1996

Protein Characterization by Mass Spectrometry

Steven Louis Cohen

Protein Characterization

by

Mass Spectrometry

A thesis presented to the faculty of
The Rockefeller University
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

by

Steven Louis Cohen

1996

Proverb

The universe
is made up
of stories,
not of atoms.
(Muriel Rukeyser)

We are the stuff of stars.
(Carl Sagan)

A man said to the Universe:
"Sir, I exist!"
"However," replied the Universe,
"The Fact has not created in me
A sense of obligation."
(Stephen Crane).

I would be the rock
about which the water is
flowing; and I would
be the water flowing
about the rock.
And I am both and neither
being flesh.
(Charles Reznikoff)

A man smiles by himself in the dark
perhaps because he can see in the dark
perhaps because he can see the dark.
(Yannis Ritsos)

"One of the greatest things in the world
is to train ourselves to see beauty
in the commonplace."
*(Charles Webster Hawthorne,
American realist painter)*

Acknowledgments

It's been said that 'it takes a village to raise a child.' It has taken a 'village' of colleagues to assist in my research and the completion of this thesis. Foremost, I am indebted to my advisor, Professor Brian Chait. Brian has provided much inspiration as well as a spirited intellectual atmosphere to pursue innovative mass spectrometry. I thank the other members of my thesis committee, particularly Professor Stephen Burley who encouraged many fruitful interactions with his laboratory; Professor William Agosta, committee chairman, for his time and critical comments on the thesis; and my external examiner, Professor Edward Ziff, from New York University, for numerous discussions and his critical comments.

I extend gratitude towards Professor Frank Field for strongly encouraging me to pursue the doctorate. Dr. Field established the Laboratory for Mass Spectrometry and Gaseous Ion Chemistry at Rockefeller over two decades ago.

I acknowledge the interactions with many laboratories, principally with Professor Burley and also Professors Jeffrey Friedman, James Darnell and Robert Roeder. I also extend gratitude to Professor John Kuriyan for his many comments and opinions concerning mass spectrometry.

I thank the many teachers I had along the way, especially Professor Peter Model, whose inquisitiveness and insightfulness were refreshing and always highly instructive.

I gratefully acknowledge the following individuals from Rockefeller for their scientific contributions and camaraderie that made the thesis possible and enjoyable: Adrian Ferré-D'Amaré (Max), Xiaoling Xie (TAFs), Jeffrey Halaas (leptin), Ketan Gajiwala (leptin), Uwe Vinkemeier (STAT), Dimitar Nikolov (TBP/TFIIB), Hal Lewis (Nova 1), Hristo Houbaviy (YY1), Lei Zhong (MEF 2C), Joe Marcotrigiano (eIF-4E), and Joe Kim (TBP). I also acknowledge Dr. Frank Collins at Amgen (leptin).

I am especially grateful for the following individuals for their assistance as well as friendship: Dr. Ron Beavis (NYU), for the MALDI data collection and viewing software; Herb Cohen for his invaluable electronics expertise; David Fenyö, for his protein digest mapping programs; Urooj Mirza, for his expertise and assistance with the electrospray ionization mass spectrometer; Salvatore Sechi for many valuable discussions on proteins and electrophoresis and teaching me SDS-PAGE; Jun Qin and Yingming Zhao, fellow graduate classmates, for their many discussions, Gladys McMilleon, for her secretarial expertise and NYC spirit, and Marta Delgado and the Dean's office for their assistance.

Many friendships were made along the way at Rockefeller that added much meaning to my work, life and the thesis. In addition to those already mentioned, I acknowledge the following: Subha Aduru, Jocelyn Chait, Tanuja Chaudhary, James Chou, Swapan Chowdhury, Ivan Haller, Vish Katta, Harry Lander, Jon Lindsey, Louis Mazzarelli, Urooj Mirza, Tom Muir, Shufang Niu, Julio

Cesar Padovan, Klaus Schneider, Ruud Steenvoorden, Susanna Thornqvist, Rong Wang, Wenzhu Zhang.

I finally thank New York City for its greatness and infinite diversions, and in particular, the Socrates Sculpture Park in Long Island City —a place for some *camera obscura* and quiet meditation.

Table of Contents

Abstract **1**

1. Introduction **3**

1.1 Protein Analysis 4
1.2 Protein Characterization by Traditional Methods 4
1.3 Protein Characterization by Mass Spectrometry (MS) 6
1.3.1 Mass Analyzer 6
1.3.2 Electron Impact Ionization 9
1.3.3 Chemical Ionization 11
1.3.4 Early Desorption Ionization Methods 11
1.3.5 Fast-Atom Bombardment Ionization 13
1.3.6 Matrix-Assisted Laser Desorption/Ionization (MALDI) 15
1.3.7 Electrospray Ionization (ESI) 19
1.3.7.1 Non-covalent Complexes by ESI-MS 21
1.4 Structural Characterization of Proteins 23
1.5 Proteolysis 24

2. Experimental **29**

2.1 Materials 29
2.1.1 Proteins, Peptides and Oligonucleotides 29
2.1.2 Proteases 30
2.1.3 MALDI Matrix and Matrix Solutions 30
2.1.4 Other Materials and Equipment 32
2.2 **Methods** 33
2.2.1 Proteolysis 33
2.2.2 Matrix Solution Preparation 33
2.2.3 Matrix-Analyte Solution Preparation 34
2.2.3.1 Dried-drop Crystallization 36
2.2.3.2 Slow Crystallization 36
2.2.3.3 Rapid Crystallization 37
2.2.4 Mass Spectrometry 38
2.2.4.1 MALDI-MS Instrumentation 38
2.2.4.2 The MALDI Probe 39
2.2.4.3 General Instrument Conditions 41
2.2.4.4 Data Collection 41
2.2.4.5 Mass Calibration 42
2.3 Study I Influence of Matrix Solution Conditions on the MALDI-MS Analysis of Peptides and Proteins 43
2.3.1 Background 43

2.3.2 Results 45
2.3.2.1 Effects of the Matrix Solution Composition 45
2.3.2.2 Effects of the Detergent n-Octylglucoside 49
2.3.2.3 Effects of Droplet Evaporation 51
2.3.2.4 Effects of the Rate of Matrix Crystal Growth 54
2.3.2.5 Mass Discrimination Effects between Proteins 57
2.3.3 Discussion 60
2.3.3.1 Slow Matrix Crystallization 60
2.3.3.2 Rapid Crystallization 61
2.3.3.3 Dried-drop Crystallization 62
2.4 Study II Special Properties of the 4HCAA/FWI Matrix Solution 64
2.4.1 Effects of Additives on the MALDI-MS of Proteins 64
2.4.2 Effects of DNA on the MALDI-MS Analysis of Proteins 66
2.4.3 Effects of Redissolving the Matrix-Analyte Crystal 69
2.5 Conclusion 71

3. Mass Spectrometry of Whole Proteins Eluted from SDS-PAGE Gels 73
3.1 Introduction 73
3.2 Materials 75
3.2.1 Proteins 75
3.3 Methods 76
3.3.1 Preparation of Proteins for SDS-PAGE 76
3.3.2 SDS-PAGE 77
3.3.3 Gel Staining 77
3.3.4 Gel Excising and Destaining 79
3.3.5 Extraction Method A 81
3.3.6 Extraction Method B 83
3.3.7 In-Gel Digests 85
3.3.8 MALDI-MS 86
3.4 Results 87
3.4.1 Eluting Whole Proteins from SDS-PAGE Gels 87
3.4.2 Extraction Method A: Elution of Whole Proteins from Gels for MS Analysis. 87
3.4.3 Slow Crystallization of the MALDI Matrix Greatly Enhances MALDI-MS Sensitivity 92
3.4.4 Extraction Method B: Eluting Whole Proteins from Gels for High Sensitivity MALDI-MS Analysis 94
3.4.5 Mass Accuracy 97

3.5 Discussion **100**
3.6 MALDI-MS of Whole Proteins from Gels: An Application **105**
3.6.1 The Characterization of Native Human Leptin 105
3.7 Conclusions **114**

4. Proteolysis as a Probe for Protein Structure: Max and Related Proteins **115**
4.1 Materials **115**
4.1.1 Proteins 115
4.1.2 DNA Oligonucleotides 117
4.1.3 Other Materials 118
4.1.4 Proteases —Preparation, Storage and Specificities 119
4.2 Methods **120**
4.2.1 Max∆C Phosphorylation 120
4.2.2 Proteolytic Digestions 121
4.3 Introduction —Max **122**
4.4 Results **131**
4.4.1 V8 Proteolysis of b/HLH/Z Max in the Absence of DNA 131
4.4.2 V8 Proteolysis of b/HLH/Z Max in the Presence of Max-specific DNA 133
4.4.3 V8 Proteolysis of b/HLH/Z Max in the Presence of non-specific DNA 136
4.4.4 V8 Proteolysis and the Effects of Ionic Strength 138
4.4.5 Proteolysis of the Leucine Zipper 139
4.4.6 Additional Proteases 142
4.4.7 The Proteolytic Protection Map 144
4.5 Discussion **146**
4.5.1 Max in the Absence of DNA 146
4.5.2 Max in the Presence of DNA 148
4.5.2.1 The N-Terminal Region 150
4.5.2.2 The Central Region 150
4.5.2.3 The C-terminal Region 151
4.5.3 Evaluating the Proteolytic Protection Assay 153
4.5.3.1 Specific vs. Non-Specific Protein-DNA Interactions 153
4.5.3.2 Dynamics of the Max-DNA Complex 154
4.5.3.3 MALDI-MS is Well Suited for the Proteolytic Protection Assay 155
4.6 Conclusions of the Max Study **157**
4.7 Proteolysis to Structurally Probe Other HLH Proteins **160**
4.7.1 SREBP, USF, Hairy 160
4.7.1.1 SREBP 161
4.7.1.2 USF 162

4.7.1.3 Hairy 162
4.8 Proteolysis to Structurally Probe Phosphorylated Max 164
4.8.1 Results 168
4.8.1.1 Proteolysis of MaxΔC and MaxΔC(+)P 169
4.8.1.2 Lys-C Digestion of MaxΔC 172
4.8.1.3 Glu-C Digestion of MaxΔC 172
4.8.1.4 Asp-N Digestion of MaxΔC 173
4.8.1.5 Subtilisin Digestion of MaxΔC 174
4.8.2 Discussion 175
4.8.2.1 Differences between MaxΔC and b/HLH/Z Max 175
4.8.2.2 The N-terminal region 176
4.8.2.3 Effects of Substrate Phosphorylation on Protease Activity 176
4.8.2.4 Lifetime of the MaxΔC-DNA Complex 177
4.8.3 Conclusions 180

5. Domain Elucidation by Mass Spectrometry 182
5.1 Materials 182
5.1.1 Proteins 182
5.1.2 Other Materials 183
5.2 Methods 187
5.2.1 Proteolysis and Cyanogen Bromide Digest 187
5.2.2 Electrospray-Ionization Mass Spectrometry 187
5.2.2.1 Denaturing Conditions/Direct Infusion 187
5.2.2.2 Denaturing Conditions/On-Line1 Desalting Cartridge 188
5.2.2.3 Non-Denaturing ESI-MS 189
5.2.3 Centricon Desalting 190
5.3 Introduction 191
5.4 Results: Domain Elucidation of Unknown Structures 196
5.4.1 Mef-2C 196
5.4.2 Leptin 200
5.4.3 TAF42/62 205
5.4.3.1 ESI-MS of Non-Covalent Interactions Between TAF42 and TAF62 208
5.4.3.2 Domain Elucidation of the TAF42/62 Complex for High-Resolution Studies 212
5.4.4 Domain Elucidation by Mass Spectrometry Additional Applications 217
5.5 Conclusions —Mass Spectrometry as a Tool for Protein X-ray Crystallography. 217
Bibliography..220

List of Figures

Figure 1.1 The elements of mass spectrometry. 7

Figure 1.2 The basic components of MALDI-TOF-MS. 16

Figure 1.3 The basic components of ESI-MS. 20

Figure 1.4 Protease-extended substrate interactions. 26

Figure 2.1 MALDI-MS matrix crystallization methods. 35

Figure 2.2 MALDI-MS of peptide mixture using different matrix solutions. 46

Figure 2.3 Effects of detergent on the MALDI-MS of a protein digest. 50

Figure 2.4 pH profile of an evaporating drop of matrix solution. 53

Figure 2.5 Effects of the rate of matrix crystal growth on MALDI-MS of a protein digest. 56

Figure 2.6 Mass discrimination effects between proteins in MALDI-MS. 59

Figure 2.7 Effects of polyethylene glycol on MALDI-MS of proteins. 65

Figure 2.8 Protein-DNA adduction by MALDI-MS. 68

Figure 2.9 Effects of redissolving the matrix crystals of a protein digest. 70

Figure 3.1 Eluting whole proteins from SDS-PAGE gels for mass spectrometric analysis. 78

Figure 3.2 MALDI-MS of proteins eluted from SDS-PAGE gels using gel Extraction Method A. 88

Figure 3.3 MALDI-MS of high mass proteins extracted from gels by Extraction Method A. 90

Figure 3.4 ESI-MS of recombinant human leptin extracted from SDS-PAGE gel by Extraction Method A. 91

Figure 3.5 Slow Crystallization of matrix greatly improves the MALDI-MS response of proteins. 93

Figure 3.6 MALDI-MS of recombinant human leptin from cooper-stained SDS-PAGE gels. 95

Figure 3.7 MALDI-MS of molecular mass markers extracted from cooper-stained SDS-PAGE gels. 96

Figure 3.8 Internal calibration improves mass accuracy. 99

Figure 3.9 MALDI-MS of immunoaffinity isolated human leptin from blood serum. 109

Figure 3.10 MALDI-MS of human leptins. 112

Figure 3.11 Trypsin digestion of human leptins. 113

Figure 4.1 HLH constructs. 116

Figure 4.2 Sequence alignment of HLH proteins. 124

Figure 4.3 The b/HLH/Z Max-DNA complex. 127

Figure 4.4 Proteolytic protection assay. 130

Figure 4.5 V8 digest of b/HLH/Z Max in the absence of DNA. 132

Figure 4.6 V8 digest of b/HLH/Z Max in the presence of E-box (+)DNA. 134

Figure 4.7 V8 digest (48 hr) of b/HLH/Z Max + E-box DNA. 135

Figure 4.8 V8 digest of b/HLH/Z Max (+) nsDNA or elevated salt. 137

Figure 4.9 Subtilisin digest (1 min) of b/HLH/Z Max (-) and (+) DNA. 140

Figure 4.10 Subtilisin digest of b/HLH/Z Max (+) E-box DNA. 141

Figure 4.11 Summary of the proteolysis of b/HLH/Z Max-DNA complex. 143

Figure 4.12 Proteolytic protection map of the b/HLH/Z Max-DNA complex 145

Figure 4.13 Calculation of percent Max dimer bound to DNA. 156

Figure 4.14 Casein kinase II consensus sites in HLH proteins. 165

Figure 4.15 MaxΔC + nsDNA MALDI-MS proteolysis summary. 170

Figure 4.16 MaxΔC + E-Box DNA MALDI-MS proteolysis summary 171

Figure 5.1 Protein constructs for domain elucidation studies. 186
Figure 5.2 Domain elucidation. 192
Figure 5.3 Domain elucidation of MaxΔC. 194
Figure 5.4 Mef2C domain elucidation. 197
Figure 5.5 Long-term stability of recombinant mouse leptin in solution. 202
Figure 5.6 Characterization of recombinant human leptin. 204
Figure 5.7 Sequence alignments of dTAF42 & dTAF62 with the histones H3 & H4. 206
Figure 5.8 ESI-MS of dTAF42/62 under denaturing conditions. 209
Figure 5.9 ESI-MS of dTAF42/62 under native conditions. 211
Figure 5.10 Domain elucidation of the dTAF42/62 complex. 215
Figure 5.11 The dTAF42/62 and histone H3/H4 structures. 216
Figure 5.12 Protein X-ray crystallography and mass spectrometry. 218

List of Tables

Table 2.1 Effects of matrix solution composition and pH on the MALDI-MS of a complex mixture of peptides. 48

Abstract

Powerful new techniques in mass spectrometry (MS) provide an unprecedented opportunity to develop methods that facilitate protein characterization. This thesis utilizes matrix-assisted laser desorption/ionization and electrospray ionization MS to explore protein structure. Starting at the 'primary' level of protein structure, an approach is developed to aid in the characterization of posttranslational modifications and processing of mature gene products. The approach involves the elution of picomole amounts of whole proteins from SDS-PAGE gels in a form suitable for MS analysis. Application of the elution method to the obesity factor leptin led to the characterization of the endogenous form of leptin in obese humans.

A biochemical procedure is developed that permits 'higher' order structural features of proteins to be probed. The method uses limited proteolysis to distinguish regions within a target protein that are solvent exposed and flexible from those that are solvent inaccessible or conformationally rigid. The DNA-binding protein Max was investigated by the procedure and several structural features were determined. The structural study of Max was extended in an investigation of the effects of N-terminal phosphorylation. Limited proteolysis was found to be of great utility in the determination of the 'precise' boundaries of protein folding domains. This method of 'domain elucidation' is explored fully and applied to proteins of unknown structure, including the *Drosophila* TAF

42/62 protein complex. Limited proteolysis/MS experiments permitted a rapid design of a compact construct of the TAF complex, crystals of which X-ray diffracted to 1.4 Å resolution.

MALDI-mass spectrometric methods were absolutely crucial to every step of the described research. However, before the methods could be effectively applied, it proved necessary to develop new and practical MS sampling handling procedures.

Chapter 1 — Introduction

What is 'protein characterization?' I asked this question before choosing these words for the thesis tittle. Is characterization the same as analysis (an alternate title)? Not really. The terms 'characterize' and 'analyze' have distinctly different definitions. *Analyze*, according to Webster's Third Edition, means to "ascertain the components of or separate into component parts" whereas *characterize* means "to describe the essential character or quality of." The former is mechanical, the latter is intimate. Accordingly, protein analysis refers simply to protein separation whereas protein characterization refers to the measuring of more intrinsic properties of proteins, such as molecular mass, amino acid modifications and structure.

I begin the thesis with a discussion of methodologies that are the traditional 'tools of the trade' in protein analysis and characterization. Therein, I introduce the 'new player', mass spectrometry (MS) —a highly developed 'mechanical' methodology that can render 'intimate' details about a molecule. For proteins, new mass spectrometric techniques can measure accurate molecular mass, identify modifications of peptide residues, and provide structural information. Finally, I present the concept of structural characterization via proteolysis, a traditional biochemical method whose utility, when combined with mass spectrometry, is greatly enhanced.

1.1 Protein Analysis

For biologists, the tool of choice to analyze proteins has been polyacrylamide gel electrophoresis (PAGE). The most common forms of PAGE are SDS (sodium dodecyl sulfate)-PAGE, isoelectric focusing (IEF), and nondenaturing PAGE. In SDS-PAGE, proteins are separated spatially in one-dimension, essentially according to their molecular masses. In IEF proteins are spatially separated in one-dimension through a pH gradient according to their isoelectric points or net charge. In two-dimensional PAGE, IEF is run in the first dimension, followed by SDS-PAGE in the second dimension. The resolving power of 2-D gels is unsurpassed, particularly in separating complex mixtures of proteins from whole-cell lysates. In nondenaturing gels, electrophoresis is performed under native conditions, important when the enzymatic activity or structure of a protein or protein complex needs to be maintained during separation.

1.2 Protein Characterization by Traditional Methods

To characterize proteins, PAGE methods are insufficient by themselves. Molecular mass estimations by SDS-PAGE are often no better than 5-10% accurate and many times grossly inaccurate. The isoelectric point, an intrinsic property of the protein, is by itself of little value. However, if an antibody against a target protein is available, Western electroblotting methods are useful to recognize the presence of the protein on the gel.

Classical methods for characterizing proteins include amino acid analysis and N-terminal sequencing. In amino acid analysis, a single, purified protein is subjected to strong acid at elevated temperatures to hydrolyze all peptide bonds and release the individual amino acid residues. Elution through an ion-exchanger identifies and quantifies the different amino acid residues. However, amino acid analysis provides no information about protein sequence, is prone to errors and requires a considerable amount of protein (>10 μg).

Protein sequencing has been made possible by the powerful, classical method of N-terminal (Edman) sequencing. Edman sequencing requires the prior steps of protein separation, usually one- or two-dimensional gels, and electroblotting. However, a substantial portion of intact intracellular eukaryotic proteins are 'blocked' at their N-terminus, usually with acetyl groups. N-terminally block proteins cannot be directly sequenced. An alternate strategy involves enzymatic digestion of the protein on the blotting membrane or in the gel, elution of the resulting peptide digest fragments for reverse-phase HPLC separation. The HPLC-separated fragments can be delivered to an automated sequencer for identification using Edman chemistry. Despite automation, Edman sequencing methods are time consuming and expensive. More significantly, success rates fall drastically when analyzing low picomole amounts of protein. In addition, although Edman sequencing can locate a covalently modified residue, the Edman method cannot normally determine the exact identity of the modification. Nonetheless, in the 'pre-genome era', Edman sequencing has

been the method of choice for providing sequence information enabling the design of cDNA probes for gene cloning.

1.3 Protein Characterization by Mass spectrometry

From a biologist's perspective, mass spectrometry is a methodology that seems to have appeared 'out of the blue.' However there is a long history that speaks of the efforts and advancements to bridge the biological community with mass spectrometry. I survey some of that history in the following section, providing some perspective and appreciation to the 'pioneers' of the field.

Mass Spectrometry

Simply put, a mass spectrometer measures the molecular mass of a molecule, accurately and better than any other methodology. All mass spectrometers consist of three principal components: an ion source, mass analyzer and detector (Figure 1.1). Development of the mass analyzer and ion source have had the most impact on mass spectrometry, particularly in biological applications.

1.3.1 Mass Analyzer

Early in the development of mass spectrometry, attention was focused principally on the mass analyzer. In the early 1900s electric and magnetic fields were used to separate positively charged 'rays' onto a photographic plate. The rays were generated by an electrical discharge. Aston developed the first mass

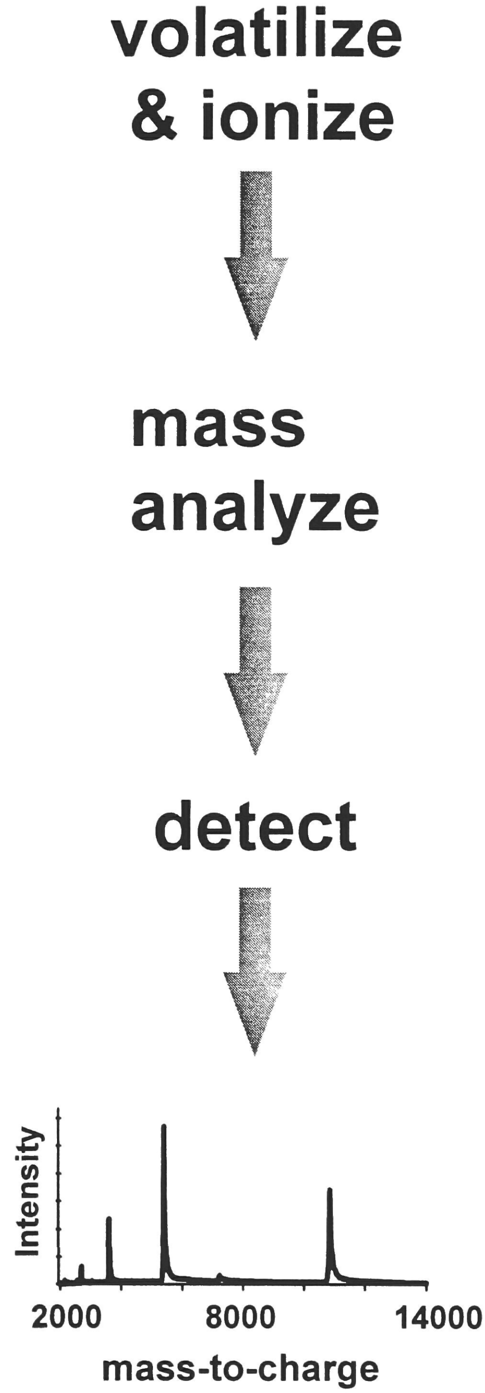


Figure 1.1 *The elements of mass spectrometry*

The essential elements of mass spectrometry are sample volatilization and ionization, followed by mass analysis and ion detection. The resulting mass spectrum is a plot of relative ion intensity vs. the mass-to-charge ratio.

spectrograph in 1919 (he coined the term). A photoplate was used to collect the separated rays that were later identified to be charged atoms (i.e., ions). The mass spectrograph represented a substantial breakthrough since it provided accurate mass-to-charge (m/z) ratios of ions and a rapid means to determine isotopic abundances of many of the elements.

In 1922 Dempster introduced the first 'single focusing' magnetic mass spectrometer, equipped with an electrometer detector. It soon became apparent that improvements in mass resolution were necessary, especially to satisfy the growing needs of the rapidly developing field of synthetic organic chemistry. Fulfilling the need was the double-focusing mass spectrometer, advanced during the 1930s and 40s. Double-focusing instruments were equipped with electric as well as magnetic sectors grouped in various arrangements and coupled together with mechanical slits. The double-focusing instrument provided angular as well as energy focusing of an ion beam, the combination of which produced resolutions exceeding 10,000. Double-focusing instruments are still used today, and though expensive, routinely provide high mass accuracy and resolution measurements of small to intermediate organic molecules (M_r <2000).

Another important group of mass analyzers, developed by Wolfgang Paul in the 1950s, was the relatively inexpensive quadrupole ion filter and the quadrupole ion trap. A much more expensive and high-performance analyzer is the Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR/MS or FT-MS), introduced by Comisarow and Marshall in the 1970s. FT-MS offers very

high resolution (exceeding 10^6 at m/z 100) and high sensitivity. Finally, the least expensive analyzer (and conceptually the easiest to understand) is the time-of-flight (TOF) mass spectrometer, developed into a practical device by Wiley and McLaren during the late 1950s. A majority of the research presented in this thesis was performed on a TOF mass spectrometer. The TOF mass spectrometer consists of a long 'field-free' flight path (typically 0.5-2 m) through which ions traverse *in vacuo* to the detector. Using fast electronics, transit times of the ions through the field-free drift region are accurately measured of all ions impinging on the detector. The result is a 'time-of-flight' (TOF) spectrum. The m/z of an ion is directly proportional to the square of its flight time, thus the TOF data is easily converted into mass-to-charge data. The TOF mass spectrometer offers low to moderate resolution, high transmission and, in principal, an unlimited m/z range.

