1258.78	1258.8	1258.4
13	391 - 396	587.27	(589.9)	585.3 ?
14	397 - 405	1030.59	-	1031.1
14/15	397 - 415	2229.11	2229.0	2229
15	406 - 415	1199.52	-	1199.7
16	416 - 416	102.06	n.d.	n.d.
17	417-419			
Homoserine				
1	1 - 1	120.07	-	-
2	2 - 21	2186.16		
3	22 - 29	873.48		873.5
4	30 - 147	13371.06		13385 ?
8	220 - 245	2853.49	2855.7	2853.xx
10	328 - 341	1601.73		1601.xx
11	342 - 380	4627.49		4628.3
3- 4	22 - 147	14225.52		14258 ?
4- 5	30 - 186	17354.22		17352.9

Appendix B: Vitae

Ole Nørregaard Jensen (ONJ) is a native of Denmark, born in the city of Odense on September 6, 1965. After graduating from high school in 1984, he went to Odense University to study the life sciences. In the summer of 1988, he received a 3-month internship at the laboratory for Protein Chemistry, Novo-Nordisk (Bagsværd, Denmark) to work on mass spectrometric analysis of recombinant human blood-coagulation Factor VII. ONJ received a Bachelors degree in Computer Science in 1988 and his *Cand. scient.* degree (Masters of Science equivalent) in Cell Biology in 1990. His work leading to a thesis entitled *Characterization of Abnormal Human Hemoglobins by Chromatographic and Mass Spectrometric Methods* was carried out under the supervision of Research Professor Peter Roepstorff at the Department of Molecular Biology, Odense University.

In September of 1990, ONJ joined the Department of Biochemistry & Biophysics at Oregon State University (OSU) to begin his graduate studies under the supervision of Professor Douglas F. Barofsky, Department of Agricultural Chemistry, OSU. During this time he became interested in protein-nucleic acid interactions – an interest that led to the present work. This thesis was completed on October 20,1994.

ONJ has received an Alexander von Humboldt Research Fellowship to continue development of his research career at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany.

PUBLICATIONS

1. O. N. JENSEN, P. HØJRUP, AND P. ROEPSTORFF. "Plasma desorption mass spectrometry as a tool in characterization of abnormal proteins. Application to variant human hemoglobins." *Anal. Biochem.* **199**, 175-183 (1991).

2. O. N. JENSEN, P. ROEPSTORFF, B. ROZYNOV, M. HORANYI, J. SZELENYI, S. R. HOLLAN, E. A. ASEEVA, AND V. A. SPIVAK. "Plasma desorption mass spectrometry of haemoglobin tryptic peptides for the characterization of a hungarian a-chain variant." *Biol. Mass Spectrom.* **20**, 579-584 (1991).

3. O. N. JENSEN AND P. ROEPSTORFF. "Application of reversed phase high performance liquid chromatography and plasma desorption mass spectrometry for the characterization of a hemoglobin variant." *Hemoglobin* **15**, 497-507 (1991).

4. O. N. JENSEN, D. F. BAROFSKY, M. C. YOUNG, P. H. VON HIPPEL, S. SWENSON, AND S. E. SEIFRIED. "Direct observation of UV-crosslinked protein-nucleic acid complexes by MALDI mass spectrometry." *Rapid Commun. Mass Spectrom.* **7**, 496-501 (1993).

5. S. C. VENDELAND, M. A. BEILSTEIN, C. L. CHEN, O. N. JENSEN, E. BAROFSKY, AND P. D. WHANGER. "Purification and properties of selenoprotein W from rat muscle." *J. Biol. Chem.* **268**, 17103-17107 (1993).

6. O. N. JENSEN, D. F. BAROFSKY, M. C. YOUNG, P. H. VON HIPPEL, S. SWENSON, AND S. E. SEIFRIED. "Mass spectrometric protocol for the analysis of UV-crosslinked protein-nucleic acid complexes." in *Techniques in Protein Chemistry V* J. W. Crabb, Ed. (Academic Press, San Diego, 1994) pp. 27-37.

7. MEYER, M., JENSEN, O.N., BAROFSKY, E., BAROFSKY, D.F., AND REED, D. "Alkylation of thioredoxin by a dihaloethane-glutathione conjugate". *Chem. Res. Toxicol.* (In press) (1994).

8. S. E. BENNETT, O. N. JENSEN, D. F. BAROFSKY, AND D. W. MOSBAUGH. "UV catalyzed crosslinking of Escherichia coli Uracil-DNA Glycosylase to DNA." *J. Biol. Chem.* **269**, 21870-21879 (1994).

9. O. N. JENSEN, S. KULKARNI, J. ALDRICH, S. E. BENNETT, D. W. MOSBAUGH, AND D. F. BAROFSKY. "Characterization of covalent peptide-oligonucleotide hybrids by mass spectrometry." *Manuscript in preparation* , (199x).

PRESENTATIONS

1. Jensen, O.N., Barofsky, E., and Barofsky, D.F. (1993). MALDI mass spectrometry: Proteins in the gas-phase. *Proceedings of the Graduate Congress 93*, Oregon State University, p. 80.

2. Meyer, M., Jensen, O.N., Arbogast, B., Barofsky, D.F., and Reed, D. (1993) Protein alkylation by dihaloethane-glutathione adducts. *Abstract Volume of the Joint Meeting of the American Society of Biochemistry and Molecular Biology and Division of Biological Chemistry of the American Chemical Society, The FASEB Journal* (April issue), San Diego.

3. Jensen, O.N., Young. M., von Hippel, P.H., Swenson. S., Seifried, S., and Barofsky, D.F. (1993). A strategy for mass spectrometric characterization of nucleotide-protein complexes after UV-induced cross-linking. *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*, San Francisco, p. 266.

4. Jensen, O.N., Barofsky, D.F., Young. M., von Hippel, P.H., Swenson. S., and Seifried, S. (1993). Mass spectrometric protocol for the analysis of UV-crosslinked protein-nucleic acid complexes. *Protein Sci. 2. Abstract Volume of the Seventh Symposium of The Protein Society*, p. 101.

5. Swenson, S., Jensen, O.N., Barofsky, D.F., and Seifried, S.E. (1994) Mapping a nucleic acid binding site: Photoaffinity-labeling, mass spectrometry, and transcription termination factor Rho. Presented at the *Biophysical Society Meeting*. March 6-10, New Orleans.

6. Jensen, O.N., Bennett, S., Mosbaugh, D.W., and Barofsky, D.F. (1994) Analysis of UV-crosslinked protein-nucleic acid complexes by MALDI mass spectrometry. Presented at *Desorption '94*, March 27-31, Bend.

7. Beilstein, M.A., Vendeland, S.C., Barofsky, E., Jensen, O.N., and Whanger, P.D. (1994) Evidence for binding of glutathione to Selenoprotein W from rat muscle. Presented at *FASEB Experimental Biology '94*. April 24-29, Anaheim.

8. Jensen, O.N., Bennett, S., Mosbaugh, and Barofsky, D.F. (1994). Mass spectrometric characterization of UV-crosslinked protein-nucleic acid complexes. Presented at the *42nd ASMS Conference on Mass Spectrometry and Allied Topics*. May 29-June 1, Chicago.

9. Jensen, O.N., Bennett, S., Mosbaugh, D.W., and Barofsky, D.F. (1994). Mapping of nucleic acid binding sites in proteins by photochemical crosslinking and mass spectrometry. Presented at the *Third International Symposium on Mass Spectrometry in the Health and Life Sciences*, September 13-18, San Francisco.