1.3.2 Electron Impact Ionization

Without a volatilization/ionization source, the mass analyzer is just a heavy and expensive paperweight. On early commercial MS instruments, ionization was commonly performed by bombarding a gaseous sample with ~70 eV electrons, a process that results in removal of an electron from the sample atom or molecule (i.e., electron ionization, EI). Involatile samples must be vaporized prior to ionization, so that EI is limited to molecules of low polarity and molecular weights. For biological samples, small and relatively non-polar molecules (e.g., fatty acids and steroids) can be analyzed by EI. Chemical

derivatization methods have been extensively developed to increase the volatility of polar compounds. A significant advancement resulted from coupling of the eluant output of a gas chromatograph to the ion source of the mass spectrometer (i.e., GC/MS). The practice of GC/MS has evolved into a large field whose activities extend to forensic, environmental, and chemical sciences. The Olympic Games is a good example of where GC/MS is applied to test participating athletes for illegal use of anabolic steroids and other drugs (Longman, 1996).

Despite the many early efforts to facilitate volatilization, EI is a high energy process, producing radical ion species that tend to undergo significant gas-phase fragmentation. Thus, the signal from an intact 'parent' ion is normally weak or not detected when using EI, particularly for compounds with $M_r > 400$ Da. However, most molecule ions exhibit a unique pattern of fragmentation, driven mainly by gas-phase radical ion chemistry. Elaborate rules governing the gas-phase fragmentation of small organic ions, pioneered by McLafferty and others, have been enumerated and are still being actively studied (McLafferty and Turecek, 1993). The fragmentation rules have permitted structural elucidation by MS of an enormous number of organic molecules, especially natural products. Today, the EI fragmentation pattern of an 'unknown' compound can be submitted to search large databases of EI mass spectra using pattern matching retrieval algorithms to facilitate identification or classification of the compound.

1.3.3 Chemical Ionization

Chemical ionization (CI), developed by Field and Munson in 1966 (Munson and Field, 1966) represented a major breakthrough in ionization methods. CI is a lower energy ('softer') process than EI —much less fragmentation is observed. Thus CI provides a facile means to measure the masses of intact, moderately volatile molecules. Usually a protonated or deprotonated species is detected in CI. For electronegative compounds, electron capture CI has proven to be a particularly sensitive method of MS analysis. CI has also bolstered the use of high resolution MS instruments for determining elemental composition of organic molecules by measuring masses to an accuracy better than 0.01 mass units. CI was soon thereafter adopted by instrument manufacturers and became, along with EI, a standard mode of ionization. The combination of gas-phase fragmentation patterns (from EI-MS) along with an accurate molecular mass (from CI-MS) has become a powerful means of analyzing organic molecules.

1.3.4 Early Desorption Ionization Methods

To begin to make a significant impact on the analysis of intractable biological molecules, a breakthrough in volatilization procedures was needed. It came with the development of desorption methods —field desorption , pioneered by Beckey and Schulten (Beckey and Schulten, 1975) and plasma desorption, discovered by Macfarlane and Torgerson (Torgerson et al., 1974).

In field desorption (FDMS), ionization of the sample *precedes* volatilization. FDMS was a softer ionization method than EI and showed promise for qualitative analysis of picogram amounts of drugs and drug metabolites, small sugars and other very polar molecules that failed to yield a parent ion by EI or CI. However, FDMS never really developed 'legs' because of sample preparation and operational difficulties associated with the method. FDMS soon became over shadowed by more facile and powerful desorption methods, one of which was plasma desorption (Torgerson et al., 1974).

In plasma desorption (PDMS) the sample is directly deposited onto a thin metal surface. The reverse side of the surface is bombarded by MeV fission fragments that originate from spontaneous decay of ^{252}Cf nuclei mounted close to the surface. The bombardment causes sample molecule ionization and volatilization. Mass analysis is by TOF-MS since ion production is pulsed, based on the random decay of ^{252}Cf nuclei. Intact parent and fragment ions are observed in PDMS. Despite technical complexities (e.g., data collection sometimes can take hours for each sample), plasma desorption was adopted by several groups around the world (including Field and Chait at Rockefeller University). PDMS was much softer than EI-MS and has proven to be a rather powerful tool in analyzing peptides (Chait, 1989) and other intractable molecules such as synthetic porphyrins (Lindsey et al., 1992). Small molecules with masses up to 24 kDa could be analyzed with this method, albeit with some difficulty. Commercialization of PDMS lagged other methods, however, so PDMS was never universally adopted by the greater mass spectrometry

community. Dyed-in-the-wool sector-type mass spectrometrists were reluctant to embrace the non-sophisticated pulsed ion source of PDMS and the low resolution time-of-flight mass analyzer.

1.3.5 Fast-Atom Bombardment Ionization

Desorption technology gradually led to the developed of new atom and ion bombardment methods. In the late 1970s Benninghoven and others introduced secondary ion mass spectrometry (SIMS), a method that provided a means to study the molecular composition of solid surfaces. A major breakthrough came with Fast Atom Bombardment (FAB) pioneered by Barber (Barber et al., 1981). FAB is a liquid SIMS method in which the analyte is dissolved in a low-volatility matrix such as glycerol. The analyte/matrix mixture is bombarded *in vacuo* by keV energy heavy atoms or ions (e.g., Xe or Cs^+). Desorption and ionization of the analyte molecules from the liquid surface produces a continuous, stable beam of ions that is suitable for analysis by scanning sector instruments. FAB provided a means to analyze otherwise intractable small and medium-sized polar molecules such as nucleotides, sugars and peptides with greater reproducibility and ease than field desorption techniques. FAB is a moderately low energy process producing parent ions (protonated or deprotonated) of peptides with masses usually up to 6 kDa and in some reports up to 25 kDa. The method was universally embraced by the MS community and was quickly commercialized.

Following historical trends, the notion of structural determination using FAB-MS took hold. In particular, peptide (and oligosaccharide) sequencing and the characterization of posttranslational modifications were actively explored. PDMS of peptides was already shown by several groups to provide sequence information due to a significant degree of fragmentation associated with the method. Peptide sequencing by FAB-MS required the use of tandem MS. Tandem MS, first conceptualized by McLafferty in the 1960s and then developed by McLafferty, Biemann and others, offered a means to selectively analyze components of a mixture using two mass spectrometers joined in series. In the first mass spectrometer, an ion is mass selected, energetically 'activated' by collisions with an inert gas and subsequently allowed to undergo unimolecular dissociation, the product ions of which are analyzed by the second mass spectrometer. Fortuitously, gas-phase peptide ions fragment primarily along the peptide backbone, providing a means to sequence the peptide. Through extensive research, the gas-phase fragmentation of peptide ions was found to be a complex function of the mode of ionization, collision energy, and the peptide sequence. Groups headed by Biemann, Cooks, Hunt, Yost & Enke, and others developed tandem MS instrumentation for sequencing peptide ions generated by FAB-MS. Despite its widespread appeal, however, FAB-MS offers only moderate sensitivity (barely subnanomole) and the sensitivity sharply falls off for large (> 2 kDa) peptides. In addition, the mass analysis of large peptides requires expensive (> $ 200,000) and intricate magnetic sector instruments. In order to compete with existing N-terminal Edman sequencing methods, the MS

sensitivities provided by existing ionization methods required about a thousand-fold improvement. This came with two concurrent ionization breakthroughs in mass spectrometry —matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). MALDI and ESI revolutionized biological MS, catapulting mass range accessibility, sample sensitivity, mass accuracy, and turn-around times to unprecedented levels. They are the main players in biological mass spectrometry today.

1.3.6 Matrix-Assisted Laser Desorption/Ionization (MALDI)

MALDI-MS was introduced by Hillenkamp and Karas in 1988 (Karas and Hillenkamp, 1988). Figure 1.2 shows a simplified schematic of the MALDI-TOF instrument designed and constructed by Chait and Beavis at Rockefeller University (Beavis and Chait, 1989). The 'magic' in MALDI occurs between the laser and the sample. A laser beam directed at 'bare' protein deposited on the probe surface would result in charring of the sample and no signal from intact protein. Hillenkamp and Karas originated the idea to embed the protein in a solid organic matrix. (A second group developed an alternative laser desorption method that made use of a viscous liquid matrix, but due to poor sensitivities, was never adopted (Tanaka et al., 1988).) For MALDI-MS, the solid matrix is a small, inert organic molecule, the crystals of which are capable of absorbing laser light followed by sublimation. The crystals are grown from solutions

MALDI-MS

matrix-assisted laser desorption (time-of-flight) mass spectrometry

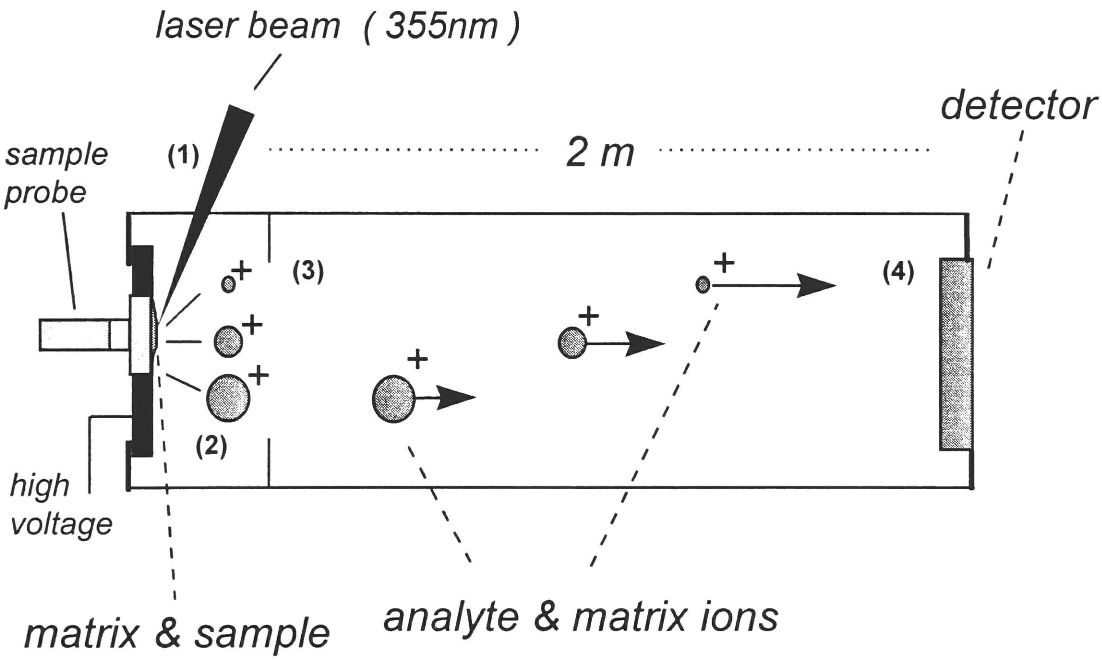


Figure 1.2 *The basic components of MALDI-TOF-MS.*

The essential elements of MALDI-TOF-MS consisting of a sample probe, laser, time-of-flight analyzer, and detector (not to scale).

(1) A pulsed laser beam is focussed on the matrix-sample co-crystals (deposited on a probe that is placed *in vacuo* at high voltage). **(2)** Molecules are desorbed by the laser into the gas phase. Ionization occurs producing ions (circlular shapes) of differing mass-to-charge (m/z) ratios (denoted by the circle size). **(3)** With each laser pulse and starting at the same time, ions accelerate out of the source into the 2 m flight tube. Transit time down the flight tube is proportional to the square-root of m/z (lower m/z ions have higher velocities, denoted by arrow lengths). **(4)** Ions strike the detector, producing an electrical signal that is amplified and sent to the computer where the precise flight time of all detected ions are stored as a time-of-flight spectrum. Mass conversion of the data results in a mass spectrum.

consisting of analyte (e.g., proteins or peptides) and the matrix dissolved in water (usually acidified) along with a portion of miscible organic solvent. The matrix must exhibit at least partial solubility in the solution. It is believed that co-crystallization of the analyte with the matrix is critical for the MALDI process. In a simple but elegant experiment, Beavis showed that protein incorporates into sinapinic acid matrix crystals in a well defined manner (Beavis and Bridson, 1993).

The precise mechanism of the MALDI process remains a mystery. In short, the matrix acts to ionize as well as transport analyte molecules into the gas phase and to provide a 'cushion' against charring by the high energy laser. In one possible scenario, the laser photon (normally a 5-10 nsec pulse) is 'absorbed' by the matrix, the result of which leads to sublimation (or ablation) of matrix molecules forming a dense matrix plume. In this ablation event, analyte molecules embedded in the matrix are entrained in the plume. Experiments done in the Chait lab as well by others indicate that the matrix plume is highly 'forward' directing, suggesting that the plume may consist of a fast moving supersonic jet of molecules. A large number of collisions within the jet would lower the internal energies of the molecules in the plume and foster analyte ionization, two events that ultimately permit the detection of intact analyte ions.

Hillenkamp and Karas originally used 266 nm UV light from an expensive Neodymium/YAG laser and the matrix nicotinic acid. The MALDI mass spectra of proteins run in nicotinic acid matrix exhibited poor resolution and extensive matrix photochemically generated adduct peaks. Chait and Beavis discovered a

group of cinnamic acid derivatives (such as sinapinic acid) that permitted the use of longer wavelength photons (355 or 337 nm) generated from a Nd/YAG laser or from a less expensive nitrogen laser. The cinnamic acid matrices yielded improved resolution and much less matrix adduction (Beavis and Chait, 1990a). Application of cinnamic acids matrices also revealed a great strength of MALDI —the ability to analyze complex mixtures of polypeptides (Beavis and Chait, 1990b). The power of MALDI-MS towards biological applications became immediately apparent with its high mass accuracy (0.01% for peptides), sensitivity (subpicomole) and analysis speed (minutes) (Hillenkamp et al, 1991; Chait and Kent, 1992; Beavis and Chait, 1996).

Despite the rapid growth of the MALDI-MS technique, there has been a need to better understand and improve the sample preparation stage of the analysis. In this regard, I undertook an extensive study of the influences of various sample preparation parameters that significantly effected the MALDI-MS of peptides and proteins (see chapter 2). The results of the matrix study directly influenced all of my subsequent research, permitting the development of a method to elute proteins from gels for high sensitivity MALDI-MS analysis (chapter 3) and the optimization of protein digestion analysis for protein structural studies (chapter 4) and domain elucidation (chapter 5).

1.3.7 *Electrospray Ionization (ESI)*

Electrospray ionization (ESI) MS was the other 'revolution' in biological MS analysis, sweeping through the MS community about the same time as MALDI-MS. ESI-MS was innovated by Fenn and coworkers (Fenn et al., 1989). The great appeal in ESI-MS is the ability to generate gas-phase peptide and protein ions directly from solution, in a simple and highly reproducible way. ESI-MS was also ideal for interfacing to the liquid chromatograph. For years, a major challenge was the interfacing of liquid chromatography (LC) to the mass spectrometer. Through the early 1980s, two methods showed some promise for an LC/MS interface —continuous-flow FAB and thermospray. Both methods, however, were beset with technical difficulties and were quickly supplanted by the development of ESI-MS. With ESI, a practical LC/MS interface was finally possible.

Figure 1.3 depicts the general features of ESI-MS. The idea of ESI is rather simple. A solution containing analyte (such as protein) is infused through a metallic hypodermic needle that is kept at a high electrical potential (~4 kV). The solution emerges from the needle at ambient pressure as a mist of small, highly charged droplets that contain the analyte. The droplets are entrained by a gas flow into the vacuum of the mass spectrometer. On the original Rockefeller electrospray design (Chowdhury et al., 1990), the flow is directed through a heated 20 cm long, 0.5 mm i.d. stainless steel capillary tube. The heating assists in droplet evaporation and ion formation. The exact mechanism of ion formation

Electrospray Ionization

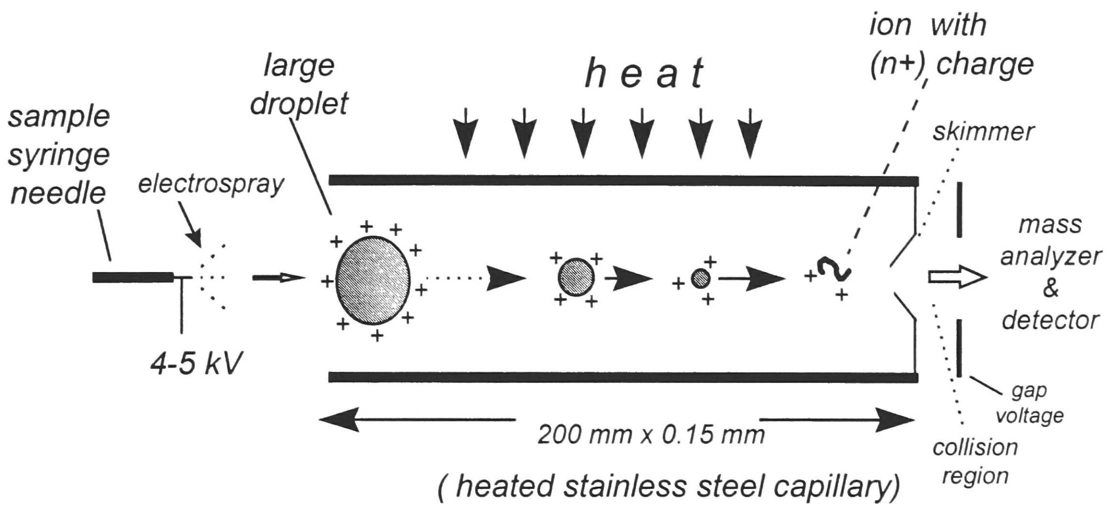


Figure 1.3 *Basic components of electrospray ionization (ESI).*

The essential elements of ESI-MS (not drawn to scale) consist of an analyte dissolved in an appropriate solution, the electrospray, a heated capillary tube, the skimmer, mass analyzer and detector.

A solution is infused through a fine needle that is placed at 4-5 kV. Under the proper flow conditions, the solution emerging from the needle forms an 'electrospray' at the tip of the needle under atmospheric pressure. Guided by an electric field, the spray is directed into a heated stainless steel capillary tube. The spray droplets are entrained into a gas flow through the capillary (due to the vacuum of the MS instrument). The heat of the capillary evaporates and shrinks the droplets, eventually leaving individual analyte ions of multiple charges and a few solvating water molecules. Passing through the skimmer, the hydrated ions collide with residual gas. The collision energy is controlled by the gap voltage. The collisions completely desolvate the ions that are then focused into the mass spectrometer for analysis and detection.

remains in dispute. One scenario envisions ion formation through a series of 'Coulombic explosions' occurring from the drop. Through evaporation, the highly charged droplet shrinks to a critical size —the so called Raleigh limit— where electrostatic repulsive forces exceed the surface forces keeping the drop together. At this point, an asymmetric Coulombic explosion occurs from the surface of the drop, releasing smaller droplets containing the analyte. Following several cycles of evaporation and Coulombic explosions, the droplet will have almost completely evaporated, leaving the analyte in a highly charged (i.e., protonated or cationized) state and surrounded by a small number of solvent molecules. Exiting the heated capillary tube, the solvated analyte ion collides with residual gas in the vacuum (now at ~1 torr) in transit to the skimmer. The skimmer is at the entrance of the mass analyzer. An adjustable electric field, maintained between the exit of the capillary and the skimmer, controls the collision energy. At low to moderate collision energies the analyte ion can be completely desolvated (Chowdhury et al., 1990), whereas at higher energies the analyte, if it is a peptide ion, is collisionally activated and made to undergo fragmentation (Katta et al., 1991).

1.3.7.1 Observing Non-Covalent Interactions by ESI-MS

Normally a protein analyzed by ESI-MS is dissolved in an acidic solution with a high amount of miscible organic solvent (i.e., largely denaturing conditions). If the acid and organic solvent is omitted (i.e., a pure water solution)

the native, folded state of the protein can be maintained in solution. Through careful adjustments of the capillary temperature and skimmer potential it is possible to observe non-covalent protein interactions in the gas phase (Katta and Chait, 1991). This can be achieved if the desolvation thermal energy (capillary temperature, Figure 1.3) and the skimmer (gap) voltage are sufficiently reduced. One example of one of the earliest non-covalent complex observed in the gas phase by ESI-MS was the heme-globin complex in native myoglobin performed in the Chait lab (Katta and Chait, 1991). The study of non-covalent interactions in the gas phase by ESI-MS (Smith and Light-Wahl, 1993) is being undertaken by several groups including Robert Anderegg (protein-peptide), Brian Chait (protein-ligand), Jack Henion (protein-ligand), Sheena Radford (protein-protein) Richard Smith (protein-DNA complexes), and Ken Standing (protein oligomeric complexes). (An excellent review of noncovalent interactions can be found in Przybylski and Glocker, 1996.) The mass analyzer used by most groups to study non-covalent interactions is the quadrupole ion filter, instruments that tend to have a limited m/z range. Non-covalent complexes sprayed from pH 7 aqueous solutions are less charged than if they were sprayed from acidic solutions (a pH at which non-covalent complexed tend to dissociate.) Low charge (z) means high m/z values (e.g., m/z > 3000) so quadrupoles that only scan up to say m/z 2400 cannot be used. Some quadrupole instruments scan up to m/z 4000. As an alternate mass analyzer to the ion filter, Richard Smith has used an FT-ICR MS device and Ken Standing has demonstrated excellent results with a less expensive ESI-time-of-flight device.

1.4 Structural Characterization of Proteins

A principal application of mass spectrometry in this thesis is focused on the characterization of protein structure. The most powerful techniques available today to study protein structure are the high-resolution methods of X-ray crystallography and multidimensional nuclear magnetic resonance (NMR) spectroscopy. X-ray crystallography has been particularly fruitful, since the early heroic efforts of Max Perutz (hemoglobin) and John Kendrew (myoglobin). The three dimensional structures of >1000 proteins have been now determined. Simultaneous advances in the technologies of recombinant DNA cloning, DNA synthesis, and the expression of foreign DNA in heterologous systems as well as powerful x-ray sources, phase-determination strategies and powerful computers have paved the way to the explosive growth in protein structure determination of the 1990s (Eisenberg and Hill, 1989). Although the atomic resolution details of protein structure provided by X-ray crystallography and NMR are unsurpassed, there are a large number of other methods that have been used to obtain protein structural information. These techniques are considered as 'low resolution' approaches that include genetic (e.g., mutational), spectroscopic (e.g., circular dichroism), chemical methods (e.g., chemical modification and hydrogen/deuterium exchange) and biochemical methods (e.g., proteolysis). The last method has played a prominent role throughout the research described in this thesis.

1.5 Proteolysis

Proteolysis is a phenomenon recognized from early times. The first protease to be isolated was trypsin by Kühne in 1890 from pancreatic extractions (Mihalyi, 1978). A better understanding of proteolysis at the molecular level came only after the purification of proteases in the 1920s. In the 1930s Linderstrøm-Lang recognized that the peptide bond was hydrolyzed during proteolysis (Linderstrøm-Lang et al., 1938). By the end of the 1950s, the concept of partial or 'limited proteolysis' evolved from a structural study of fibrous proteins (Harrington et al., 1959). The authors of this short study (one of whom was Peter Von Hippel) cleverly foresaw a use of proteolytic enzymes as a way of distinguishing 'ordered' and 'disordered' regions of proteins.

There are two categories of proteases based on their mode of cleavage. Exoproteases cleave amino acid residues sequentially from the N-terminal (aminopeptidases) or C-terminal (carboxypeptidases), whereas endoproteases cleave the peptide bond between residues within a polypeptide chain. Cleavage specificity of a protease varies from highly specific (e.g., endoprotease Lys-C targets lysines) to broadly specific (e.g., subtilisin targets almost any amino acid) (Hames and Rickwood, 1989)

The active site of a protease is where substrate peptide bond hydrolysis occurs. A great deal of effort has been devoted to understanding the precise mechanism of hydrolysis (Hames and Rickwood, 1989). Important for the thesis

research is the 'size' of the active site. In a seminal study, Schecter and Berger employed biochemical means to probe the size of the active site of the protease papain (Schechter and Berger, 1967). From their findings and guided by the X-ray analysis of another enzyme, lysozyme, the authors proposed a general model of the proteolytic enzyme-substrate complex (Figure 1.4) (Schechter and Berger, 1967). The model suggested that the active site of papain protease extends beyond the catalytic site (C in Figure 1.4) to include seven enzyme subsites (S) that coordinate interaction between the enzyme and an extended form of the substrate peptide (P) that includes six residues. The intermolecular interactions were shown to increase enzyme-subtrate binding. More recent experimental (Fontana, 1989) and modeling studies (Hubbard et al., 1994) support the notion that a substantial amount of conformational flexibility of the substrate is a prime determinant for limited proteolysis. In contrast, another modeling study downplays the significance of substrate flexibility and suggests that surface exposure of the polypeptide substrate is a better determinant for predicting sites of limited proteolysis (Novotny and Bruccoleri, 1987). These differences notwithstanding, a survey of the literature generally favors the notion that a combination of surface exposure and polypeptide flexibility of the substrate are highly correlated to the sites of limited proteolysis of proteins (Mihalyi, 1978; Fontana, 1989).

When a protein is subjected to denaturing conditions, the protein unfolds, exposing the polypeptide backbone to solution. Because many proteases will

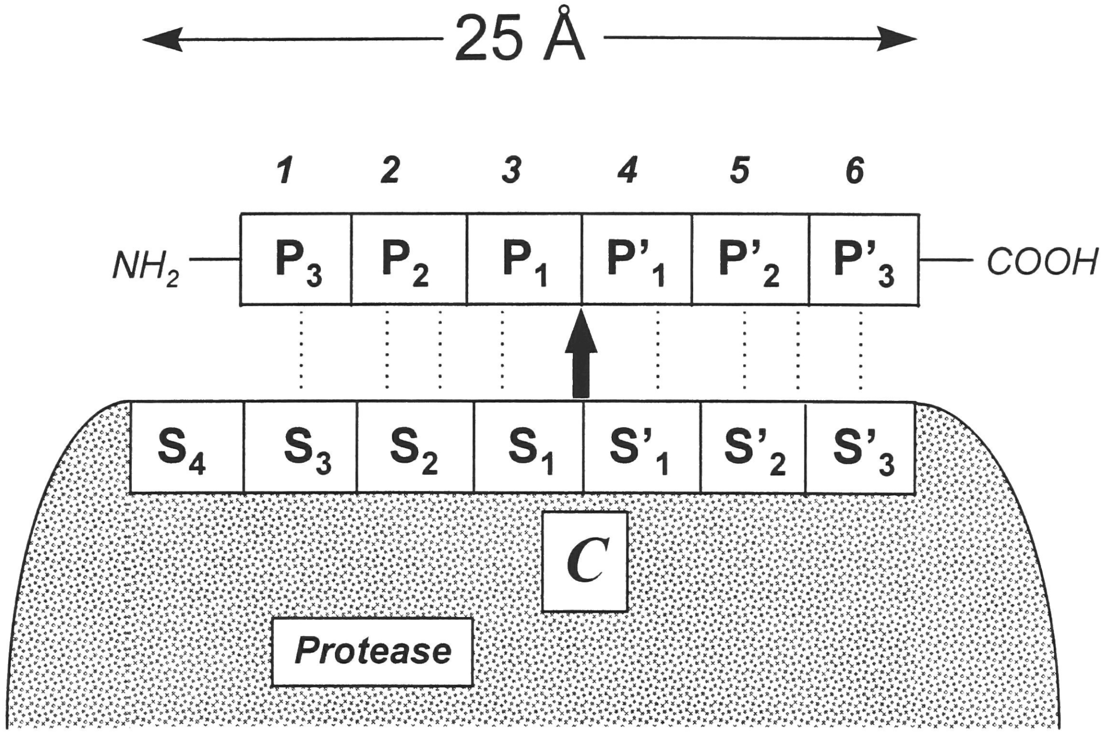


Figure 1.4 *Protease-extended substrate interactions.*

Schematic representation of a possible papain protease-substrate interaction.

The active site of papain protease is composed of seven subsites,**S** located on and numbered from both sides of the catalytic site, **C**. The 'primed' sites are on the C-terminal side of the catalytic site. The hexapeptide substrate positions, **P**, have the same numbering as the papain subsites they occupy. The arrow points to the target of proteolytic activity in the substrate (P_1/P'_1 scissile peptide bond) by the enzyme. The dotted lines imply general protease-substrate binding interactions required for complex formation and efficient hydrolysis. The length of the complex (25 Å) is based on a fully extended substrate peptide chain. (Adapted from Schechter and Berger, 1967.)

remain folded and active under mild to moderate denaturing conditions (Hames and Rickwood, 1989), the target protein will eventually give way to complete digestion by protease action. This form of digestion, usually referred to as peptide mapping, has for years facilitated protein identification, cDNA probe design, and protein sequence verification.

Conversely, subjecting a protein to proteolytic attack under non-denaturing (native) conditions greatly limits the activity of the protease. Under non-denaturing conditions, the protein is folded, and relevant protein-protein interactions are maintained. Polypeptide chain that is buried within the protein structure (such as a globular domain) is inaccessible to proteolytic attack. Residues that are exposed on the surface of the protein, however, may or may not resist proteolytic hydrolysis. According to the models discussed above, conformational flexibility of the polypeptide chain adjacent to the cleavage site is an important determinant for proteolysis. Thus solvent exposed residues within an α-helix or β-sheet should show resistance to proteolytic cleavage, whereas residues within extended flexible loops are expected to be highly susceptible.

The factors discussed above form the basis for using proteases to probe proteins for structure. There is a long history employing proteases as probes for structure, from studying conformational changes, protein unfolding and refolding events as well as for domain elucidation (Mihalyi, 1978; Hames and Rickwood, 1989). A majority of these studies relied on SDS-PAGE and N-terminal sequencing methods to analyze the digests. In this thesis, MS analysis is used

instead. Because of the speed and mass accuracy of MS, a large number of digests can be performed, revealing a deeper a more complete 'picture' of proteolytic action on a protein.

Chapter 2 — Experimental

This chapter details the materials and methods for MALDI-MS analysis used throughout my thesis research. It continues with the presentation of two studies concerning properties of the MALDI matrix and the matrix preparation step, a fundamental stage in the MALDI-MS analysis of peptides and proteins. Study I concerns the influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. Study II discusses special properties of the matrix solution.

2.1 Materials

2.1.1 Proteins, Peptides and Oligonucleotides

Max, USF, Hairy, SREBP1, Mef 2C, TAFs 42/62, leptin, TBP, TFIIB, and eIF-4C recombinant proteins were provided by various members of the laboratory of Stephen K. Burley (Rockefeller University) (see chapters 4 and 5 for sequences of selected proteins). Additional recombinant leptin samples also came from the laboratory of Jeffrey M. Friedman. Synthetic DNA oligonucteotides were from the Burley laboratory. Sequences of the proteins and DNA oligonucleotides and details of their preparation are given in the material sections found at the beginning of chapters 4 and 5. Dynorphins 1-11, 1-13, and 1-17 (mass calibrants) were provided by the Kreek laboratory. Bovine

cytochrome C, ubiquitin and equine myoglobin (protein controls and mass calibrants) were from Sigma (St. Louis, MO).

2.1.2 Proteases

Sequencing grade endoproteases Arg-C (Clostripain; *Clostridium histolyticum*), Glu-C (*Staphylococcus aureus* V8), Lys-C (*Pseudomonas aeruginosa*); and modified trypsin (bovine pancreas) were purchased from Promega (Madison, WI). Sequencing grade endoproteases Glu-C, Asp-N (*Pseudomonas fragi*) chymotrypsin (bovine pancreas), and subtilisin (*Bacillus subtilis*) were obtained from Boehringer Mannheim (Indianapolis, IN). A second form of sequencing grade Arg-C (mouse submaxillary gland) was purchased from Sigma. Preparation, storage and specificities of the proteases are given in Chapter 4.

2.1.3 MALDI Matrix and Matrix Solutions

The matrix used in a majority of the experiments to analyze peptides and proteins by MALDI-MS was 4-hydroxy-α-cyanocinnamic acid (4HCCA) obtained from Aldrich (Milwaukee, WI). 4HCCA was used almost exclusively throughout the studies because of its excellent sensitivity (Beavis et al, 1992; Chou, et al, 1994). In addition, I found that 4HCCA consistently yielded highly reproducible results. The 4HCCA was purchased several years ago (lot #HX06228PV) and was used without further purification. All other lots of 4HCCA from Aldrich or

Sigma were found to contain high levels of sodium and potassium salts that can be removed by recrystallization in water-methanol (1:1 v/v).

Other matrices employed were sinapinic (sinapic) acid (SA) (Aldrich) and 2,5-dihydroxy benzoic acid (DHB) (Sigma) that required no further purification. SA was best for proteins and peptides >2 kDa. DHB was best for analyzing phosphorylated peptides. Despite the popularity of DHB by some practitioners of MALDI-MS, I found DHB difficult to use for peptide analysis. The MALDI-MS signal, when using the DHB matrix, varied from spot-to-spot. I list some of my observations: (1) The MALDI-MS of a mixture of peptides from an enzyme digest gave ion peaks for the peptides whose relative intensities differed significantly from one spot to another of the same analyte-DHB sample. (2) The use of DHB matrix required much higher laser power than required for 4HCCA to observe peptide signal —higher laser intensities give rise to broader peaks and lower resolutions. (3) In contrast to 4HCCA, DHB crystals can neither be 'washed' nor used in slow crystallization. MALDI-MS of oligonucleotides were usually performed in 3-hydroxypyridine-2-carboxylic acid (3-HPA) (Lancaster, Windam, NH). The water for the matrix and sample solution preparations was supplied by a MilliQ UV Plus water purification system (Millipore, Bedford, MA). Formic acid (100% HPLC grade) was from Sigma and trifluoroacetic acid (TFA; HPLC/spectrograde) was from Pierce (Rockford, Ill). All other acids and organic solvents were HPLC grade or better.

2.1.4 Other Materials and Equipment

Cyanogen bromide was from Sigma. n-Octylglucoside was from Boehringer Mannheim. Pre-weighed amounts of N-octylglucoside were lyophilized into small Eppendorf tubes and stored desiccated at 4°C until needed. When DTT and urea was required, solutions were prepared fresh. Stock solutions of buffers and salts (analytical grade or better) were prepared and stored at 4°C or RT, normally in 125 mL glass bottles or 20 mL borosilicate glass scintillation vials (VWR) with plastic caps (avoid aluminum or paper lined caps). Some buffers required fresh preparation such as ammonium bicarbonate.

Microtubes and pipet tips were of the highest quality. This is a particularly important point when handling small volumes (few μL) of sample and matrix solutions. Mass spectrometry (ESI or MALDI) has a great ability to detect trace quantities of impurities and contaminants that can interfere partially or completely with the detection of the analyte. Two sizes of microtubes were used throughout —0.65 mL polyallomer micro-centrifuge tube (PGS Scientifics, Gaithersburg, MD) and 1.5 mL microtube (Starsedt, Newton, NC). Avoid the use of colored microtubes; they appear to contain high levels of contaminants that interfere with MALDI-MS. Three sizes of microtips (in racks) were used: 100-1000 μL (RT200) and 20-250 μL (RT20) tips and 1-10 μL tips (Eppendorf brand, Fisher Scientific). Less expensive 1-10 μL microtips are available (USA/Scientific Plastics, Ocala, Fl)but are not as rugged as the Eppendorf brand.

2.2 Methods

2.2.1 Proteolysis

A complete description of the proteolysis protocols is given in the methods section of Chapter 4. All protein digests were conducted in 0.65 μL microcentrifuge tubes. Typically, 15-30 μL of a 30 μM sample of recombinant protein in a solution containing buffer and salt was subjected to enzymatic digest at 25-37°C without any agitation. Enzyme:protein ratios ranged from 1:20-1:100 (w/w). For MALDI-MS analysis, 0.5 μL aliquots of the digest solution were mixed with 15 μL of MALDI matrix solution (see below) after a specified time of digestion, ranging from 30 s to 2 days. The acidity of the matrix solution (pH < 3) was sufficient to completely quench any further digestion.

2.2.2 Matrix Solution Preparation

Solutions of cinnamic acid derivatives (4HCCA, SA) were prepared saturated. They were made at room temperature by adding the solid matrix to organic solvent, followed by the addition of water and acid (as required). Microcentrifuge tubes (1.5 mL) were used for the solution preparation. Each mixture was thoroughly vortexed and centrifuged, leaving a clear, working matrix solution. The solubility of 4HCCA was dependent on the solvent composition, ranging from 5 mM (water:methanol, 2:1 v/v), 29 mM (formic acid:water:isopropanol, 1:3:2 v/v/v) to 74 mM (water:acetonitrile, 1:1 v/v). The

solubilities were estimated experimentally. Although formic acid can modify polypeptides through formylation, I have found that matrix solutions which contained formic acid at levels ≤ 17 % by volume (such as the 1:3:2 v/v/v solution mentioned above) did not give rise to formylation side reactions. SA matrix was usually prepared (saturated) in 0.1% TFA:ACN (2:1 v/v) and had about twice the solubility of 4HCCA dissolved in a solution consisting of the same solvents. A matrix solution of 4HCCA could be used for up to a week, although it is advisable to use a fresh preparation for critical work. Sinapinic acid matrix solution should be prepared fresh each day. When used, DHB matrix solutions were unsaturated. DHB was first prepared saturated in 0.1% TFA/ACN, H_2O/ACN or H_2O /MeOH. The saturated DHB solution is then diluted 1:4 making the working (unsaturated) matrix solution. See elsewhere for more details on using DHB as a matrix (Qin, 1996). All pH measurements of the matrix solutions were performed at room temperature with a combination pH microelectrode (model MI-415; Microelectrodes, Inc., Bedford, NH) and pH meter (model PHM 95; Radiometer, Copenhagen). The pH was also measured, albeit less accurately, with pH test strips (pH 0.0-6.0) (Sigma).

2.2.3 Matrix-Analyte Solution Preparation

The most extensively used matrix-sample preparation procedure I used (for 4HCCA and SA) was the dried-drop method (Karas and Hillenkamp, 1988). I also used slow (Xiang and Beavis, 1993) and rapid (Weinberger et al., 1993) crystallization. The methods are shown in Figure 2.1 and discussed below.

Figure 2.1 *MALDI-MS matrix crystallization methods*

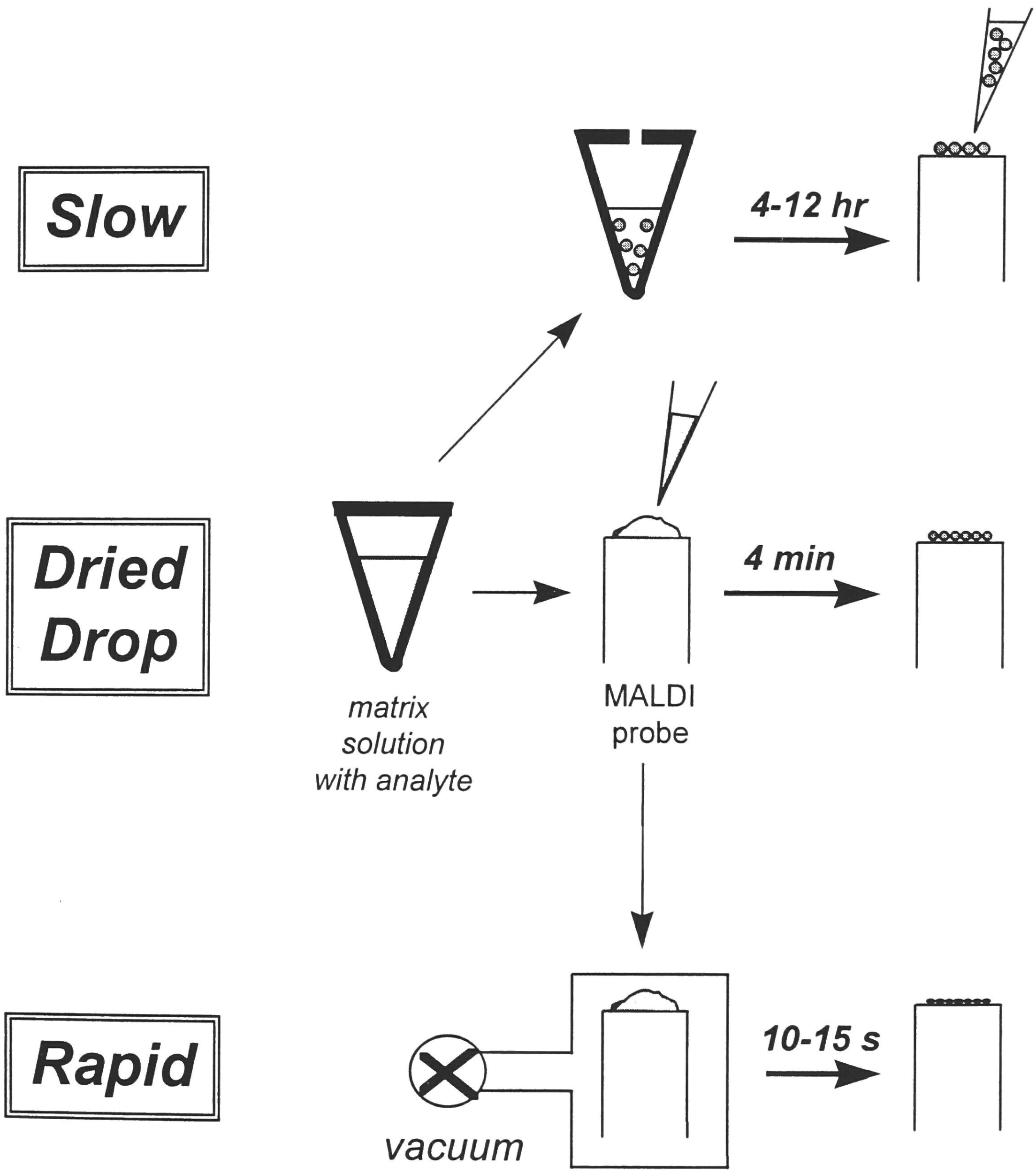
Three methods of matrix crystallization for the matrix 4HCCA are shown. Initially, the analyte (protein, peptide or mixture) is dissolved into the matrix solution in a small microcentrifuge tube. Three procedures can be used to obtain analyte-matrix co-crystals for deposition on a probe tip for MALDI-MS analysis —slow, dried-drop and rapid crystallization. See text for full details.

SLOW: The matrix-analyte solution is allowed to slowly (4-12 hr) evaporate directly from the tube. Analyte-matrix co-crystals, formed in the solution during the evaporation are retrieved and deposited on the probe for MALDI-MS analysis.

DRIED-DROP: The matrix-analyte solution is aliquoted onto the MALDI probe tip and allowed to air evaporate (~4 min) to form a bed of crystals.

RAPID: The matrix-analyte solution is aliquoted onto the MALDI probe tip that is immediately subjected to vacuum ($\leq$ 30 millitorr). The solution quickly evaporates (~10-15 s) leaving a fine bed of matrix crystals.

Matrix Crystallization



2.2.3.1 Dried-drop Crystallization (Karas and Hillenkamp, 1988): The dried-drop method of matrix crystallization was used throughout except where noted. In the dried-drop method, a small volume (~1 μL) of the analyte (e.g., peptide, protein or enzymatic digest) is mixed with a small volume (~10-30 μL) of matrix solution. Mixing (vortexing) should be performed in 0.65 mL microcentrifuge tubes. The microtube must be of high quality, free of plasticizers and other contaminants that can compromise high sensitivity MS measurements. Avoid the use of colored microtubes. After thoroughly vortexing, a small aliquot (<1 μL) of the matrix-analyte solution is deposited onto the MALDI-MS probe and allowed to air evaporate (~4 min). Normally, a uniform layer of fine, granular matrix crystals remain adhering to the probe. The crystals can be washed to help remove involatile salt residues. Cold water (2-4 μL) is placed over the crystals for a few seconds and removed by vacuum suction. (The 4HCCA or SA matrices are practically insoluble in pure cold water.) The microcrystalline surface layer is ready for MALDI-MS analysis.

2.2.3.2 Slow Crystallization Method (Xiang and Beavis, 1993): Although quite tolerant to the presence of involatile additives commonly found in biological samples (e.g., salts and buffers), the dried-drop method often fails to produce matrix crystals when the concentrations of additives are very high (i.e., typically >0.25 M) or liquid additives (e.g., glycerol) exceed a few percent in the matrix solution. To overcome the shortcomings of the dried-drop method, Beavis

introduced slow crystallization (Figure 2.1) as an alternate method for obtaining analyte-doped matrix crystals (Xiang and Beavis, 1993). In the slow crystallization approach, approximately 15-30 μL of the sample-matrix solution is left at room temperature in the microcentrifuge tube. A small hole (~1 mm diameter) is pierced through the cap of the tube permitting partial evaporation of the matrix solution. After 4-12 h, a portion of the solution evaporates, leaving small granular crystals of matrix adhering along the sides of the tube, submerged in the remaining supernatant solution. I have slightly modified the protocol to hasten crystallization. Instead of a small hole punched through the top of the microtube, the tube is left fully open and vortexed for 1-4 hr on a Tomy shaker (model MT-360, 36-sample microtube mixer; Tomy Tech, Palo Alto, CA). Vortexing produces a suspension of fine matrix crystals, much finer than the coarse crystals obtained using the original slow crystallization protocols (without vortexing). After crystals have formed, the tube is centrifuged, the supernatant removed, cold water is added and vortexed briefly to wash the crystals. The washing/centrifuging steps are repeated 2-3 times. The crystals are scraped off the walls of the tube with a small pipette tip and taken up in ~1 μL of water that is deposited directly onto the MALDI probe tip. Evaporation of the water leaves a bed of matrix crystals adhering to the probe ready for MALDI-MS analysis.

2.2.3.3 Rapid Crystallization Method (Weinberger et al., 1993): An aliquot (0.5 μL) of the sample-matrix solution is deposited onto the MALDI probe tip

which is immediately introduced into a vacuum (~20-30 mtorr) that is maintained by a mechanical rotary pump. The Speedvac (model SC110, Savant Inst. Co., Farmingdale, NY), *used without centrifugation*, is a convenient tool for this protocol. Secure a clean aluminum MALDI probe tip to its holder. Place the holder into the Speedvac, spot the matrix solution onto the probe tip and immediately close the lid to the Speedvac and apply vacuum. Viewing through the transparent cover of the Speedvac instrument, the evaporation was complete within 10-15 s leaving a fine, powdery layer of matrix crystals. With evaporation times much beyond 20 s, the effects attributed to rapid crystallization (discussed in Study I) become less pronounced. Peptides and proteins analyzed using the rapid crystallization method tend to exhibit extensive alkali cation adduction. The adduction is reduced by rinsing the crystals on the probe 3-4 times with portions (~5 μL) of cold water.

Investigation of the matrix-analyte preparation and crystallization steps developed into a major study, the results of which had significant impact on the research described in my thesis. I devote two separate sections to these matrix solution studies at the end of this chapter

2.2.4 Mass Spectrometry

2.2.4.1. MALDI-MS-Instrumentation: MALDI-MS was performed on a linear time-of-flight device constructed earlier at Rockefeller by Chait and Beavis (Beavis and Chait, 1989; Beavis and Chait, 1990a; Beavis and Chait, 1990b).

The bare essentials of the instrument consist of an ion source, a flight tube and a ion detector (discussed in more detail in chapter 1 and Figure 1.2). The three Chait and Beavis papers (above) describe the instrument in full detail. A good source for practical information concerning MALDI-MS can be found in Beavis and Chait (1996). Here I will discuss aspects of the instrument that were particularly relevant for my thesis research.

2.2.4.2 The MALDI Probe: Sample introduction into the ion source was via a sample probe. The probe was essentially a metal rod at the end of which was secured a sample probe tip. Aliquots (0.5-1 μL) of analyte-matrix solutions were deposited onto the probe tip for the dried-drop and rapid crystallization methods. Matrix crystals taken up in liquid were placed on the probe for the slow crystallization method (see Figure 2.1). I accumulated in excess of eight thousand spectra over a four year period, so the probe tip design had a significant impact on speed and ease of data collection. For example, MALDI-MS measurements made before Dec., 1994 were performed using a single-stage stainless steel sample probe tip (~2 mm diameter), whereas all subsequent measurements were made using a 10-position aluminum multiple probe tip. The ability to perform MALDI analysis on up to ten samples per introduction into the ion source with the multiple tip proved to be an invaluable experimental feature. The multiple probe design was based on contributions from all members of the laboratory. The probe assembly and probe tips were constructed by John

Doherty, supervisor of the Rockefeller University Instrument Shop. A large number of the aluminum sample probe tips were fabricated at low cost, allowing storage of selected samples for re-analysis, if necessary.

A note concerning the multiple probe tips. Cleaning of the tips (using gloves) is as follows. Rinse tip with acetone and methanol (from wash bottles), then sand the surface of the tip, face down, with 4/0 grade emery paper (Gamzon Bros., Inc., 21 W. 46th Street, NYC) wetted with isopropanol. Rinse with acetone/MeOH, repeat sanding, rinse again and wipe with Kimwipe. DO NOT USE WATER FOR CLEANING THE PROBE! (Water-cleaning of the probe results in MALDI mass spectral peaks with extensive copper adduction, especially if the matrix solution is 4HCCA + TFA/ACN or H2O/ACN)). Dab a bit of Apiezon high vacuum grease "L" (Manchester, England) onto the clean, dry tip surface, wipe across whole surface and rub off to leave a ultrathin layer of grease. The grease is necessary to prevent excess spreading of certain solvents having low surface tensions, particularly formic acid and isopropanol. The grease doesn't produce problematic background peaks in MALDI-MS at the pmol sensitivity level. In lieu of the grease, the probe tip can be rubbed hard across a piece of Teflon a few times. A thin layer of Teflon coats the surface. However, I find copper adduction peaks (especially using water/acetonitrile matrix solution) more prevalent in the MALDI spectra run from Teflon-coated than from the grease-coated tips. (I am grateful to Dr. Ivan Haller of the Rockefeller Mass Spectrometry group for suggesting the Teflon-coating idea.)

2.2.4.3. General Instrument Conditions: A prepared probe tip was inserted into the ion source and allowed to reach high vacuum. Background pressure within the instrument, measured by an ion gauge located below the source, normally reached better than 5 x 10^{-7} torr. A Neodymium:YAG laser (Lumonics Inc., Ontario, Canada) was set to deliver 355 nm wavelength pulses (approximately 10 ns duration) onto the sample at a rate of 2.5 Hz. Each laser shot produced a full mass spectrum. The spectra presented in this thesis represent averages of 100-200 laser shots. Averaging improved the signal-to-noise ratios and sampling statistics. The ion acceleration potential was +30 kV and the flight tube length was 2 m. An electrostatic particle guide mounted within the flight tube was used to compensate ion beam divergence and maintain good sensitivity levels when analyzing proteins. Ion detection and signal amplification was through a conversion microchannel plate detector-discrete dynode multiplier assembly (Beavis and Chait, 1991). The multiplier voltage ranged between 3.5-4 kV; the higher voltages are appropriate for proteins and high sensitivity measurements.

2.2.4.4. Data Collection. Signal from the electron multiplier was amplified and sent to the data collection system. For MALDI spectra collected before Aug., 1995, the amplified signal from single laser shots was visually monitored with a digital oscilloscope (model 7200A, LeCroy Corporation, Chestnut Ridge, NY). Simultaneously, the signal was digitized by a transient recorder (LeCroy, model

TR8828D) and stored as time-of-flight data on a Vax 4000 workstation (Digital Equipment Corporation, Woburn, MA). The data was viewed and manipulated using the program APPLE written by Ronald Beavis and modified by Wenzhu Zhang and David Fenyö. For MALDI spectra collected after Aug. 1, 1995, the digital oscilloscope was a Model 9350M, (LeCroy) and the mass spectra were collected on a Pentium PC using the program AQUIRE in a Windows (3.1, 95 or NT) environment. The collected spectra were viewed on the PC using the program M/Z, which is publicly available on the World Wide Web (URL http://mcphar04.med.nyu.edu/software/contents.htm). Both AQUIRE and M/Z were written by Ronald Beavis and were quite powerful and simple to use. Like the multiple probe, the AQUIRE and M/Z programs greatly facilitated my research. Spectra collected on the VAX is also readable by the program M/Z. Mass spectra run on M/Z can be conveniently converted into 'metafiles' for import into Microsoft Powerpoint graphics program. All 8000+ spectra are digitally stored (archived) on a single 1.2 Gbyte Verbatim Optical Disk (Club Mac, Irvine, CA). Access to the files (spectra, papers, etc.) stored on the Optical Disk is through the Silicon Graphics computer (SGI) using protocols written by Ruud Steenvorden. Files can be transferred between computers on the internet via FTP. In the future, all of my files will also be stored on a recordable CD.

2.2.4.5 Mass Calibration: Raw data collected from the MALDI instrument (either on the VAX or PC) was in the form of time-of-flight data. The TOF data

was mass converted either by external calibration or by using internal ion peaks of known mass —on the VAX using the APPLE program or on the PC using M/Z.

2.3 Study I: Influence of Matrix Solution Conditions on the MALDI-MS Analysis of Peptides and Proteins.

The matrix-analyte preparation step required an extensive investigation in order to optimize the MALDI-MS analysis of the complex peptide mixtures produced by enzymatic digests of proteins. The results of this study impacted greatly on my research. I devote the rest of this chapter to detailing this study (Cohen and Chait, 1996).

2.3.1 Background

There are many factors that influence the quality of MALDI mass spectra. Because MALDI-MS is a relatively new technique, many of these factors have been little explored. For example, it is generally accepted that the matrix solvent composition and sample-matrix preparation procedures greatly influence the MALDI-MS response. To date, a number of different sample-matrix preparation methods have been developed, such as the dried-drop (Karas and Hillenkamp, 1988) and slow crystallization procedure (Xiang and Beavis, 1993) (discussed above and shown in Figure 2.1). Slow crystallization represented a significant improvement over the dried-drop method in the tolerance of MALDI to high levels of involatile additives (Xiang and Beavis, 1993). In addition, I have observed

other features of slow crystallization that proved to be important for the matrix-solution study (described below) as well as for the development of methods to analyze whole proteins from gels using MALDI-MS (Chapter 3). More recent studies have demonstrated the uses of a polycrystalline thin film of matrix as a useful preparation procedure (Vorm et al., 1994; Xiang and Beavis, 1994). There are other factors such as the choice of matrix (Chou et al., 1994; Gusev et al., 1995) and matrix additives (Billeci and Stults, 1993; Gusev et al., 1995) that have been claimed to significantly influence the MALDI-MS response of peptides and proteins. The additives consist of simple sugars such as fucose. I've had little success in obtaining any positive results using the matrix additive fucose. Despite over five years of development of MALDI-MS, an understanding of the influences of other factors such as the matrix solvent composition on the quality of the MALDI spectra have been lacking —much of the knowledge of these factors has been anecdotal. To this end, I undertook a detailed investigation of the matrix 4HCCA. Dramatic mass discrimination effects in the MALDI-MS response of peptides, proteins and their mixtures were observed. The discrimination effects are greatly influenced by the composition of the sample-matrix solution, the solution pH, and the time scales used for growing the sample-matrix co-crystals. My findings provided guidance to optimize sample-matrix preparation procedures, permitting successful analysis of the majority of components in complex mixtures of peptides and proteins.

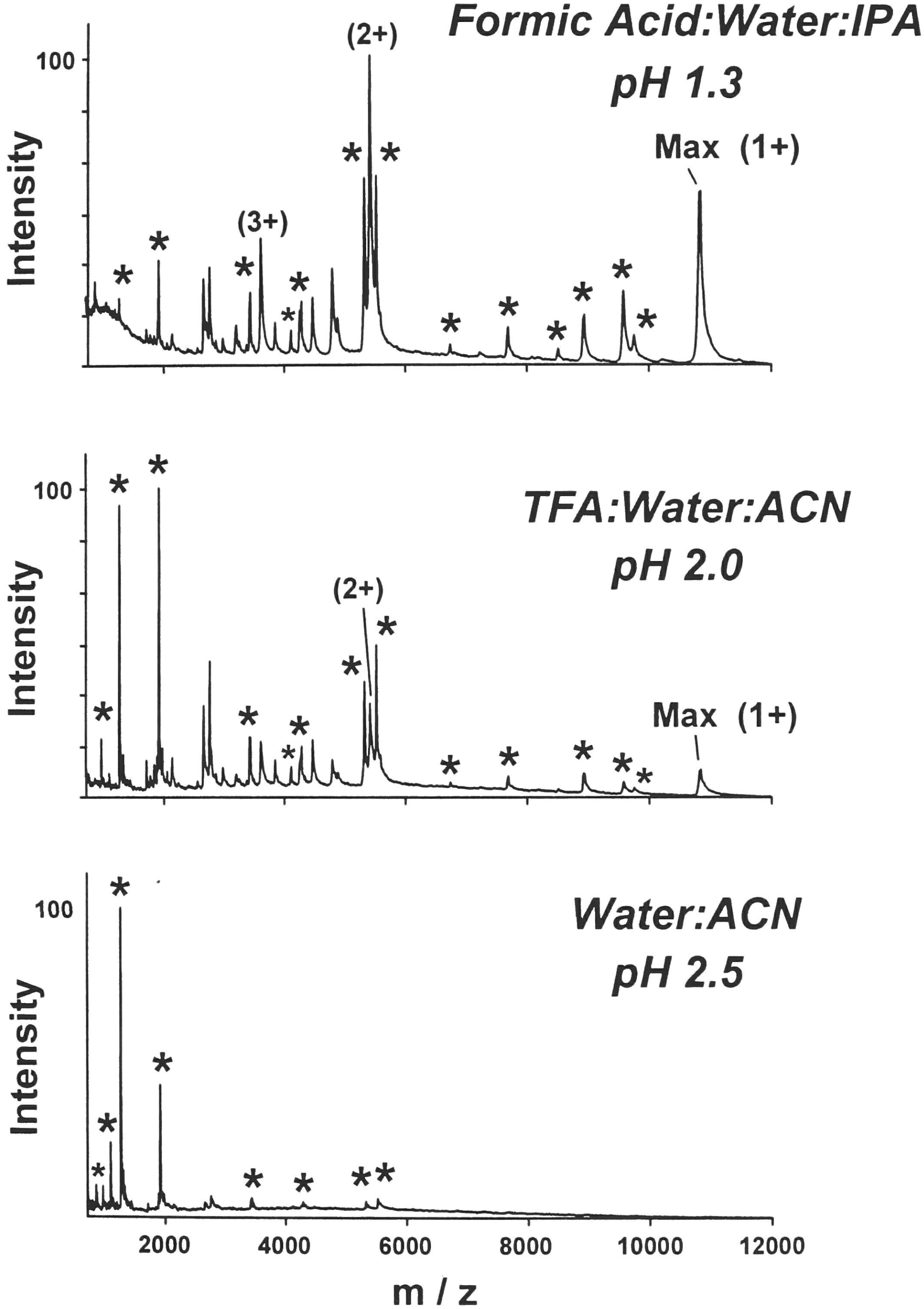
2.3.2 Results

2.3.2.1 Effects of the Matrix Solution Composition

I have found striking effects of the matrix solution composition on the MALDI-MS of peptide and protein mixtures. An example of these effects is illustrated in Figure 2.2, which compares mass spectra of a mixture consisting of peptides and proteins prepared in three different 4HCCA matrix solutions using the dried-drop method (Karas and Hillenkamp, 1988) (see Method section). The mixture of peptides was prepared by partial V8 digestion of b/HLH/Z Max (10.8 kDa) (Cohen et al, 1995). b/HLH/Z Max is the DNA binding domain of the Max proto-oncogenic transcription factor (Prendergast et al, 1991; Blackwood and Eisenman, 1991) (see Chapter 4). Figure 2.2a shows the mass spectrum of the digest obtained from a matrix solution consisting of 1:3:2 (v/v/v) mixture of formic acid:water:isopropanol (pH 1.3) saturated with 4HCCA. I refer to this formulation as FWI. The FWI matrix solution was originally developed by Klaus Schneider in the Chait laboratory (Martin de Llano et al., 1993) and has been used with considerable success by myself and others in the laboratory for protein analysis. Figure 2.2b shows the results obtained from a saturated 4HCCA solution made with a 2:1 (v/v) mixture of 0.1% trifluoroacetic acid:acetonitrile (TFA/ACN) with pH 2.0. The TFA/ACN matrix solution formulation is commonly used for peptide and protein analysis by many practitioners of MALDI-MS (Beavis and Chait, 1990a). A comparison of Figures 2.2a-b shows that the MALDI response from

Figure 2.2 *MALDI-MS of peptide mixture using different matrix solutions.*

Comparison of the positive ion MALDI mass spectra of the partial V8 digest of b/HLH/Z Max obtained from three different 4HCCA matrix solutions using the dried-drop method. The matrix solution compositions consist of **(a)** formic acid:water:isopropanol (1:3:2 v/v/v); **(b)** 0.1% trifluoroacetic acid:acetonitrile (2:1 v/v); and **(c)** water:acetonitrile (2:1 v/v). Both partial and complete digest fragments are observed. Peaks that are marked 1+, 2+, and 3+ designate singly, doubly, and triply protonated undigested Max. Peaks arising from singly protonated fragments are marked by stars (*); multiply protonated fragments are denoted by crosses (+). The sequences of the fragment ions are given in Figures 4.5b,c.



Formic Acid:Water:IPA
pH 1.3
(2+)
(3+)
Max (1+)
TFA:Water:ACN
pH 2.0
(2+)
Max (1+)
Water:ACN
pH 2.5
100
Intensity
2000
4000
6000
8000
10000
12000
m / z

the sample prepared in the FWI solution favors the appearance of high mass components (>6 kDa) compared to the response from that prepared in the TFA/ACN solution. The third spectrum (Figure 2.2c) was obtained from a saturated 4HCCA solution consisting of a 2:1 (v/v) mixture of water:acetonitrile (pH 2.5), where the acidity derives from the moderately acidic 4HCCA matrix (pK_A ~4 in water). This latter preparation yields a spectrum dominated by low mass (<2 kDa) but devoid of high mass peaks (>6 kDa). I have analyzed partial digests of many other proteins including cytochrome C, TAFs 42/62, USF, and leptin using the three 4HCCA matrix preparations described above and have found mass discrimination effects similar to those presented in Figure 2.2

The differences in the MALDI response in Figure 2.2 can be attributed to factors relating to the matrix solution composition and pH. These effects were further explored by examining the mass spectrometric response of the Max digest using a large variety of matrix solution compositions (Table 2.1). The acidities of these matrix solutions range from pH 1.1 (4HCCA/0.1N HC1/acetonitrile) to pH 2.9 (4HCCA/water/methanol). The results listed in Table 2.1 follow the same trends observed in Figure 2.2, showing a strong correlation between the matrix solution acidity and the MALDI-MS response to the different components in the digest. Highly acidic matrix solutions (pH <1.8), regardless of the identity of the added acid or organic solvent, showed weak or no signal for peptides <2 kDa and mainly favored the appearance of components with masses >2 kDa. In addition, matrix solutions that contained formic acid

TABLE 2.1 *Effects of matrix solution composition and pH on the MALDI-MS of a complex mixture of peptides*[a]

Composition of 4HCCA Solution	pH	Response[c]
0.1 N HCl / ACN (2:1)[b]	1.1	(l), **i**, h
1 % TFA / ACN (2:1)	1.1	(l), **i**, h
Formic acid / Water /ACN (1:3:2)	1.1	(l), i, **h**
Formic acid / Water / IPA (1:3:2)	1.3	(l), i, **h**
0.5 % TFA / ACN (2:1)	1.4	(l), **i**, h
Formic acid / Water MeOH (1:3:2)	1.6	(l), i, **h**
0.01N HCl / ACN (2:1)	1.9	l, **i**, h
Formic acid / Water / IPA (0.2:3:2)	2.0	l, **i**, h
0.1 % TFA / ACN (2:1)	2.0	l, **i**, h
0.1 % TFA / MeOH (2:1)	2.0	l, **i**, h
0.1 % TFA / IPA (2:1)	2.1	l, **i**, h
2 M Acetic acid / ACN (2:1)	2.3	l, **i**, h
Water / ACN (1:1)	2.4[d]	**l**, i, (h)
Water / ACN (2:1)	2.5[d]	**l**, i, (h)
Water / IPA (2:1)	2.7[d]	**l**, i, (h)
Water / EtOH (2:1)	2.8[d]	**l**, i, (h)
Water / MeOH (2:1)	2.9[d]	**l**, i, (h)

a. Peptide mixture is a partial V8 digest of b/HLH/Z Max protein (see chapter 4).
b. The matrix solutions were prepared by volume.
c. Response refers to a measure of relative peptide intensity, estimated from the mass spectra obtained for each matrix solution. The response is divided over three mass ranges designated by the letters l = low (0.8 - 2 kDa); i = intermediate (2 - 6 kDa); h = high (6 - 11 kDa). For each solution the three mass ranges (l, i, or h) are assigned a response level —the underlined letter in ***boldface*** designates the mass range exhibiting the strongest relative response; letters in *normal face* indicate a moderate response and the letter in *parenthesis* indicates a weak or no response.
d. The acidity of these solutions derives from the moderately acidic 4HCCA (pK_A~4).

and whose pH was <1.8 consistently yielded the strongest response to high mass components. Matrix solutions with pH between 1.8-2.3 exhibited mass spectrometric peaks that spanned the greatest latitude in mass range, although tending to favor the low and intermediate over the high mass components. The matrix solutions with pH >2.3 (i.e., no added acids), regardless of the identity of the organic solvent, consistently showed few components above 6 kDa and strongly favored the appearance of peptides below 2 kDa.

2.3.2.2 Effects of the Detergent n-Octylglucoside

I speculated that the above-described pH-dependent discrimination effects (Figure 2.2 and Table 2.1) may arise from differences in polypeptide solubility, and tested this hypothesis by monitoring the effects of the addition of n-octylglucoside on the MALDI-MS response of the V8 digest of b/HLH/Z Max. n-Octylglucoside is a non-ionic detergent that is known to aid in polypeptide and protein solubilization (Baron and Thompson, 1975). It has been previously shown that the presence of n-octylglucoside in the matrix solution enhances the MALDI-MS response of a large peptides in digest mixtures (Vorm et al., 1993). Figure 2.3 displays the MALDI spectra of the V8 digest of the Max protein. The spectra were obtained from a 4HCCA/water:acetonitrile (2:1 v/v) matrix solution in the absence (Figure 2.3a) and presence (Figure 2.3b) of 20 mM n-octylglucoside. In the absence of detergent from the matrix solution, high-mass discrimination effects are observed. In the presence of n-octylglucoside,

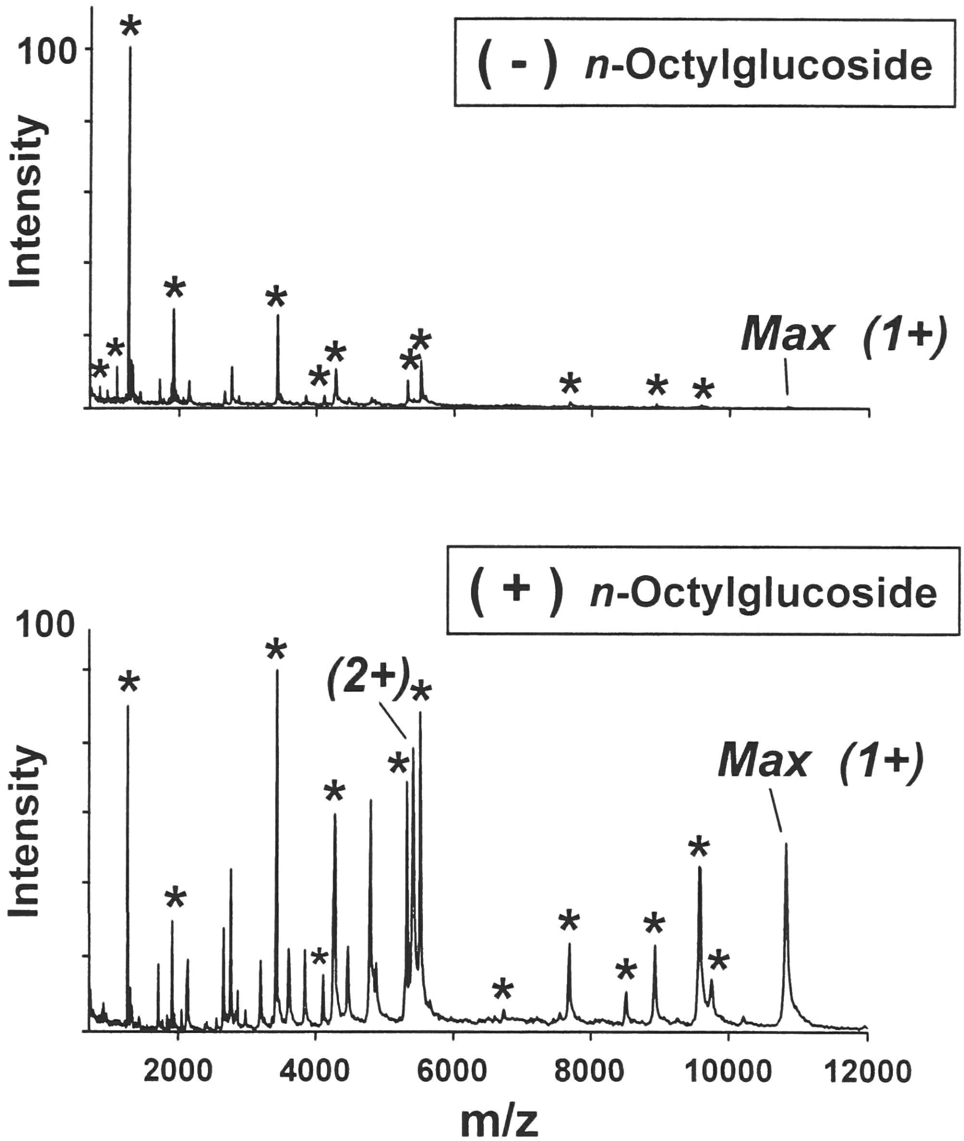


Figure 2.3 *Effects of detergent on MALDI-MS of a protein digest.*

Effects of n-octylglucoside on the MALDI mass spectra of a partial (2 min) V8 digest of b/HLH/Z Max. Spectra were obtained from a matrix solution consisting of 4HCAA-water/acetonitrile (2:1 v/v) in the **(a)** absence or **(b)** presence of 20 mM n-octylglucoside using the dried-drop method of matrix crystallization. See Figure 2.2 caption for a full description of the symbols.

the high mass discrimination effects are significantly reduced. Virtually all of the high mass polypeptides are 'restored' in the spectrum (Figure 2.3b), which strikingly now closely resembles the spectrum obtained using the FWI matrix solution (Figure 2.2a). These results suggest that the observed pH-dependent matrix solution mass discrimination effects may be related to polypeptide solubility.

2.3.2.3 Effects of Droplet Evaporation

I suspected that the evaporation of the droplet of matrix solution was accompanied by a change in pH, an effect that may influence the incorporation of analyte into the matrix crystals. Initial evaporation of the drop involves preferential loss of organic solvent. As the organic solvent evaporates, the solubility of the matrix, a moderately weak acid, falls sharply. A simple relationship describing the matrix dissolution and dissociation in a solution of water/acetonitrile is given in equation 1.

$$\underbrace{\mathrm{MH\ (s)} + \mathrm{H_2O/ACN} \rightleftharpoons \mathrm{MH\ (aq)}}_{\text{Crystallization / Dissolution}} + \mathrm{H_2O/ACN} \rightleftharpoons \mathrm{M^{-}\ (aq)} + \mathrm{H_3O^{+}/ACN} \qquad (1)$$

In equation 1 the matrix (M), with its acidic hydrogen, is denoted as MH.

I performed a simple experiment to measure the pH of evaporating drops of matrix solution. In this experiment, the pH of droplets of matrix solution (initial volume, 5 μL) was monitored as they were allowed to air evaporate at room temperature. A glass microelectrode measured the pH. Results of the experiment are given graphically in Figure 2.4, which shows droplet pH as a function of time. The results indicate that matrix solutions prepared without added acid (e.g., 4HCCA in H_2O/ACN) experience an *increase* in pH as the droplet evaporates. In contrast, Figure 2.4 shows that matrix solutions with added acids (e.g., 4HCCA in FWI or TFA/ACN) experience a *decrease* in pH with droplet evaporation. I rationalize these results as follows. In the absence of added acid, the matrix is the dominant source of H_3O^+ in the droplet (Eq. 1). As organic solvent evaporates, Eq. 1 is driven towards the left (i.e., matrix crystallization) resulting in the removal of H_3O^+ from solution and pH elevation. In contrast, the presence of added (strong) acids ensures a high concentration of H_3O^+. As organic solvent evaporates and Eq. 1 is driven towards the left, the strong acids remaining in the droplet maintain the supply of H_3O^+ ions. The H_3O^+ ions concentrate in the shrinking droplet, leading to the decline in pH.

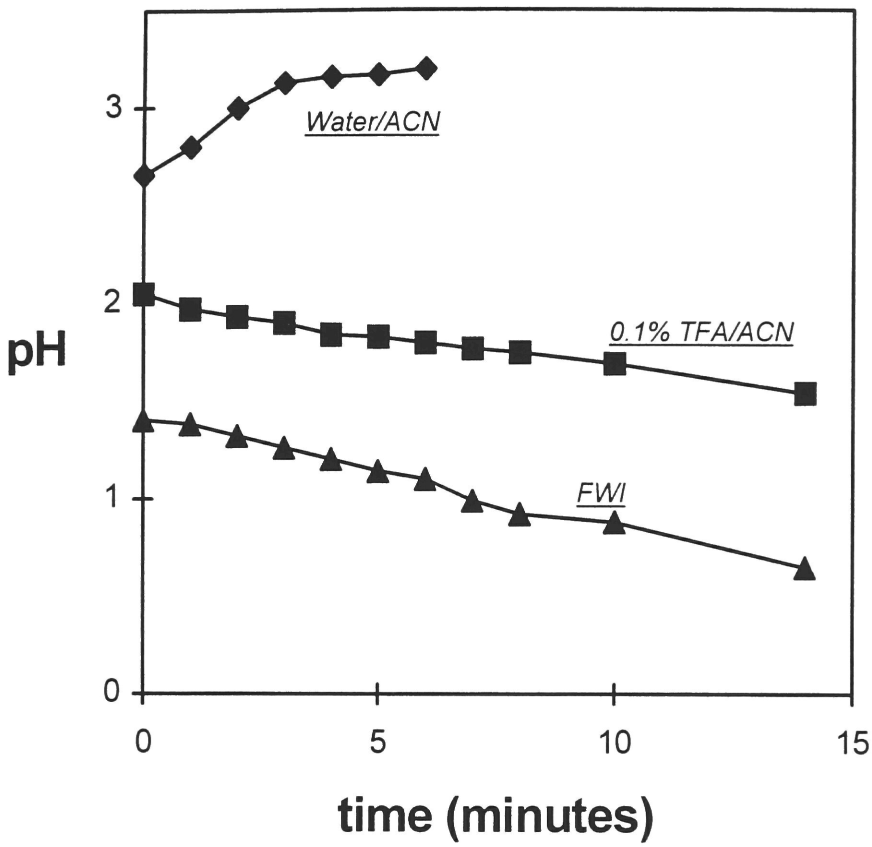


Figure 2.4 *pH profile of an evaporating drop of matrix solution.*

The pH profile of evaporating drops (5 μL initial volume; RT) of three different matrix solutions (saturated with the matrix 4HCCA) monitored by a pH microelectrode. The matrix solutions consisted of the following solvents: water/ACN (2:1 v/v) (*diamond symbols);* 0.1 % TFA/ACN (2:1 v/v) (*square*); and formic acid/ water/isopropanol (FWI) (1:3:2 v/v/v) (*triangle*). Evaporation was nearly complete by ~6 min for the water/ACN and 14 min for the other two matrix solutions.

2.3.2.4 Effects of the Rate of Matrix Crystal Growth

In addition to the matrix solvent composition, and pH changes during droplet evaporation, I have found that an important element of the sample-matrix preparation procedure is the rate at which matrix crystallization occurs. The most commonly used strategy to grow analyte-doped matrix crystals involves variations of the dried-drop method, a procedure that usually depends on a moderately quick (minutes) and complete evaporation of a small drop (~1 μL) of analyte-containing matrix solution (Karas and Hillenkamp, 1988). A second approach in growing matrix crystals is slow matrix crystallization (Xiang and Beavis, 1993), a method in which a moderate volume of matrix solution (~15 μL) is allowed to slowly (hours) and partially evaporate, a process that is accompanied by matrix crystal formation within the remaining matrix solution. A third approach involves a rapid (seconds) and complete evaporation of a small drop of matrix solution (~1 μL) to obtain matrix crystals (Weinberger et al., 1993).

Using the three matrix preparation methods, I explored the effect on the MALDI-MS response when utilizing the three different time scales described above to grow analyte-doped matrix crystals. An example is given in Figure 2.5, which compares MALDI spectra of a peptide mixture using the slow, dried-drop and rapid crystallization methods for two different matrix solution compositions. The peptide mixture derives from a *complete* V8 digest of b/HLH/Z Max. The complete digest of Max produces principally six peptide fragments, labeled 1 - 6 in Figure 2.5. The mass spectra in Figures 2.5a-c were obtained from a sample-

matrix solution made from the 4HCCA/formic acid:water:isopropanol (1:3:2 v/v/v) (FWI) formulation, while spectra in Figures 2.5d-f were obtained from a sample-matrix solution made from 4HCCA/water:acetonitrile (2:1 v/v). The spectrum in Figure 2.5a, obtained from the FWI matrix solution using slow crystallization, shows exclusively high mass peptide fragments 5 and 6 and is dominated by the latter. No traces of fragments 1-4 appear in the spectrum. The spectrum in Figure 2.5b, obtained from the same FWI solution using the dried-drop method, also only shows fragments 5 and 6. In stark contrast, the spectrum in Figure 2.5c, obtained from the FWI solution using the rapid crystallization technique, shows all six fragments, a strong indication that rapid matrix crystallization can 'overcome' the low mass discrimination effects observed from either slow crystallization (Figure 2.5a) or the dried-drop methods (Figure 2.5b).

Results using the 4HCCA/water-acetonitrile matrix solution (Figures 2.5d-f) reveal characteristics similar to those observed with FWI. The spectrum in Figure 2.5d, obtained using slow crystallization, shows only the two high mass fragments 5 and 6 (c.f. Figure 2.5a) whereas the spectrum in Figure 2.5a, obtained from the dried-drop method, shows all 6 fragments. The spectrum in Figure 2.5f, obtained from rapid crystallization, exhibits all six fragments and is dominated by the lower mass fragments 1-4.

The results illustrated in Figure 2.5 indicate that the rates of matrix crystal growth have striking effects on the MALDI-MS response of different mass components in a polypeptide mixture. I have performed similar experiments with

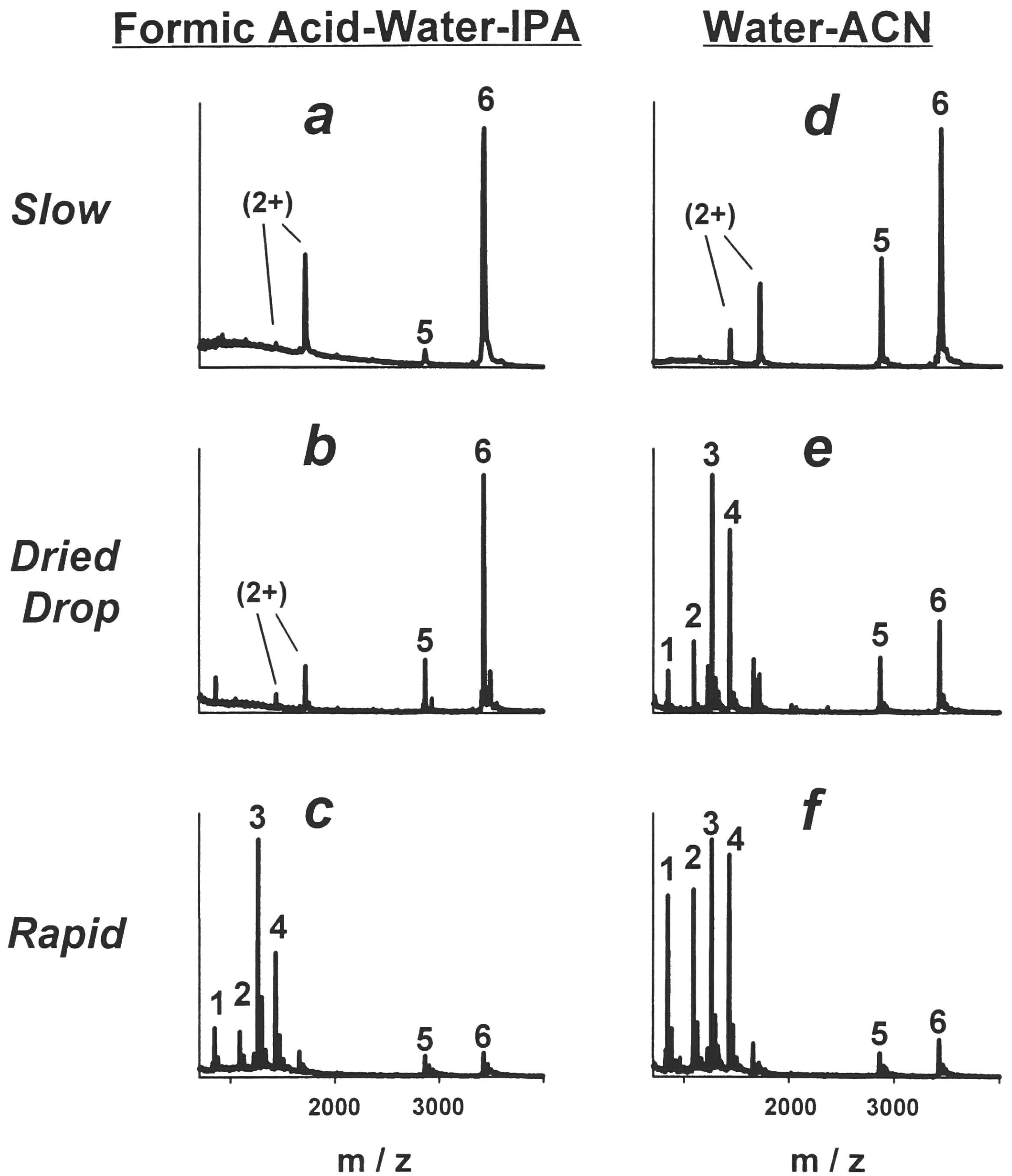


Figure 2.5 *Effects of the rate of matrix crystal growth on the MALDI-MS of a protein digest.*

Positive ion spectra of the complete V8 digest of b/HLH/Z Max obtained from two matrix solutions: **(a-c)** 4HCCA-formic acid/water/isopropanol (1:3:2 v/v/v) and **(d-f)** 4HCCA-water/acetonitrile (2:1 v/v). Results using three rates of matrix crystallization are illustrated: top panels (a,d), slow crystallization; middle panels (d,e), dried-drop; and bottom panels (c,f) rapid crystallization. A complete V8 digestion of Max results in six fragments that appear in the spectra as singly protonated peptides (labeled 1-6).

other polypeptides including a synthetic peptide ladder and enzymatic digestions of human leptin. The results from these experiments mirror the findings illustrated in Figure 2.5.

2.3.2.5 Mass Discrimination Effects between Proteins

I have shown that there are appreciable MALDI-MS mass discrimination effects from complex mixtures of peptides and proteins (Figures 2.2-2.5). The components of these mixtures ranged in mass from 0.8 to 11 kDa. I have also observed strong mass discrimination effects between small proteins whose mass differ from each other by < 1 kD. An example of this phenomenon is illustrated in Figure 2.5, which compares spectra run under various conditions of a protein that consists of two non-covalently interacting subunits with molecular masses 9.0 kDa and 9.9 kDa. These subunits are recombinant, truncated forms of the *Drosophila* proteins TAF62 (p62) and TAF42 (p42), factors that have been implicated in activated transcription (Kokubo et al., 1994) (also see chapter 5). The proteins share similar amino acid composition: the number of positively charged residues in p42 & p62 are 14 & 14, respectively; negatively charged residues 13 & 11; polar uncharged residues: 21 & 22; non-polar residues: 38 & 36. The pI (isoelectric point) values were also estimated (Sillero and Ribeiro, 1989) p42, pI=6.5 and p62, pI=9.9.

Figure 2.6 compares the MALDI response of the proteins using two matrix crystallization methods run in three different matrix solution preparations. Figure

2.6a-c shows results of the dried-drop method, while Figures 2.6d-f shows results using the slow crystallization procedure. The three preparations consist of 4HCCA in formic acid:water:isopropanol (1:3:2 v/v/v; FWI) (Figures 2.6a & d); 0.1% TFA:acetonitrile (2:1 v/v; TFA-ACN) (Figures 2.6b & e); and TFA-ACN with 8 mM n-octylglucoside (Figures 2.6c & f). The spectra reveal dramatic discrimination effects that are influenced by the matrix solution composition and the rates of growth of the matrix crystals. The spectrum obtained from the FWI matrix solution and the dried-drop method (Figure 2.6a) shows results that approach the 1:1 stoichiometry expected for the p42 and p62 subunits of the protein. In contrast, the spectrum obtained from the TFA-ACN matrix solution using the dried-drop method (Figure 2.6b) shows a strong signal from p62 and a weak signal for p42. The addition of n-octylglucoside to the TFA-ACN matrix solution restores the response of p42 (Figure 2.6c) to levels similar to that observed with FWI (Figure 2.6a). I contrast the dried-drop results (Figure 2.6a-c) with the findings using slow crystallization (Figures 2.6d-f). Again, discrimination effects are observed between the two proteins, but surprisingly in reversed order compared to the dried-drop method (compare Figure 2.6a with d and b with e). The addition of n-octylglucoside in the TFA matrix solution again restores the response of the discriminated subunit —this time p62 (Figure 2.6f).

I have observed discrimination effects using other 4HCCA matrix solution compositions (such as those listed in Table 2.1) between p42 and p62 as well as for other proteins (e.g., Max, TATA binding protein and TFIIB). In general, for

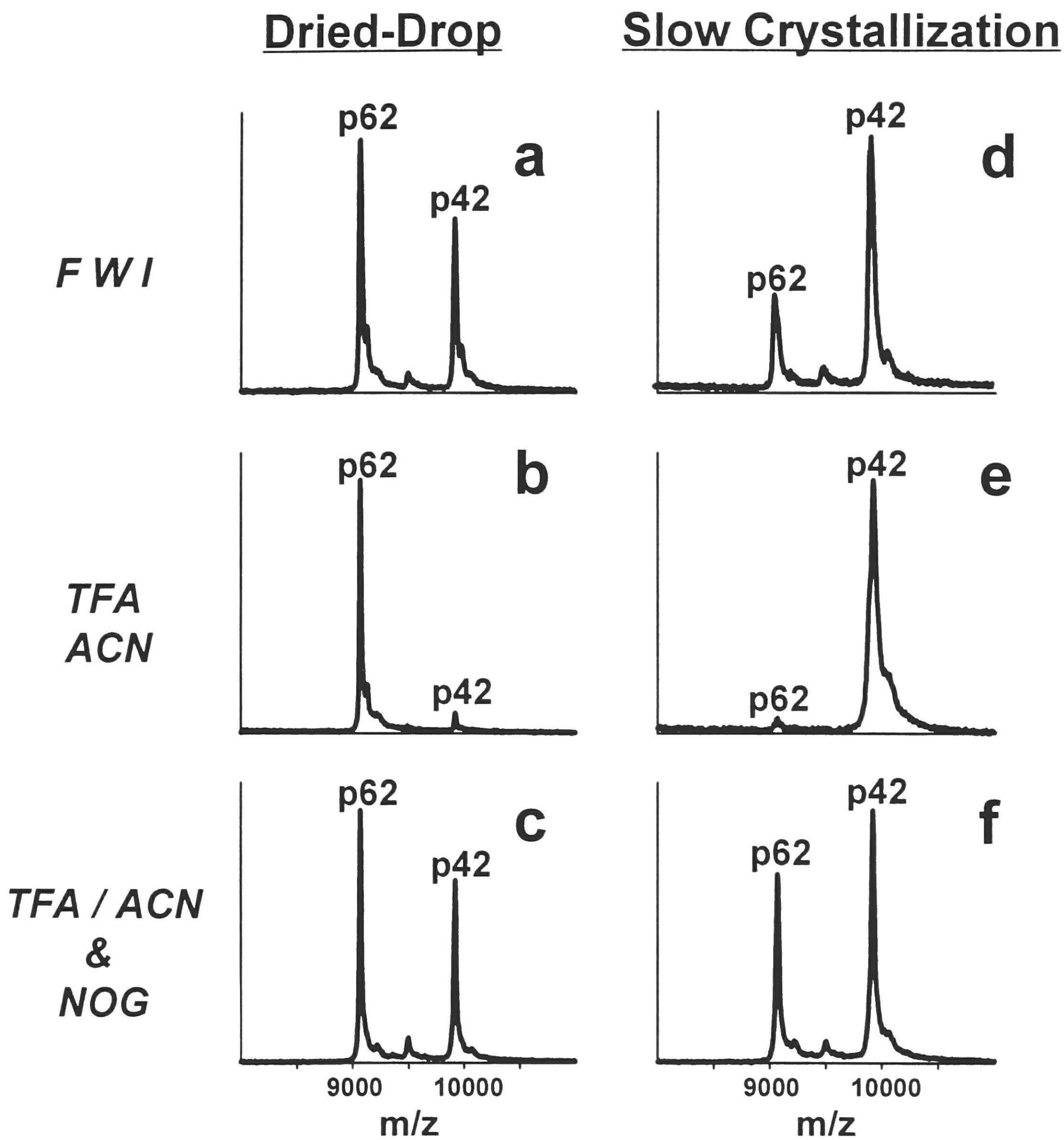


Figure 2.6 *Mass discrimination effects between proteins in MALDI-MS.*

Mass discrimination effects between truncated forms of the proteins TAF 42 and TAF 62 (labeled as p42 and p62 in the spectra). MALDI mass spectra were obtained using the dried-drop method (left column) or the slow crystallization method (right column). Spectra were obtained from three 4HCCA matrix solutions: top panels (**a,d**), formic acid/water/isopropanol (1:3:2 v/v/v); middle panels (**d,e**), 0.1% TFA/ACN (2:1 v/v); and bottom panels (**c,f**), 0.1% TFA/ACN (2:1 v/v) with 8 mM n-octylglucoside (NOG). Only a portion of the spectra that display the singly protonated proteins is shown.

mixtures of proteins of comparable masses that exhibit a propensity towards the discrimination effects illustrated in Figure 2.6, I find that: (1) matrix solutions that are prepared with formic acid and whose pH is < 1.8 are particularly effective in yielding MALDI spectra that more closely reflect the true stoichiometries of the proteins in solution; (2) matrix solutions containing TFA, acetic acid or HCl exhibit a tendency towards mass discrimination between the proteins; and (3) a moderate level of n-octylglucoside in the matrix solution can often reduce the mass discrimination effects.

2.3.3 Discussion

I have demonstrated that MALDI-MS analysis of complex mixtures of polypeptides (using the matrix 4HCCA) is accompanied by strong mass discrimination effects that are related to the sample-matrix preparation procedures. These effects are highly dependent on the rates at which the sample-matrix co-crystals are grown —slow, rapid or intermediate (dried-drop).

2.3.3.1 Slow Matrix Crystallization (hours)

The slow crystallization procedure favors the appearance of high mass components of complex mixtures, regardless of the matrix solution composition and pH (see Figures 2.5a & d). I hypothesize that during matrix crystallization, polypeptide components partition between the growing matrix crystal lattice and

the bulk solution. The tendency to partition between the crystal and the solution suggests a resemblance to the partitioning of polypeptides between a reverse-phase column and the solution phase in reverse-phase liquid chromatography. This idea is based upon a detailed physico-chemical investigation of the matrix sinapinic acid (a cinnamic acid derivative) undertaken by Beavis and Bridson (Beavis and Bridson, 1993). In this study the authors showed that polypeptides were attaching to the $(10\bar{3})$ face of the growing matrix crystal, a surface that was determined to be non-polar. They thus believed that 'hydrophobic' interactions could be a significant mode of polypeptide attachment to the matrix crystal. Borrowing from the concepts of reverse-phase liquid chromatography, the propensity of a polypeptide to adhere to the matrix crystal lattice would be a function of the mass of the peptide, a relationship that is consistent with our findings that show crystallization strongly discriminates against the incorporation of low mass components of mixtures.

2.3.3.2 Rapid Matrix Crystallization (seconds)

The rapid crystallization method favors the appearance of low mass peptides, regardless of the composition or pH of the matrix solution (Figures 2.5c & f). I believe that under rapid crystallization conditions, there is insufficient time for extensive partitioning of peptide components between the growing matrix crystal and the solution phase (as postulated for slow crystallization). Under these circumstances both low and high mass components in solution quickly become 'captured' in the matrix crystals, a phenomenon that would be largely

independent of the matrix solution composition and pH. Spectra obtained from rapid crystallization of complex mixtures of peptides show a relatively stronger response for low mass than for high mass peptide components. The difference in response between high and low mass peptides probably reflect a greater MALDI-MS sensitivity for the low mass peptides.

2.3.3.3 Dried-drop Matrix Crystallization (minutes)

Unlike the findings observed from the slow and rapid crystallization methods, the dried-drop method exhibits strong pH- and composition-dependent mass discrimination effects (Figures 2.2, 2.3, 2.5b & e and Table 2.1). I cannot fully rationalize these results. Although the microscopic events within a small drying drop of matrix solution have not been well characterized, a number of phenomena are likely to be important. (1) The organic solvent is expected to volatilize quickly from a drying drop of matrix solution, hastening matrix crystallization and leaving the solution enriched in the aqueous phase. (2) The pH of the droplet changes during its evaporation (Figure 2.4). (3) Related to (1) and (2) above are polypeptide solubility and a partitioning between the solution and matrix crystal phases. The results using n-octylglucoside suggest that polypeptide solubility may play an important role in the discrimination effects, especially from matrix solutions with pH > 2.3 (Figure 2.3). In addition, peptide solubility may be the reason for the particular benefits of using formic acid in the matrix solution, since formic acid is known for its strong polypeptide solvating properties (Heukeshoven and Dernick, 1982). (4) It is not clear why highly

acidic matrix solutions (pH < 1.8), regardless of the identity of the acid, would cause discrimination against the low mass peptides (cf. Figures 2.5b & e). The low mass discrimination may reflect a partitioning phenomenon as described for slow crystallization. In this scheme, low mass peptides would favor remaining in the acidic matrix solution instead of partitioning into the growing matrix lattice. It is also not clear how the *changing* pH of the drying drop would affect the partitioning events.

There are other factors that influence the MALDI-MS response, including peptide polarity and basicity (Liao and Allison, 1994). I and others have found that highly acidic peptides often yield poor or no response by MALDI-MS (Juhasz and Biemann, 1994). The matrix solution discrimination effects that I have characterized are observed irrespective of these other factors.

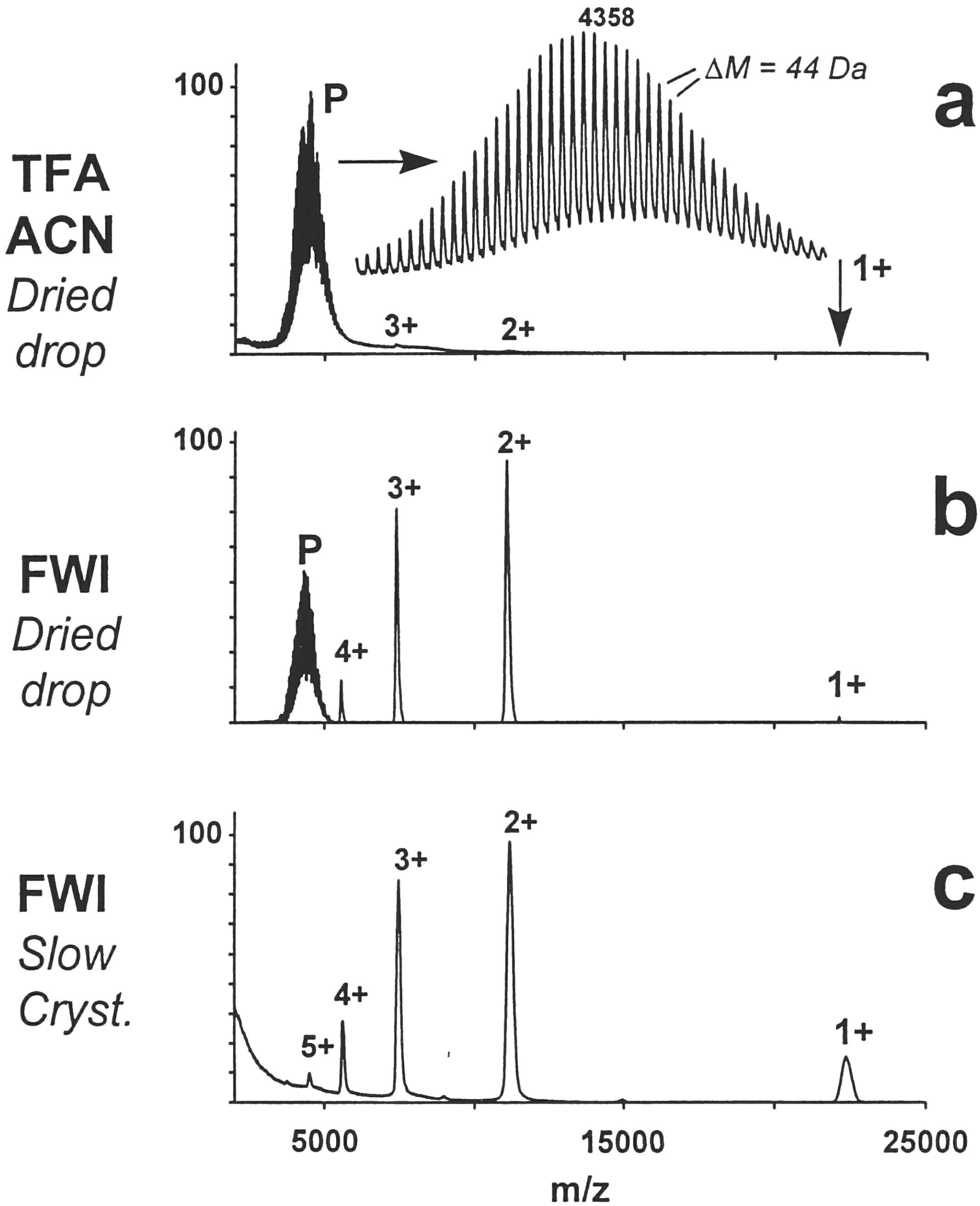
I include a note pertaining to the interpretation of my findings as it concerns quantitation and relative vs. absolute intensity of mass spectral peaks. First, my data interpretation has been based on the premise that the matrix preparation procedures affect mainly the extent to which an analyte incorporates into the matrix and on the assumption that such incorporation does not significantly effect the laser desorption and ionization properties of the analyte. Second, accurate quantitation of peptides and proteins by MALDI-MS is often quite difficult, highly dependent on the laser fluence levels and the nature of the polypeptide (unpubl. results and Tang et al., 1993). In all of my experiments the laser fluence was set slightly above the desorption/ionization threshold and varied only marginally between each sample. I have reduced the dependence

on quantitation and on the need for absolute sensitivity measurements by relying on comparative analyses of the mass spectra of samples consisting of the same components and interpreting the data based on changes in the relative peak heights. Any small variation in laser fluence did not greatly affect the relative intensities.

2.4 STUDY II: Special Properties of the 4HCCA/FWI Matrix Solution

2.4.1 Effects of Additives on the MALDI-MS of Proteins

There are special attributes concerning the 4HCCA/FWI matrix solution in addition to those discussed in the previous section. (1) 4HCCA/FWI was a rather 'robust' matrix. In the face of elevated levels of involatile additives, FWI out-performed the TFA/ACN matrix combination for protein analysis. An example is seen in Figure 2.7 which shows the MALDI-MS of a recombinant form of the eukaryotic initiation factor 4E (eIF-4E) binding protein (a.k.a. the small RNA cap-binding protein) (22.1 kDa). The protein sample contains polyethylene glycol (PEG) 4000, a commonly employed precipitating agent used in protein crystallization trials. The concentration of PEG in the matrix solution, after addition of protein, was about 2% by volume. The MALDI-MS results using the dried-drop method of matrix crystallization (Figures 2.7a-b) are quite striking. The TFA/ACN matrix solution spectrum (Figure 2.7a) is dominated by PEG



<u>Figure 2.7</u> *Effects of polyethylene glycol on MALDI-MS of proteins.*

MALDI-MS of a recombinant form of the small RNA cap binding protein (22.1 kDa). The matrix solvent composition and crystallization procedures effect the MALDI-MS response of the protein. The matrix solutions contain ~2% (by volume) polyethylene (PEG) 4000. 4HCCA matrix solution and crystallization methods are: **(a)** 0.1% TFA/ACN (2:1 v/v) by the dried-drop method; **(b)** Formic acid/water/isopropanol (FWI) 1:3:2 (v/v/v) by the dried drop method and **(c)** FWI by the slow crystallization method. PEG polymer (P) appears as a broad distribution of peaks (inset in **a**) centered at m/z 4358. Individual peaks within the polymer cluster are separated by 44 Da (C_2H_4O). The protein appears as a singly (1+) and multiply charged species (2+-5+).

polymer peaks and hardly shows any protein signal. In contrast, the FWI matrix solution shows a reduction in the signal from the PEG, but now with strong signal from the eIF-4E protein (Figure 2.7b). Slow crystallization (Xiang and Beavis, 1993) from the FWI matrix solution improves the signal-to-noise even further (Figure 2.7c). (Slow crystallization will be discussed more extensively in chapter 3.)

The 'robustness' of the 4HCCA/FWI matrix combination has proven to be quite beneficial in many situations, most significantly in the analysis of protein crystals used for high resolution X-ray crystallographic structural studies. I have use MALDI-MS to determine protein homogeneity, degradation products, and stoichiometry (for multi-component structures) of the mother liquor as well as the protein crystals for a number structural studies. The mother liquor or the protein crystals were directly dissolved into MALDI matrix solution. Polymeric additives such as PEG are normally present in the mother liquor as well as the protein crystals, so 4HCCA/FWI is the best choice of matrix solution (see Figure 2.7). I have applied this approach to several protein crystallization projects including the TBP-TFIIB-DNA ternary complex (Nikolav/Burely) (Nikolov et al, 1995), TAF 42/62 (Xie/Burley) (Xie et al, 1996) and the eIF-4C protein (Marcotrigiano/Burley).

2.4.2 Effects of DNA on the MALDI-MS of Proteins.

4HCCA/FWI has particularly interesting properties when analyzing protein samples that contain DNA oliogonucleotides (oligos). I highlight these findings

here. (1) 4HCCA gives a better protein MALDI response than sinapic acid in the presence of oliogs. regardless of the matrix solution composition. (2) FWI as a solvent system is superior than TFA/ACN to observe protein signal in the presence of oligos. (3) The presence of DNA oligos. tends to degrade the protein MALDI signal (positive ion MALDI) in a fashion that is dependent on the oligo. size and concentration —DNA oligos. longer than ~20 base pairs and an oligo:protein ratio (molar) greater than ~four begin to adversely affect the MALDI response from the protein. (4) Protein/DNA oligo. mixtures run in 4HCCA/FWI (or SA/FWI) give rise to protein-DNA adduct peaks that are completely absent when the same mixture is run using a TFA/CAN solvent system. The adduct peaks have a distinct 'shape' that can serve as a unique signature for the presence of DNA. (Normally, MALDI analysis of DNA oligos. is performed with non-cinnamic acid derivative matrices run in the negative ion mode, a combination that is unsuitable for protein analysis.) An example of the protein-DNA adduction is given in Figure 2.8 that compares the MALDI spectra of b/HLH/Z Max and a 14-mer double-stranded DNA oligonucleotide run in 4HCCA matrix with either TFA/ACN or FWI. The oligo. sequence is 5'-AGGTCACGTGACCT-3' and the protein:ssDNA mole ratio is 1:4. The spectra from both matrix solutions show the presence of the Max protein. However, the spectrum from the FWI matrix solution (Figure 2.8/RIGHT) exhibits a series of small peaks that correlate to Max-DNA adduct ions. Close inspection of these adduct peaks reveals an interesting asymmetry —the high-mass side of the

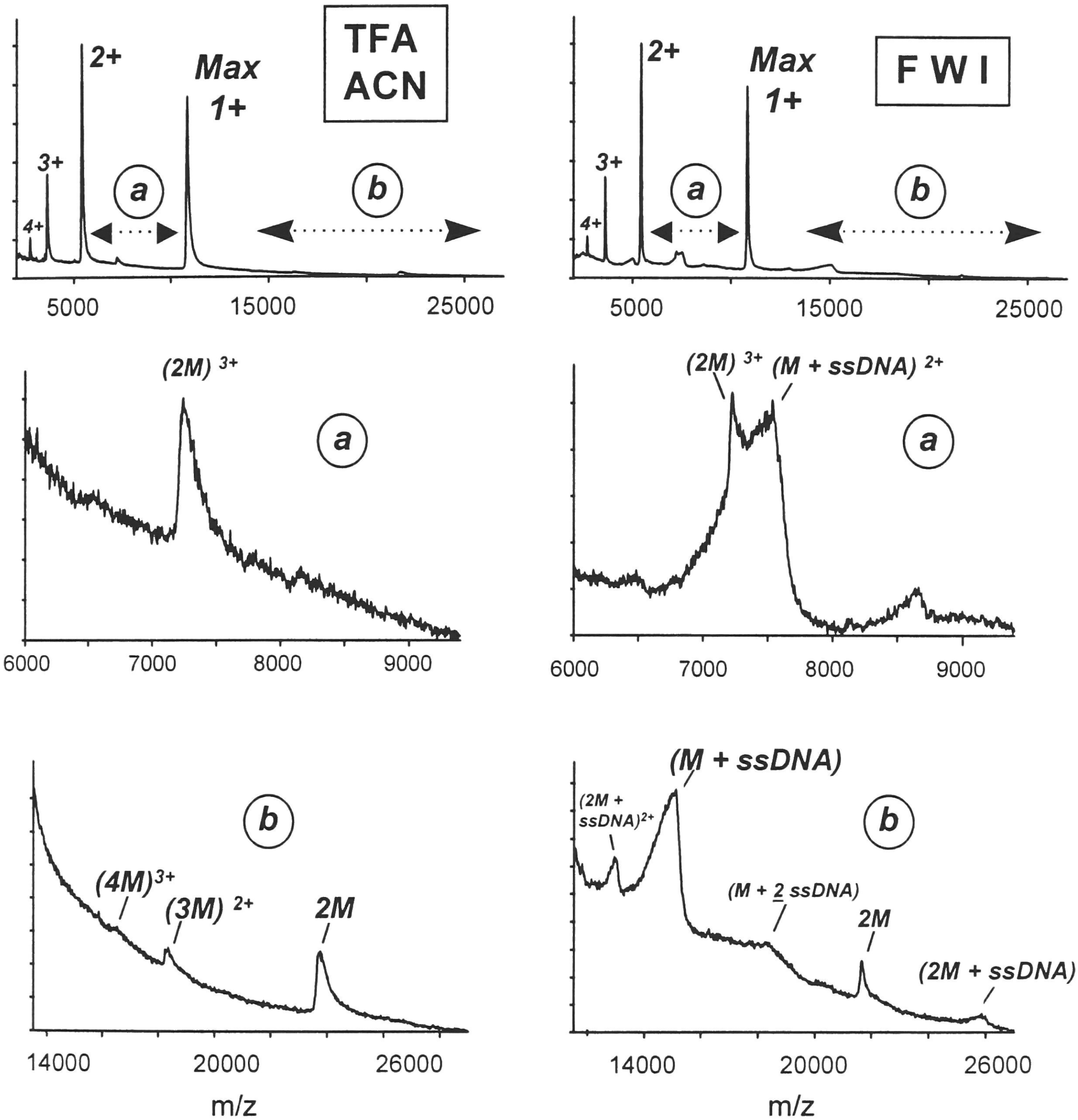


Figure 2.8 *Protein-DNA adduction by MALDI-MS.*

MALDI-MS of b/HLH/Z Max and a 14-mer DNA oligonucleotide using 4HCCA as the matrix. Spectra obtained from two different matrix solutions are shown: **LEFT COLUMN:** 0.1% TFA/ACN (2:1 v/v) and **RIGHT COLUMN:** formic acid/water/isopropanol (FWI) (1:3:2 v/v/v). *TOP* panels show the complete mass spectra from the two matrix solutions. Two regions in each spectrum at the *TOP* (labeled **a** and **b**) are shown expanded in the *MIDDLE* and *BOTTOM* panels. Max monomer is labeled M, oligomers are 2-4M and all peaks are singly charge unless noted otherwise. The Max-DNA adduct peaks are absent in TFA/ACN (LEFT) spectrum but strongly apparent in FWI spectrum (RIGHT).

peaks show a steep slope while the low mass side show significant 'tailing'. The low-mass side of protein peaks normally show a sharp slope. The tailing may be seen in Figure 2.8 for the FWI preparation may reflect gas-phase decomposition of the DNA in the adduct ion occurring just after ion formation following laser desorption but before emerging from the acceleration region of the ion source. Facile DNA fragmentation in the gas phase has been shown to occur in the linear TOF instrument, particularly at the guanosine residues of DNA oligos. (Schneider and Chait, 1993) for negative ion MALDI using cinnamic acid derivative matrices. The four guanosine residues in the 14-mer oligo. used for Figure 2.8 (see sequence above) could account for extensive gas-phase fragmentation. If the guanosine fragmentation hypothesis is correct, my results suggest that the gas-phase protein-DNA interaction is rather strong. However, it remains unclear why the Max-DNA adducts appear when using FWI but not with TFA/ACN (Figure 2.8). From a practical perspective, the protein-DNA adduct peak 'signature' was useful for the TBP-TFIIB-DNA ternary complex crystallization project (Nikolov et al., 1995). MALDI-MS quickly determined candidate crystals that contained all three components of the complex, information that was highly valuable in expediting completion of the project.

2.4.3 Effects of Redissolving the Matrix-Analyte Crystal

A final example of the benefits of using 4HCCA/FWI matrix solution is shown in Figure 2.9. The MALDI-MS of an endoprotease Asp-N digest of b/HLH/Z Max is performed in three ways. The spectrum in Figure 2.9a was from

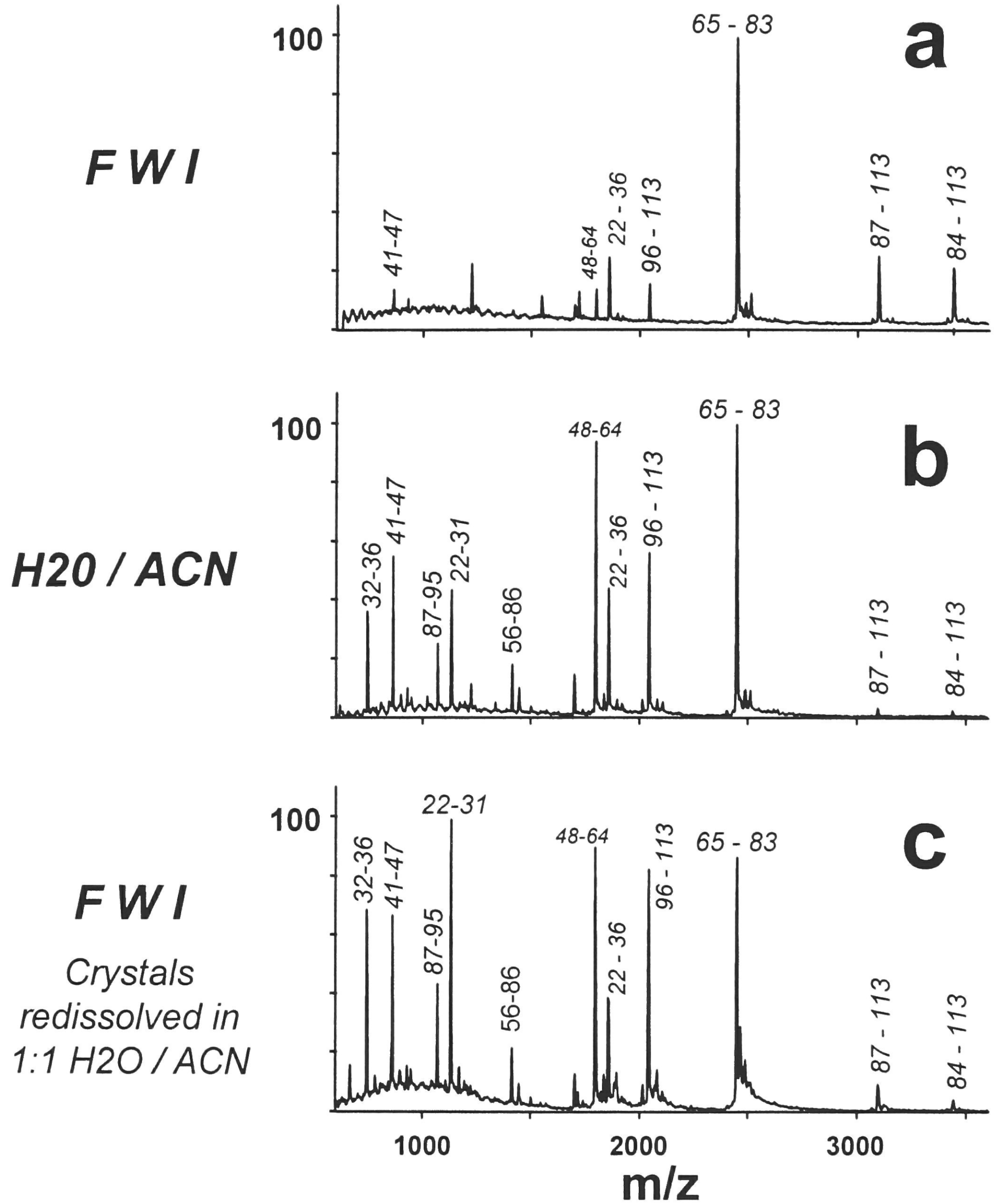


Figure 2.9 *Effects of redissolving the matrix crystals of a protein digest.*

MALDI-MS of a complete endo Asp-N digest of b/HLH/Z Max using 4HCCA matrix and dried-drop matrix crystallization. Spectrum of the digest in **(a)** was obtained using a matrix solution consisting of formic acid/water/isopropanol (1:3:2 by volume). Note the low-mass discrimination. In **(b)** the matrix solution was water/ACN (2:1 v/v). More low mass peaks appear. In **(c)** the matrix crystals used to obtain the spectrum in **(a)** were redissolved in ~1 μL of water/ACN (1:1 v/v) and allowed to recystallize *directly on the probe*. The AspN digest peaks of Max are labeled according to their sequence.

4HCCA/FWI and in Figure 2.9b from a 4HCCA-H_2O/ACN matrix solution using the dried-drop method of matrix crystallization. The pH-dependent discrimination effects (discussed previously) between the low and high mass peptides are clearly apparent —FWI favors high mass whereas H_2O/ACN favors low mass peptides. In Figure 2.9c, the spectrum was from the same dried-drop matrix crystals used to obtain Figure 2.9a *except* that the matrix crystals were redissolved and recrystallized directly on the probe tip with 1 μL of water/acetonitrile (1:1 v/v). Low mass peptides are now clearly observed in Figure 2.9c.

2.5 Conclusions

A major strength of MALDI mass spectrometry is its facility for analyzing complex mixtures of peptides and proteins. I have demonstrated that the choice of the matrix solvent system and the rate of matrix crystal growth add a certain 'chromatographic-like' dimension to the sample-matrix preparation, and that to take full advantage of the power of MALDI, careful attention must be paid to the sample-matrix preparation procedures. Having extensively investigated these effects, I suggest here some general guidelines for matrix solution preparation (using the matrix 4HCCA and the dried-drop method) (1) The analysis of proteins and high mass peptides (> 3 kDa) is best performed using matrix solutions that include formic acid and have a pH < 1.8. (2) The analysis of small peptides (< 3 kDa) is best carried out using matrix solutions that do not have

added acids. (3) For a complex mixture of peptides and proteins, a single matrix solution (pH ~2) can be used (such as TFA/ACN). However, to obtain a more complete coverage of components in the mixture, it is advisable to run the mixture separately in the two matrix solutions suggested in (1) and (2) above. (4) The presence of formic acid (such as in FWI) lends particularly useful features for analyzing protein samples with oligonucleotides or high levels of involatile additives. (5) Overall, optimizing the MALDI-MS signal is somewhat sample-dependent, requiring a fine tuning of the matrix solvent composition and matrix crystallization protocols.

Chapter 3 — Mass Spectrometry of Whole Proteins Eluted from SDS-PAGE Gels

3.1 Introduction

Gel electrophoresis plays a significant role in biological research (Hames and Rickwood, 1990). The notable strength of gel electrophoresis is its ability to provide separation of mixtures of large biomolecules at high resolving powers in a relatively timely and inexpensive manner. Isoelectric focusing (IEF) and sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels, used individually (1-D gels) or in combination (2-D gels), are powerful electrophoretic methods that provide microscale analyses of complex mixtures of proteins. However, the information concerning a particular protein 'spot' on a gel is limited. IEF reveals the relative acidity or basicity (i.e., pI) of a protein whereas SDS-PAGE provides an 'apparent' molecular mass of the protein. Mass accuracy by SDS-PAGE, however, is no better than 5% and more often than not is grossly lower.

Matrix-assisted laser desorption/ionization (MALDI) (MS) and electrospray ionization (ESI) MS can provide rapid, sensitive, and accurate analysis of proteins. Mass accuracies are usually better than 0.1 % (Beavis and Chait, 1990b; Chait and Kent, 1992), representing at least a 50-fold improvement over the mass accuracy of SDS-PAGE. Thus, the coupling of SDS-PAGE (protein separation) and mass spectrometry (mass accuracy) is a potentially powerful

analytical combination. For example, the interfacing of one- and two-dimensional polyacrylamide gel EF with MS is experiencing a surge of activity in the areas of protein identification, using database searching algorithms (Qin et al., 1996b; Wilm et al., 1996b; Yates III et al., 1996b), and in characterization of posttranslational modifications of proteins (reviewed in Burlingame et al., 1996). These applications employ various types of gel separation and mass spectrometric techniques that depend on *in-situ* chemical or proteolytic digestion of the protein of interest followed by elution of the resulting peptide fragments for MS analysis.

Techniques are also being developed for the MS analysis of *whole* proteins from gels. However, retrieving large polypeptides and proteins from gels is inherently more difficult than recovering small and moderately sized peptides. In addition, with the use of SDS-PAGE, the presence of SDS detergent and protein visualization dye (e.g., Coomassie Brilliant Blue) severely compromises mass spectrometric response to proteins (S. L. Cohen, unpublished results; Strupat et al., 1994; Loo et al., 1996a). A variety of methods have been introduced to overcome these limitations and permit MS analysis of whole proteins from gels. One strategy uses electroblotting and electroelution techniques. For example, plasma desorption-MS (Klarskov and Roepstorff, 1993), UV-MALDI-MS (Vestling and Fenselau, 1994) or IR-MALDI-MS (Strupat et al., 1994) can be directly run on proteins that have been electroblotted onto polymeric membranes. UV-MALDI-MS has also been applied to electroblotted proteins recovered from a dissolvable nitrocellulose membrane

(Liang et al., 1996), and ESI-MS analysis of proteins electroeluted from SDS-PAGE gels (le Maire et al, 1993; Nakayama et al, 1996; Schuhmacher et al, 1996). A more recent approach applies UV-MALDI-MS to the analysis of proteins directly from ultrathin polyacrylamide gels (Loo et al., 1996a).

I developed an alternative strategy to mass spectrometrically analyze whole proteins from gels, a strategy that evolved principally from the MADLI matrix studies described in chapter 2. The method consists of three components: the use of conventional SDS-PAGE gels, reversible negative staining procedures, and passive elution of proteins from the gel followed by MALDI- or ESI-MS analysis. For high sensitivity applications, the method is restricted to MALDI-MS and involves the use of slow matrix crystallization (Xiang and Beavis, 1993). This strategy offers the ability to analyze soluble proteins from gels down to the low picomole level (1-10 pmol), providing a powerful means for characterizing endogenous forms of mature proteins separated by SDS-PAGE (Cohen and Chait, 1997).

3.2 Materials

3.2.1 Proteins

Horse skeletal muscle myoglobin (+heme = 17,568 Da; 95-100% pure; Sigma) was accurately weighed on a microbalance (Cahn 4400, Orion Instruments, Boston, MA) and dissolved in aqueous 0.1 % TFA (pH 2) (Pierce, Rockford, IL) to make a 50 μM protein solution. Recombinant mouse leptin was

provided by Dr. Jeffrey Friedman (Rockefeller University). The leptin consisted of wild-type residues 22-167 with an amino-terminal methionine (16,134 Da) (Zhang et al., 1994). The presence of a single intramolecular disulfide bond in mouse leptin was determined by performing in-solution CNBr digestion as well as enzymatic proteolysis (Halaas et al., 1995). The leptin concentration was 1 mg/mL (62 μM; prepared in a standard phosphate-buffered saline solution, pH 7.4) and was determined spectrophotometrically using an extinction coefficient of 0.16 $(mg\text{-}cm/mL)^{-1}$ (personal communication, Ketan Gajiwala, Rockefeller University). For qualitative studies, a broad range protein mass marker mixture (0.1-0.2 mg/mL each protein) was purchased from New England BioLabs (Beverly, MA) and bovine transferrin and bovine carbonic anhydrase II were obtained from Sigma.

3.3 Methods

3.3.1 Preparation of proteins for SDS-PAGE

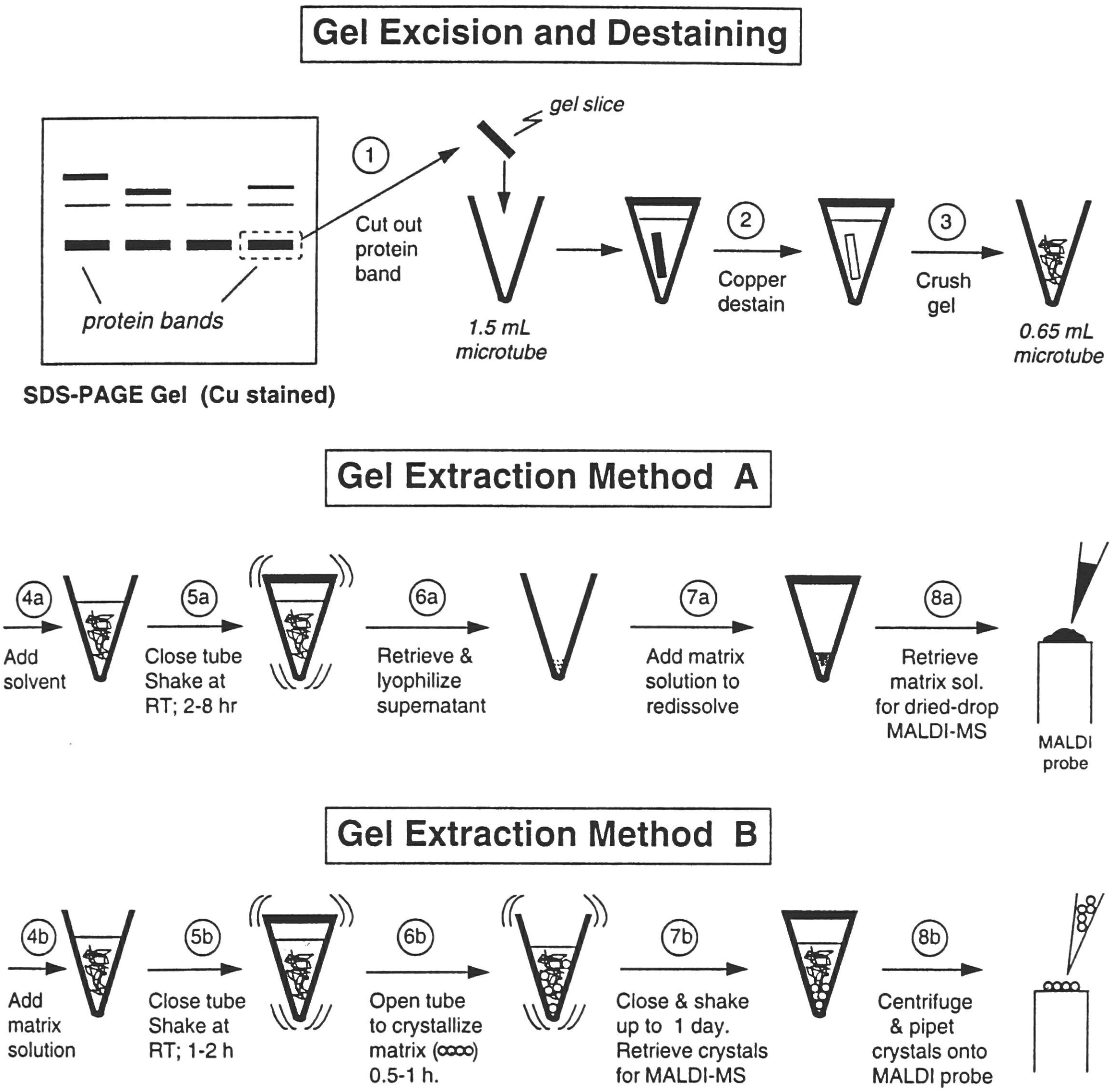
Proteins were aliquoted into 0.65 mL micro-centrifuge tubes and ammonium phosphate or acetate buffer was added to bring the volume to 10 μL. An equal amount (10 μL) of Tris-Glycine sample buffer (2X concentrated) (Novex, San Diego, CA) was added to each tube. The sample buffer contained bromophenol blue to visualize the electrophoresis dye front. The protein/sample buffer solutions were vortexed and centrifuged prior to loading onto the gel.

3.3.2 SDS-PAGE

Gel electrophoresis was performed according to Laemmli (Laemmli, 1970). Gloves were worn at all times when handling the gels. Precast 14 % Tris-Glycine gels (1 mm thick, 10 wells) (Novex) were used throughout, although gels with other acrylamide concentrations, thickness and number of wells can be used, depending on the desired molecular mass range, sample volumes, and electrophoresis equipment available. However, only Tris-Glycine gels are appropriate for the present application because the destaining process is based on the Tris buffer system (Lee et al., 1987). Electrophoresis was carried out on an Xcell II Mini-Cell device (Novex) at room temperature. The running buffer was prepared from a 10X Tris-Glycine SDS solution (Novex). Following assembly of the Mini-Cell with a fresh gel and addition of the running buffer, the protein/sample-running buffer solutions (20 μL; see above) were loaded into the sample wells with the aid of gel loading tips (Novex). Electrophoresis was carried out at 120 volts and an initial current of ~32 mA using a PowerPac 1000 power supply (Bio-Rad; Hercules, CA), and was terminated when the dye front reached close to the bottom of the gel. Typical run times were 1.5-2 h.

3.3.3 Gel Staining

Immediately following electrophoresis, the gel can be rinsed in water, although I recommend against rinsing when handling low molecular mass proteins (<20 kDa). Protein bands were visualized by soaking the entire gel in



<u>Figure 3.1</u> *Eluting whole proteins from SDS-PAGE gels for mass spectrometric analysis.*

The strategy is comprised of two parts. The first consists of gel excising, destaining, and crushing (TOP ROW). The second part consists of gel extraction, involving either passive elution of the gel using an acidic solution **(MIDDLE ROW, Gel Extraction Method A)** or passive elution combined with a slow matrix crystallization using a MALDI matrix solution **(BOTTOM ROW, Gel Extraction Method B).** See text for a full description.

45 μL of copper staining solution (supplied 10X from Bio-Rad) for five minutes with gentle rocking (Lee et al., 1987). Alternatively, zinc staining (Dzandu et al., 1988) or imidazole-zinc staining (Ortiz et al., 1992) protocols can be used. After staining, the gels were thoroughly rinsed in deionized water. Because copper and zinc staining are negative stains, protein bands appear translucent and are best viewed against a black background. Bands with as little as 17 ng (1 pmol) of myoglobin or recombinant leptin were visualized by copper staining.

3.3.4 Gel Excising and Destaining (Figure 3.1)

Destaining was performed on individual gel slices containing a single protein band as depicted in Figure 3.1 and each step is described as follows.

Step 1: Protein bands were excised from the stained gel with a clean, sharp straight-edge razor. The gel slices were excised as close as possible to the boundaries of the protein band. The resulting blue-tinted gel slices typically had the dimensions 1-1.5 mm x 6-7 mm and were best handled with flat-end tweezers. Protein extraction should be performed immediately after destaining. If a delay is anticipated, it is best to leave the gel slices stained and stored moist (but not soaking) at 4°C in a sealed microtube until needed.

Step 2: The destaining process completely removes the visualization stain as well as the SDS detergent (Lee et al., 1987). Each gel slice was treated in a

separate 1.5 mL micro-centrifuge tube using standard copper-destaining protocols (Bio-Rad). Zinc-stained gels can also be destained using the same protocols. Destaining was performed by a three-step soaking process using a copper Tris-Glycine destain solution (Bio-Rad). The first soaking of the gel is carried out in a solution consisting of 100 μL destain buffer and 900 μL water. The tube is sealed and vortexed (Tomy MT-360 36-sample microtube mixer; Tomy Tech, Palo Alto, CA) for 5 minutes at room temperature. During the first destaining period the solution turns pale translucent blue (for copper-stained gels) and may become slightly foamy. The gel will retain a faint blue tint. The second soaking is performed in fresh solution of 100 μL destain and 900 μL water with 10 minutes of vortexing. At the end of the second soaking, the destain solution and gel will be nearly colorless. The final soaking is performed in a solution of 50 μL destain and 950 μL water with 5 minutes of vortexing. At the end of the third soaking, the destain solution as well as the gel are clear. The destained gel slice is rinsed in deionized water and is now ready for extraction or in-gel digestion (see below).

Step 3: Prior to protein extraction, the gel slice is crushed. Before crushing, the destained gel is gently blotted free of any excess water clinging to the gel with a lint-free tissue, a step that facilitates crushing of the gel. With the aid of tweezers, the gel slice is placed into a 0.65 mL microtube (PGC Scientifics, Gaithersburg, MD). The microtube should be of high quality, free of plasticizers

and contaminants that can compromise high sensitivity MS measurements. The gel slice is manually crushed at the bottom of the microtube for a few seconds with a sharp-pointed dental tool. Mechanical homogenization of the gel can be used instead of the manual crushing, but I recommend against it since the resulting small gel pieces tend to clog pipettor tips. Others have also noted that homogenization introduces 'poisons' that severely impair mass spectrometric protein analysis (personal communication, Bill Jordan). At this point, either of two methods described below (Extraction Methods A or B) can be used to elute the protein from the gel.

3.3.5 Extraction Method A (Figure 3.1, Steps 4a-8a)

This method of extraction elutes protein from the gel directly into an aqueous solution and is best suited for gel loadings exceeding 25 pmol. Solution compositions that I have examined for protein extraction include formic acid/water/2-propanol (1:3:2 v/v/v) (FWI); formic acid/water (1:5 v/v); water/2-propanol (2:1 v/v); and pure water. Optimal extraction efficiency occurred with the FWI combination, whereas no extraction was observed with pure water. The relative extraction efficiencies are based on a qualitative assessment of the MALDI-MS signal-to-noise ratios. Other acids and water-miscible organic solvents (e.g., 0.1 % TFA and acetonitrile) can be used for extraction.

Step 4a: Sufficient extraction solution (~30-40 μL) is added to the crushed gel to completely cover the gel pieces.

Step 5a: The tube is closed and vigorously shaken or vortexed at room temperature typically for a period of 2-8 hr. Although not essential for protein elution, shaking/vortexing facilitates protein extraction. Generally, the greater the protein loading or the smaller the protein, the shorter the vortexing time required for extraction. The gel extraction times for the spectra in Figure 3.2 is 4 hr whereby for Figure 3.3, it is 12 hr.

Step 6a: After the vortexing period, the microtube is centrifuged and the supernatant is retrieved with a micropipettor equipped with a 1-10 μL tip, avoiding taking up any of the gel pieces. The crushed gel is washed once with an equal volume of fresh extraction solution that is combined with the supernatant.

Steps 7a and 8a: If the protein is to be analyzed by MALDI-MS, the extraction solution is lyophilized and a MALDI matrix solution (2 μL for Figure 3.2 and 3.3) is added to redissolve and retrieve the protein for MS analysis using the dried-drop method of matrix crystallization (see chapter 2). The matrix solution is FWI saturated with 4-hydroxy-α-cyano-cinnamic acid (4HCCA) (Cohen and Chait,

1996). A calibrant protein, if required, can also be added to the solution at this point.

If the protein is to be analyzed by ESI-MS, the extraction solution is injected into a protein desalting cartridge (Michrom BioResources, Inc., Auburn, CA) that is mounted to the ESI capillary inlet. (The desalting cartridge is discussed in detail in chapter 2, Methods). The loaded cartridge is flushed several times with a 2 % acetic acid wash solution. Following the washing, the protein is eluted from the cartridge into the electrospray source with a solution consisting of 7% acetonitrile/2.5% acetic acid at 6 μL/min. ESI-MS was performed on a TSQ-700 triple quadrupole (Finnigan Corp., San Jose, CA) using standard ESI conditions for analyzing proteins (Mirza et al., 1995) (see chapter 5 "Methods" section for ESI conditions).

3.3.6 Extraction Method B (Figure 3.1, Steps 4b-8b)

This method of extraction elutes protein from gels directly into a MALDI-MS matrix solution and is best suited for gel loadings below 25 pmol of protein or for proteins that fail to passively elute using Extraction Method A (described above). In Method B the matrix is allowed to undergo slow crystallization (Xiang and Beavis, 1993) during the protein extraction period. The matrix used throughout was 4HCCA, although sinapinic acid can also be used. The moderately low solubility of these cinnamic acid derivatives in the matrix solution is an important characteristic for efficient slow crystallization (Xiang and Beavis,

1993). The most frequently used matrix solution consisted of a saturated solution of 4HCCA in FWI. Other solvent systems were also examined (e.g., 0.1% TFA:acetonitrile, 2:1 v/v) (see Table 2.1 for suggestions). Gel extraction with the matrix solution (Figure 3.1) is as follows:

Step 4b: After crushing the destained gel (see above), enough matrix solution (~20-50 μL) is added to cover the gel pieces. A calibrant protein, if necessary, can be added at this point to the solution.

Step 5b: The tube is closed and vigorously shaken or vortexed at room temperature for 1-2 h.

Step 6b: Matrix crystallization is induced by leaving open the top of the microtube (0.5-1 h) with vortexing. The matrix solution will become slightly cloudy with a suspension of 4HCCA crystals. An aliquot (1 μL) of the milky supernatant can be immediately retrieved and deposited onto the MALDI probe for MS analysis while the remaining sample is closed and left vortexing for additional extraction. Avoid taking up any gel pieces while pipetting and avoid crushing the matrix crystals once they are on the probe.

Steps 7b and 8b: If the MALDI-MS response is weak from matrix crystals that were immediately retrieved following their formation, the signal may improve from

crystals that have undergone a longer period (~4-24 h) of vortexing in the gel/matrix suspension. During extended vortexing periods, the matrix crystals may precipitate to the bottom as a fluffy yellow 'pellet', cling to the sides of the microtube, and/or remain suspended in solution. Crystals can be retrieved from any of these locations within the tube with a micropipettor for MALDI-MS analysis. Also, during the vortexing period volume losses of the solution due to evaporation and aliquot sampling will occur. To offset the volume losses, small amounts (5-10 μL) of *matrix-free* extraction solution (i.e., formic acid-water-IPA 1:3:2, v/v/v) or just acidified water (i.e., formic acid/water, 1:5 v/v) can be added to keep the gel pieces completely submerged in solution. For the results presented in the thesis, vortexing times were: 4 h (Figure 3.9), 6 h (Figure 3.6), 8 h (Figure 3.8) to 12-19 h (Figure 3.7).

3.3.7 In-Gel Digests

Gel slices for digestion were destained and crushed as described above (steps 1-3, Figure 3.1). **Trypsin digest:** To the crushed gel was added 60 μL of 10 mM ammonium acetate buffer (pH 7) (enough to at least completely submerge the gel pieces) and 250 ng of trypsin. The tube was closed and vortexed 4 h at room temperature. Two approaches (I or II) were used to map the digest. In the first method (I), the gel and supernatant *together* were lyophilized to dryness after which 40 μL of matrix solution (4HCCA/FWI) was added to rehydrate the gel pieces. Gel extraction Method B was followed to

obtain matrix crystals for MALDI-MS analysis. Alternatively, in the second method (II) the original digest supernatant was retrieved and retained. The gel pieces were extracted once with 40 μL of 0.1% TFA/ACN (1:1 v/v) and once with 40 μL of pure acetonitrile. The two extracts were combined with the digest supernatant and lyophilized to dryness. The lyophilized digest was redissolved with matrix solution (2-3 μL) followed by the dried-drop method to obtain matrix crystals. The lyophilized digest can also be redissolved in 5-10 μL of 0.1% TFA/ACN (1:1 v/v), 1 μL of which was combined with an equal volume of matrix solution for dried-drop matrix crystallization (Qin, 1996). **Cyanogen Bromide Digest:** To the crushed gel was added 50 μL of 0.1 N HCl and a small crystal of CNBr. The tube was left shaking in the dark at room temperature for 12 h. Following the digestion, the supernatant was retrieved and the gel pieces retained. The supernatant was lyophilized and redissolved in 3 μL of 4HCCA/FWI matrix solution for dried-drop MALDI-MS. To the retained gel pieces was added 4HCCA/FWI matrix solution. Gel extraction Method B was followed to obtain matrix crystals for MALDI-MS analysis.

3.3.8 MALDI-MS

The dried-drop method of matrix crystallization (see chapter 2) (Karas and Hillenkamp, 1988) were used to obtain the spectra in Figures 3.2 and 3.3a Slow crystallization (Xiang and Beavis, 1993) (as described in chapter 2, 'Experimental') was used to obtain the spectrum in Figure 3.3b.

3.4 Results

3.4.1 Eluting Whole Proteins from SDS-PAGE Gels.

My strategy for eluting whole proteins from SDS-PAGE gels is presented in Figure 3.1. Full details of the procedures are given in the Methods section above. The approach starts with an SDS-PAGE gel on which proteins have been resolved. Protein bands are visualized with a reversible negative stain such as copper (Lee et al., 1987) or zinc (Dzandu et al., 1988; Ortiz et al., 1992) staining. The bands are excised from the gel and destained. The destaining procedure completely removes the stain and SDS detergent. The gel is then crushed and the protein extracted by one of the protocols I term Extraction Methods A and B. I have found that for gel loadings $\geq$ 25 pmol of protein, the protein passively elutes directly into a simple extraction solution (Extraction Method A). For gel loadings < 25 pmol of protein, the inclusion of a MALDI matrix in the extraction solution is important to ensure high sensitivity (Extraction Method B).

3.4.2 Extraction Method A: Elution of Whole Proteins from Gels for MS Analysis

This method involves a passive elution of the protein into solution. Following extraction, the solution is retrieved from the gel pieces and lyophilized. Saturated matrix solution is added to the lyophilized protein and MALDI-MS is performed on the matrix crystals prepared by the dried-drop method. Figure 3.2

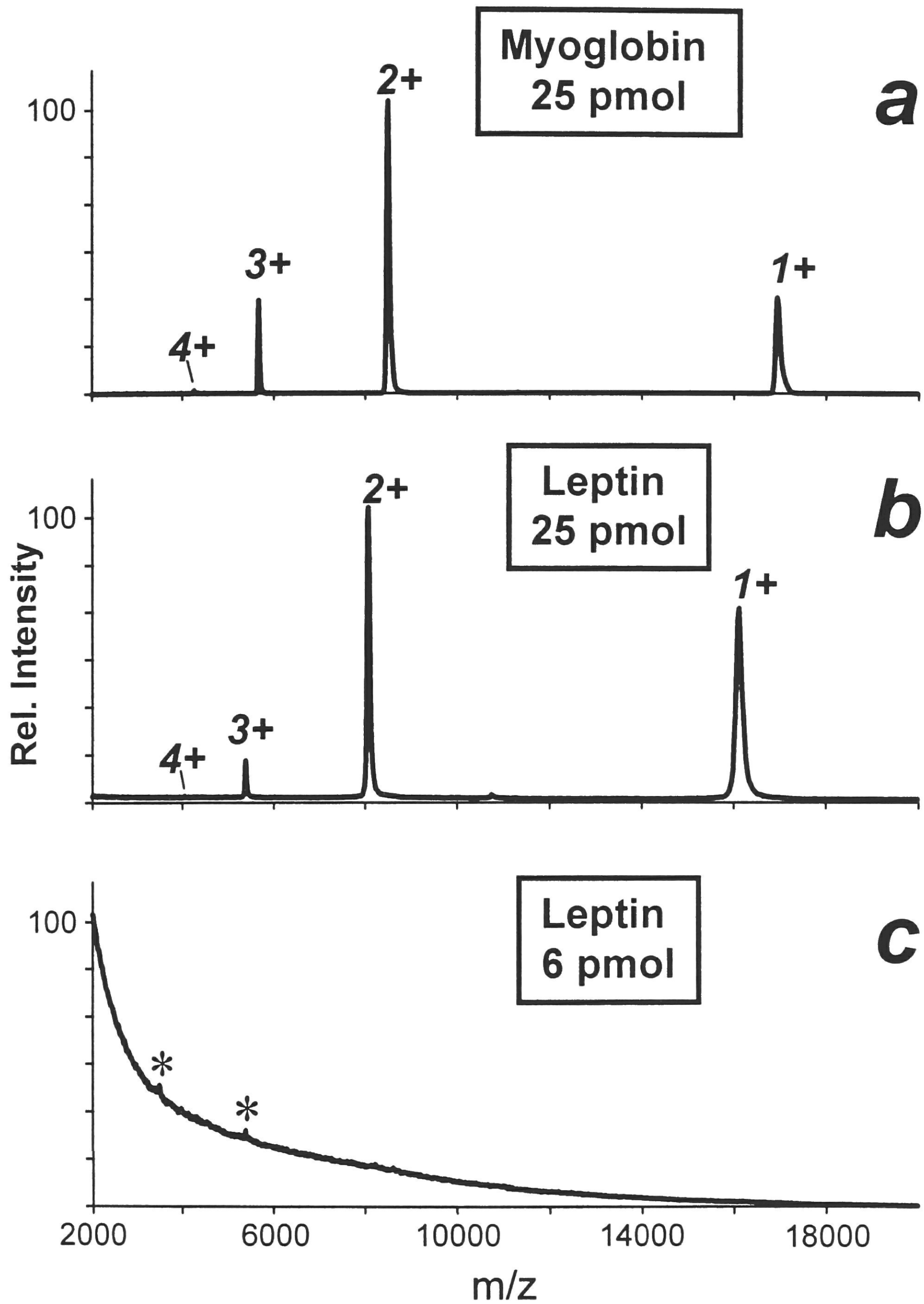


Figure 3.2 *MALDI-MS of proteins eluted from SDS-PAGE gels using gel Extraction Method A.*

The extraction solution was formic acid/water/2-propanol (1:3:2 v/v/v). MALDI spectra of three gel loadings are shown: **(a)** 25 pmol equine myoglobin (peaks are apomyoglobin, M_r = 16,951); **(b)** 25 pmol, and **(c)** 6 pmol of recombinant mouse leptin (M_r = 16,134). The small peaks labeled with an asterisk are low-level impurities.

shows MALDI mass spectra of equine myoglobin and recombinant mouse leptin extracted by Method A. The spectra in Figures 3.2a,b, obtained from the extraction of gels loaded with 25 pmol of the two proteins, show excellent signal-to-noise ratios, suggesting favorable protein extraction by Method A. However, because accurate quantitation by MALDI-MS is difficult, I did not estimate the recovery yields. MALDI mass spectra were also obtained for 50 pmol gel loadings of bovine carbonic anhydrase II (29 kDa) and bovine transferrin (78 kDa) using Extraction Method A and are shown in Figure 3.3. As an alternative to the MALDI-MS technique, ESI-MS can be used to analyze the extracted protein. Good quality ESI mass spectra from the extraction of 50 pmol gel loadings of recombinant leptin and myoglobin here obtained. The ESI-MS spectrum of recombinant leptin is shown in Figure 3.4

The utility of Extraction Method A, however, is limited to gel loadings $\geq$ 25 pmol. As observed in Figure 3.2c, the MALDI-MS signal significantly drops for gel loadings approaching the 6 pmol level of leptin. A similar fall-off in sensitivity is also observed with myoglobin (data not shown). I attributed the fall-off in signal to a reduction in protein extraction recovery yields. To offset the poor recovery yields from gels with low picomole loadings, I slightly modified the passive extraction method. The modification makes use of MALDI-MS analysis and exploits the power of slow matrix crystallization as described below.

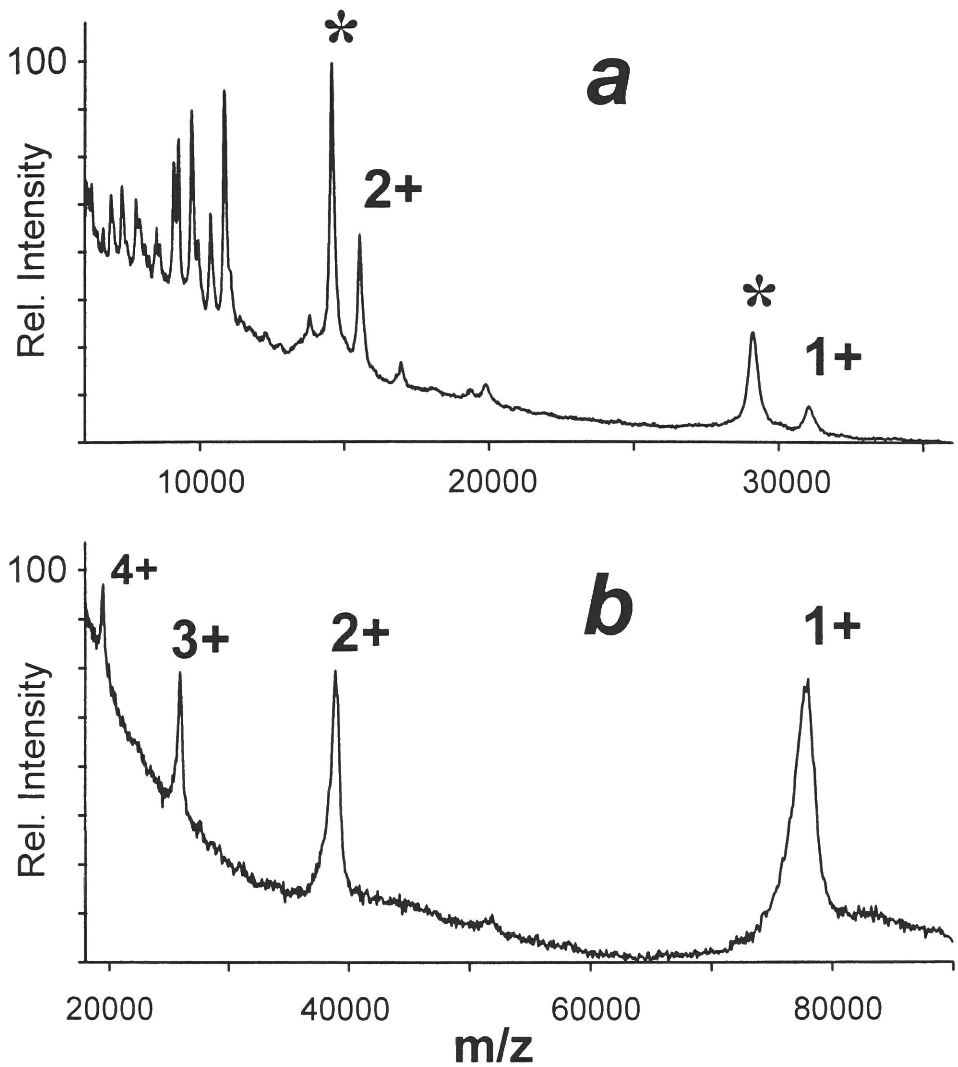


Figure 3.3 *MALDI-MS of high mass proteins extracted from SDS-PAGE gels by Extraction Method A.*

(a) Bovine carbonic anhydrase II, BCA II, (29 kDa) and ***(b)*** bovine transferrin (78 kDa). 50 pmol gel loadings. The BCA II spectrum is dominated by the hydrolysis product (*) of acid cleavage at an Asp/Pro bond. The acid hydrolysis occurred during gel extraction with the strongly acidic extraction solution.

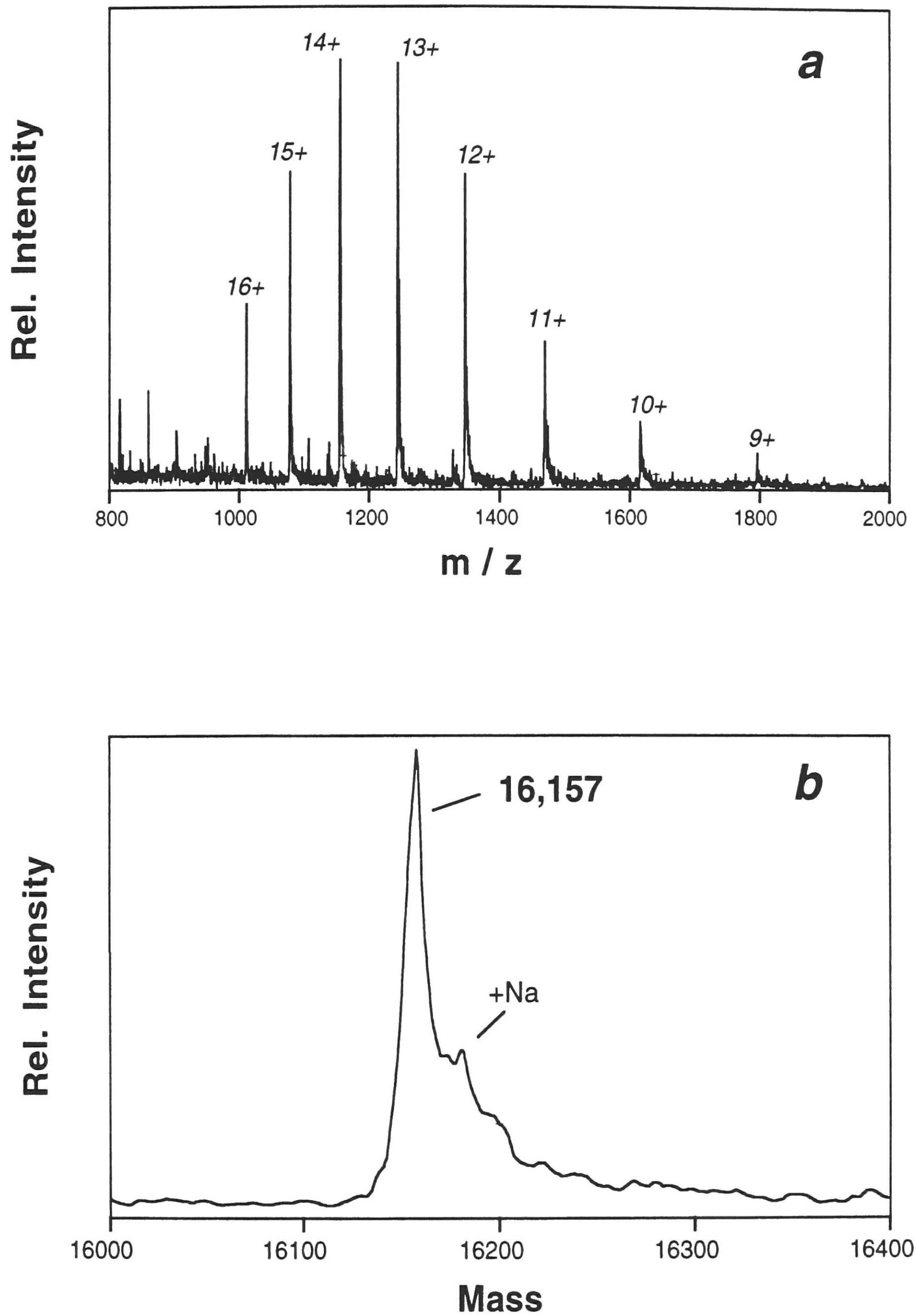


Figure 3.4 *ESI-MS of recombinant human leptin eluted from SDS-PAGE gel by Extraction Method A.*

(a) Raw data labeled with multiple protonation charge states, and ***(b)*** deconvoluted data of *(a)* labeled with the measured mass of the leptin. Expected mass is 16,155 Da. The sodium adduct peak is marked.

3.4.3 *Slow Crystallization of the MALDI Matrix Greatly Enhances MALDI-MS Sensitivity.*

MALDI-MS depends on growing analyte-doped matrix crystals. Several methods of matrix crystallization are available (Beavis and Chait, 1996), the most widely used of which is the dried-drop method (Karas and Hillenkamp, 1988) (described in chapter 2). In the dried-drop method, proteins are dissolved in a saturated MALDI matrix solution. Ideally, the protein concentration in the matrix solution should be > 0.1 μM to ensure a good MALDI-MS response. Below 0.1 μM of protein in the matrix solution, there is a non-linear drop-off in the MALDI-MS response (Figure 3.5a). Figure 3.5a shows the MALDI mass spectrum of recombinant mouse leptin using the dried-drop method (1 μL of matrix solution containing 0.01 μM protein). No signal from the protein is observed. In striking contrast, when slow crystallization (Xiang and Beavis, 1993) is used to produce matrix crystals (30 μL of matrix solution containing 0.01 μM protein), I observe a moderately strong MS response (Figure 35b). A similar pattern of behavior is seen from a matrix solution containing 0.01μM equine myoglobin (data not shown). Slow matrix crystallization thus provides a convenient means to concentrate protein from the solution into the matrix, a property that can be exploited to enhance MALDI-MS sensitivity. (One example of the MALDI signal enhancement effect by slow crystallization has already been shown in Figure 2.7b-c).

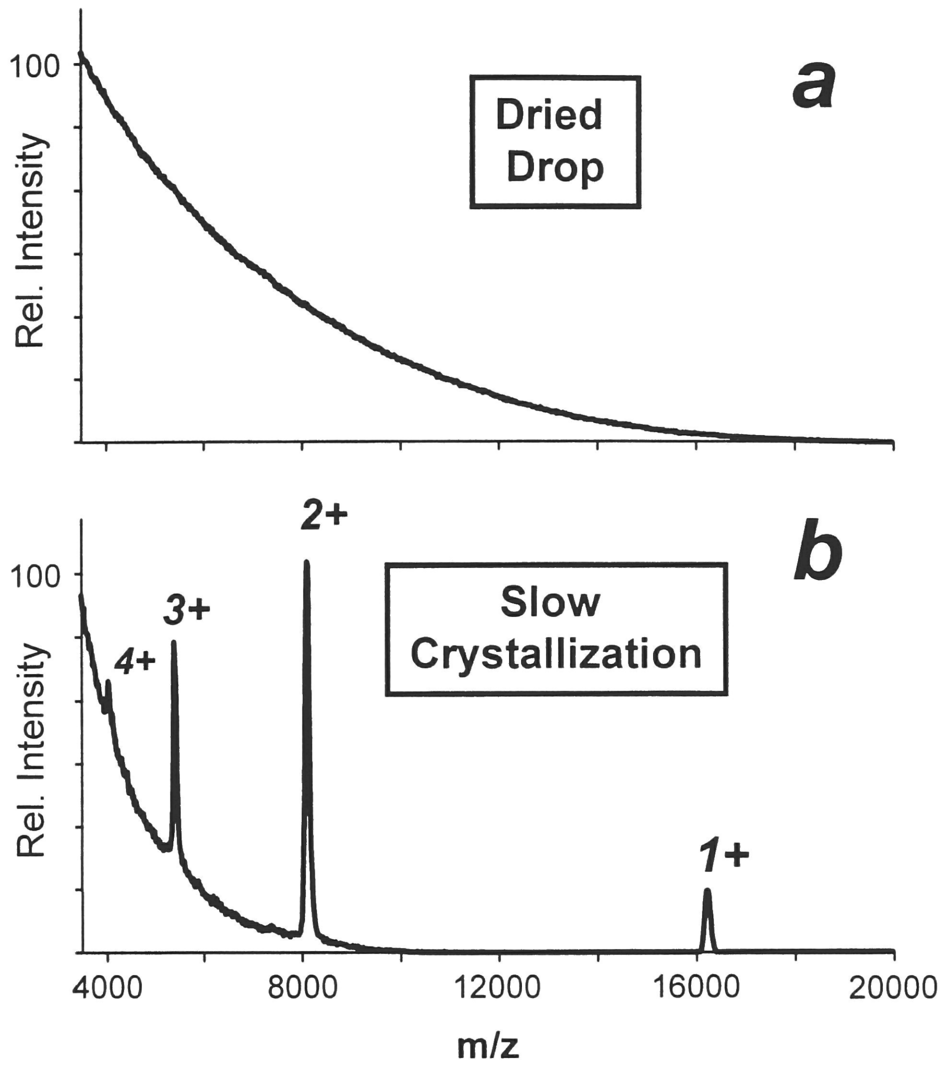


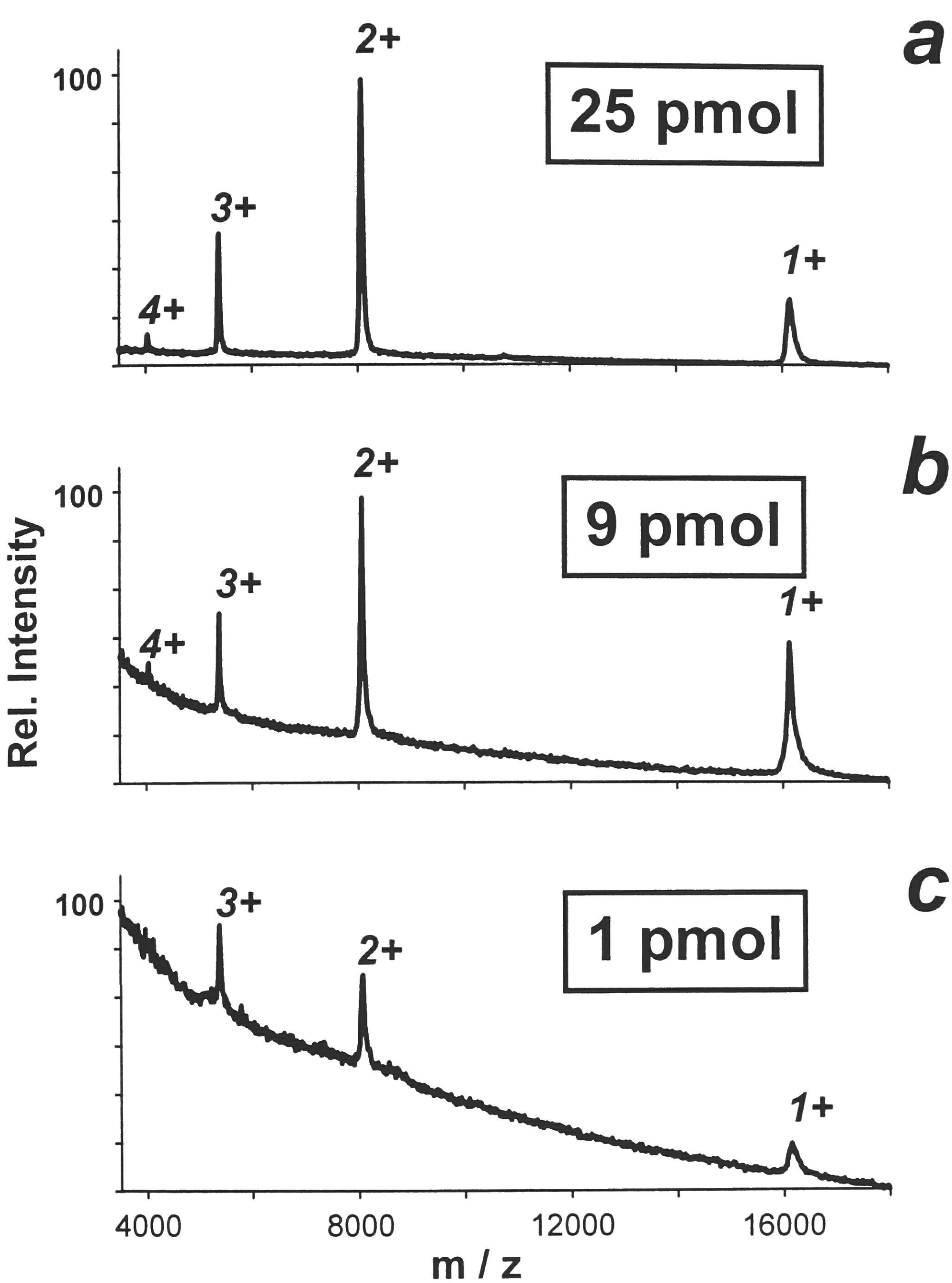
Figure 3.5 *Slow crystallization greatly improves the MALDI-MS response of proteins.*

MALDI-MS of recombinant mouse leptin at a very low concentration in the matrix solution (0.01 μM). Using **(a)** conventional dried-drop method, and **(b)** slow crystallization method of MALDI-MS analysis.

3.4.4 Extraction Method B: Eluting Whole Proteins from Gels for High Sensitivity MALDI-MS Analysis

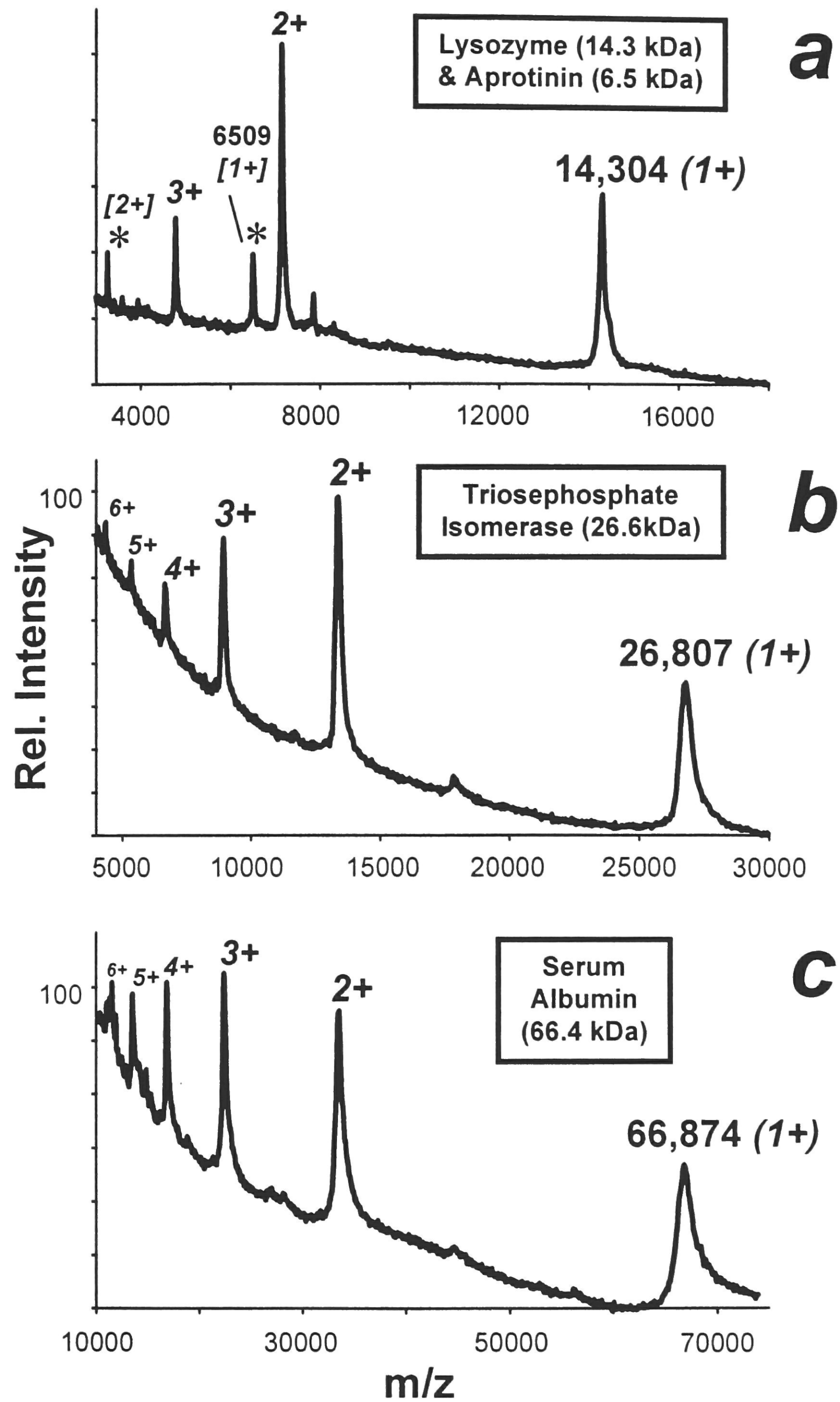
To achieve high sensitivity analysis of proteins originating from gels, I combined passive gel elution (Extraction Method A) with slow matrix crystallization. The result is a new approach to mass spectrometrically analyzing proteins from gels (Figure 3.1, Extraction Method B). In Method B the eluting solvent is a MALDI matrix solution. The matrix solution permits passive extraction of the gel along with a nearly simultaneous slow crystallization of the matrix. Under vortexing conditions, the matrix crystals remain finely dispersed in solution with the gel pieces. Thus, protein that elutes from the gel has a chance of co-crystallizing with the matrix. Figure 3.6 depicts MALDI mass spectra of recombinant mouse leptin using Extraction Method B. The spectra show the effects of decreasing gel loadings of leptin, from 25 pmol (Figure 3.6a) to 1 pmol (Figure 3.6c). Extraction Method B maintains good sensitivity even at 1 pmol gel loading of leptin, a level at which direct elution (Extraction Method A) completely failed to give any MALDI-MS signal for the protein (Figure 3.2c). (See Cohen and Chait, 1997, Fig. 4, for a photo of a Cu-stained gel with 25-1 pmol protein.)

The MALDI mass spectra of a selection of protein mass markers obtained using Extraction Method B is shown in Figure 3.7. Based on the marker set manufacturer's specification, the gel loadings were calculated at 200 ng for each protein. This represents 14 pmol of lysozyme (14.3 kDa; Figure 3.7a), 7 pmol of triosephosphate isomerase (26.6 kDa; Figure 3.7b) and 3 pmol of bovine serum



<u>Figure 3.6</u> *MALDI-MS of recombinant human leptin from copper-stained SDS-PAGE gels.*

MALDI-MS of recombinant mouse leptin (M_r=16,134) eluted from SDS-PAGE gels using Extraction Method B. The MALDI matrix solution used to extract the gel slices was formic acid/water/2-propanol (1:3:2 v/v/v) saturated with 4HCCA. The gel loadings represent **(a)** 25 pmol, **(b)** 9 pmol, and **(c)** 1 pmol of leptin.



<u>Figure 3.7</u> *MALDI-MS of molecular mass markers extracted from copper-stained SDS-PAGE gels.*

MALDI-MS of protein mass markers eluted from SDS-PAGE gels using Extraction Method B. The markers are **(a)** chicken egg white lysozyme; **(b)** triosephosphate isomerase; and **(c)** bovine serum albumin. The spectrum in (a) also shows peaks (labeled by an asterisk) corresponding to the marker protein aprotinin (6.5 kDa) that co-migrated and eluted along with lysozyme. Masses above the 1+ pks. are derived from external calibration.

albumin (66.4 kDa; Figure 3.7c). Figure 3.7a exhibits an additional set of peaks whose mass correlates to the mass marker aprotinin (6.5 kDa), indicating that aprotinin co-migrated with lysozyme on the SDS-PAGE gel.

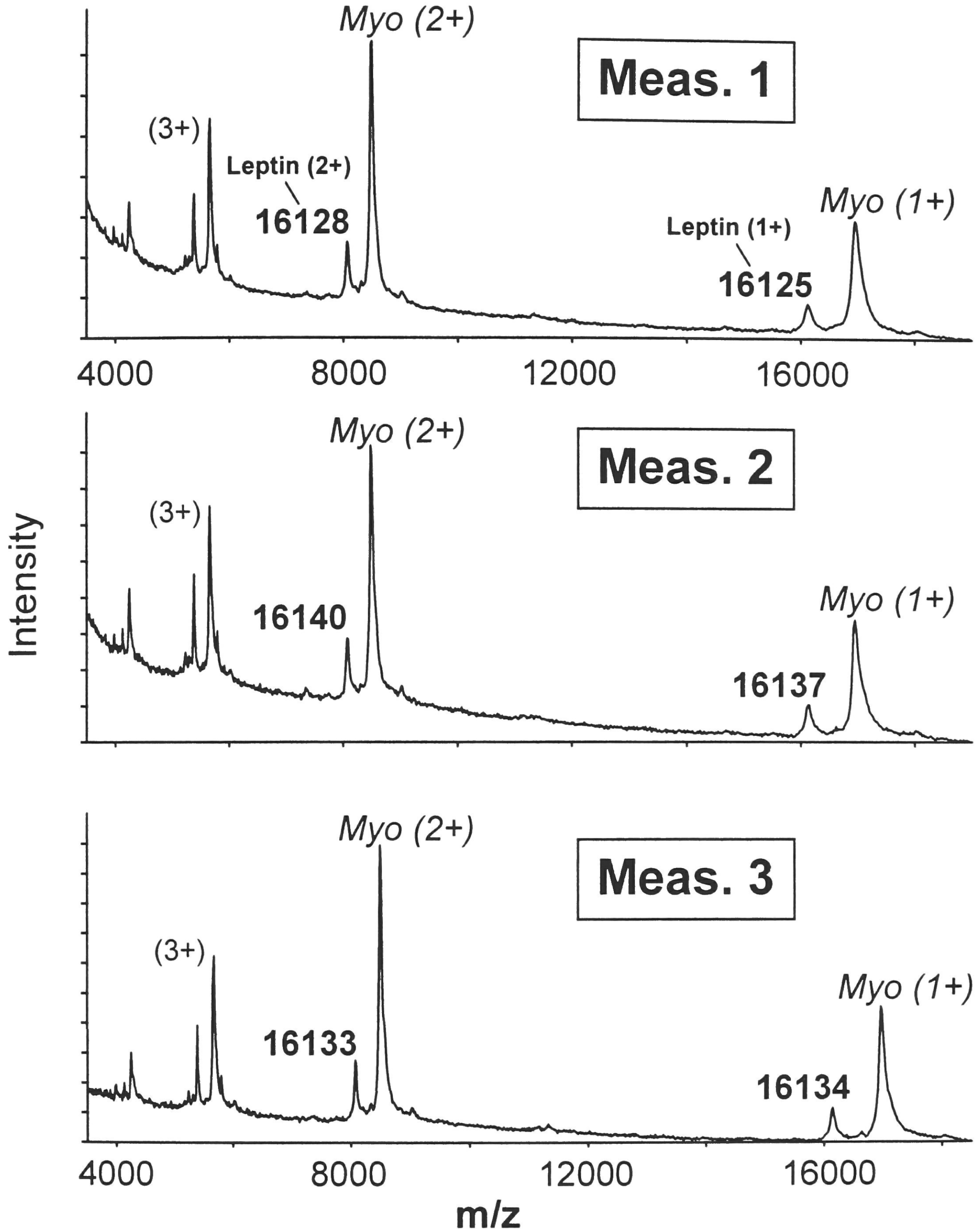
3.4.5 Mass Accuracy

The MALDI mass spectra shown in Figures 3.2-3.7 are externally calibrated. External calibration of proteins using the dried-drop method normally yields mass accuracies in the simple linear time-of-flight mass spectrometer no better than 0.1 %. With slow crystallization, other factors contribute to further reduce mass accuracy. For example, slow crystallization tends to produce larger and more heterogeneous matrix crystals than crystals formed by the dried-drop method. Also, a complication of using Extraction Method B is the inadvertent retrieval of microscopic gel pieces along with matrix crystals (Figure 3.1, step 8b). Large matrix crystals and the inclusion of gel pieces on the MALDI probe may act to distort the electric field in the ion source. These effects lead to peak broadening and centroid shifts that ultimately reduce the mass accuracy. The use of internal calibrants leads to substantial improvements in mass accuracy. With Extraction Method A, a protein of known molecular mass can be included in the matrix solution prior to the dried-drop crystallization (Figure 3.1, step 7a). Using Extraction Method B, the protein calibrant must be included before slow crystallization of the matrix (Figure 3.1, step 4b). The amount of calibrant to be added depends on the level of sensitivity and must be established on a trial-and-error basis.

albumin (66.4 kDa; Figure 3.7c). Figure 3.7a exhibits an additional set of peaks whose mass correlates to the mass marker aprotinin (6.5 kDa), indicating that aprotinin co-migrated with lysozyme on the SDS-PAGE gel.

3.4.5 Mass Accuracy

The MALDI mass spectra shown in Figures 3.2-3.7 are externally calibrated. External calibration of proteins using the dried-drop method normally yields mass accuracies in the simple linear time-of-flight mass spectrometer no better than 0.1%. With slow crystallization, other factors contribute to further reduce mass accuracy. For example, slow crystallization tends to produce larger and more heterogeneous matrix crystals than crystals formed by the dried-drop method. Also, a complication of using Extraction Method B is the inadvertent retrieval of microscopic gel pieces along with matrix crystals (Figure 3.1, step 8b). Large matrix crystals and the inclusion of gel pieces on the MALDI probe may act to distort the electric field in the ion source. These effects lead to peak broadening and centroid shifts that ultimately reduce the mass accuracy. The use of internal calibrants leads to substantial improvements in mass accuracy. With Extraction Method A, a protein of known molecular mass can be included in the matrix solution prior to the dried-drop crystallization (Figure 3.1, step 7a). Using Extraction Method B, the protein calibrant must be included before slow crystallization of the matrix (Figure 3.1, step 4b). The amount of calibrant to be added depends on the level of sensitivity and must be established on a trial-and-error basis.



<u>Figure 3.8</u> *Internal calibration improves mass accuracy.*

MALDI-MS of recombinant mouse leptin (calc. mass 16,134 Da) eluted from SDS-PAGE gel using Extraction Method B with equine myoglobin added as internal calibrant. Three repeat measurements are shown. The 1+ and 2+ leptin peaks, bracketed by the 1+ & 2+ myoglobin (Myo) peaks, are labeled with their measured mass.

found relatively good MS reproducibility. As an example, separate extractions of freshly prepared gel slices, containing 3-25 pmol loadings of recombinant mouse leptin, were performed at different times over the four-month period. The resulting MALDI mass spectra of the extracted gels using Method B yielded similar signal-to-noise ratios consistent to those shown in Figures 3.6 and 3.8. I finally note that in order to hasten the analysis of a large number of protein gel bands, the gel destaining/extraction procedures can be scaled up by processing the samples in a parallel manner.

3.5 Discussion

My approach to mass spectrometrically analyze whole proteins from SDS-PAGE gels depends on passive elution methods. The most efficient and frequently used passive gel elution employs buffer systems that normally contain SDS detergent (Hager and Burgess, 1980). However, the presence of ionic detergents, such as SDS, usually severely compromises MS response by both ESI and MALDI techniques (Beavis and Chait, 1990a & b; Mock et al., 1992). For example, I have observed that even 0.1 % of SDS in the matrix solution is 'lethal' for MALDI-MS analysis using 4HCCA matrix. Difficulty with SDS is equally true for ESI-MS. Thus, it is imperative to eliminate SDS detergent. When microgram quantities of protein are involved, detergent removal is normally carried out by protein precipitation with organic solvents such as cold

acetone (Hager and Burgess, 1980) or by ion pairing methods (Henderson et al., 1979). However, the precipitation procedures are lengthy and usually become inefficient for submicrogram quantities of protein.

Other approaches to eliminate detergent involve electroblotting proteins onto membranes, but this step usually requires staining procedures that may interfere with MS analysis. Of particular interest is the MALDI-MS of proteins *directly* from a membrane or from the gel itself. The use of infra-red (IR) MALDI-MS is one such example. Proteins are blotted onto membranes, matrix is added, allowed to dry and the entire membrane is placed into the mass spectrometer (Strupat et al., 1994). An infra-red laser is set to 'scan' the gel and desorb protein (ions) from the blotted gel bands for mass analysis. An IR laser is used (instead of the conventional UV-MALDI laser) because of the greater penetration depths (i.e., through the membrane) of IR compared with UV radiation. The field of IR-MALDI is still in the developmental stages. A second example of a 'direct' MALDI method involves the use of ultrathin SDS-PAGE gels (available from Pharmacia) (Loo et al., 1996a) and conventional UV-MALDI-MS. In this method, after resolving the proteins, the gel is soaked in a MALDI matrix solution and is carefully dried. The entire gel is placed in the mass spectrometer and is scanned by the UV-MALDI laser followed by mass analysis of desorbed protein ions. The developers of this method claim 1 pmol sensitivity for myoglobin from an ultrathin SDS-PAGE gel, although they give little evidence to support such an assertion (Loo et al., 1996a). In addition, there are several difficulties associated

with ultrathin gels. Because of their low sample loading capacities (~low μL), ultrathin gels are not widely used. The gel drying procedures are intricate and the dried gels are very delicate to handle, difficulties that could jeopardize the analysis of precious protein samples.

Part of my aim in this study was to develop a means to mass spectrometrically analyze submicrogram (low picomole) amounts of whole proteins loaded onto conventional gels with minimal sample handling. Reverse staining procedures using transition metal salts provided the desirable qualities to achieve my goal (Qin et al., 1996a). Copper staining of SDS-PAGE gels shows moderately high sensitivities (~10 ng protein) (Lee et al., 1987 and Bio-Rad, Bulletin No. 1820). More importantly, the copper destaining protocols are simple and result in complete removal of the copper stain and SDS detergent with minimal loss of protein, leaving the gel prepared for passive elution. An earlier investigation of ESI-MS of whole proteins from gels employed copper staining combined with an electroelution of the proteins (le Maire et al., 1993). This study used preparative SDS-PAGE gels and protein quantities exceeded 200 pmol. Two recent studies have demonstrated ESI-MS of proteins electroeluted from Coomassie Blue-stained (Schuhmacher et al, 1996) and copper-stained gels (Nakayama et al., 1996). In the former study, large quantities of protein (>300 pmol) were loaded onto the gels. In the latter study, the authors claim low picomole sensitivities but fail to indicate their quantification methods.

I found that passive elution of destained gels using the solvent mixture FWI (Extraction Method A) was sufficient to extract soluble proteins for MALDI- or ESI-MS analysis for protein loadings ≥ 25 pmol (~0.5 μg of myoglobin or leptin; Figure 3.2a,b). FWI was selected because of its excellent protein solubilizing properties —demonstrated, for example, in reverse-phase HPLC of proteins (Heukeshoven and Dernick, 1982). Extraction solutions of differing compositions have been employed by others, including the mixture formic acid/acetonitrile/2-propanol/water (FAPH) (50:25:15:10, by volume) (Feick and Shiozawa, 1990). FAPH was noted particularly for its ability to elute hydrophobic proteins, such as membrane proteins, with high efficiency. Typical recoveries of hydrophilic (soluble) or hydrophobic (membrane) proteins (5-10 μg loadings on SDS-PAGE gels) using passive elution by FAPH was reported to exceed 90 % using radioactively labeled polypeptides (Feick and Shiozawa, 1990). I examined FAPH as an Extraction Method A eluant for gels containing 25 and 6 pmol of either recombinant mouse leptin or myoglobin and found MALDI-MS signals comparable to those observed by eluting with FWI (i.e., Figure 3.2).

Below the 25 pmol gel loading of soluble protein, Extraction Method A showed a rapid non-linear decline in the mass spectrometric response (e.g., Figure 3.2c). For Figure 3.2c, I attribute the disappearance of the MALDI-MS signal to insignificant protein recovery from the gel at low picomole levels of gel loading. The recovery losses are probably due to irreversible adsorption of eluted protein onto the microtube and dispenser tip surfaces, loss mechanisms

known to dominate for low picomole amounts of protein (personal communication, Salvatore Sechi, Rockefeller University). To counter protein losses at low picomole levels, I combined the passive elution procedure with a second process —slow crystallization of the MALDI matrix (Xiang and Beavis, 1993).

Beavis initially described slow crystallization as a powerful strategy to grow analyte-doped matrix crystals in the presence of high levels of involatile additives, conditions that normally preclude matrix crystallization by the dried-drop method (Xiang and Beavis, 1993). I have observed an additional benefit of slow crystallization —a significant improvement in the MALDI response of proteins present in the matrix solution (especially when present in low amounts) over that observed using the dried-drop method (compare spectra in Figure 3.5). Thus, slow crystallization is a powerful means of concentrating protein into matrix crystals, providing the rationale behind Extraction Method B —protein that elutes from the gel has a high chance of being 'captured' by matrix crystals present in the extraction solution.

There are several possible shortcomings with the method of elution and MALDI-MS on picomole amounts of proteins from gels. Hydrophobic proteins or very large proteins may resist passive elution. However, the use of different solvent systems (e.g., FAPH, see Results section) (Feick and Shiozawa, 1990) or the inclusion of certain non-ionic detergents (e.g., n-octylglucoside) may assist in the elution. Also, as discussed previously, obtaining sufficient mass

accuracies using Extraction Method B can be a significant concern because of the heterogeneous matrix layer. Careful selection of internal calibrants and repetitive measurements usually provide mass accuracies better then 0.1 %. It is likely that the newly developed pulsed ion extraction methods for the MALDI ion source (Brown and Lennon, 1995; Vestal et al., 1995; Whittal and Liang, 1995) will provide a means to 'correct' the deleterious effects of the heterogeneous matrix layer. Preliminary evidence for improvement in mass accuracies has been shown (Loo et al., 1996b).

3.6 MALDI-MS of Whole Proteins from Gels —An Application

3.6.1 The characterization of native human leptin.

The present method of eluting picomole amounts of proteins from gels for MS analysis has a particular application for the study of endogenous (native) forms of proteins. This application is particularly relevant for eukaryotic proteins that normally undergo extensive *in vivo* processing events following protein translation (e.g., proteolytic cleavage, disulfide formation, phosphorylation, and glycosylation). Characterizing posttranslational modifications is routinely performed by evaluating proteolytic digests by MS peptide mapping. However, peptide coverage of protein digestions by MS is frequently incomplete, a situation that can lead to inconclusive results. Knowledge of the accurate molecular mass of the intact protein along with information from digestions and

peptide mapping can lead to more definitive results. This approach was used to characterize endogenous human leptin, the recently cloned obesity factor (Cohen et al., 1996).

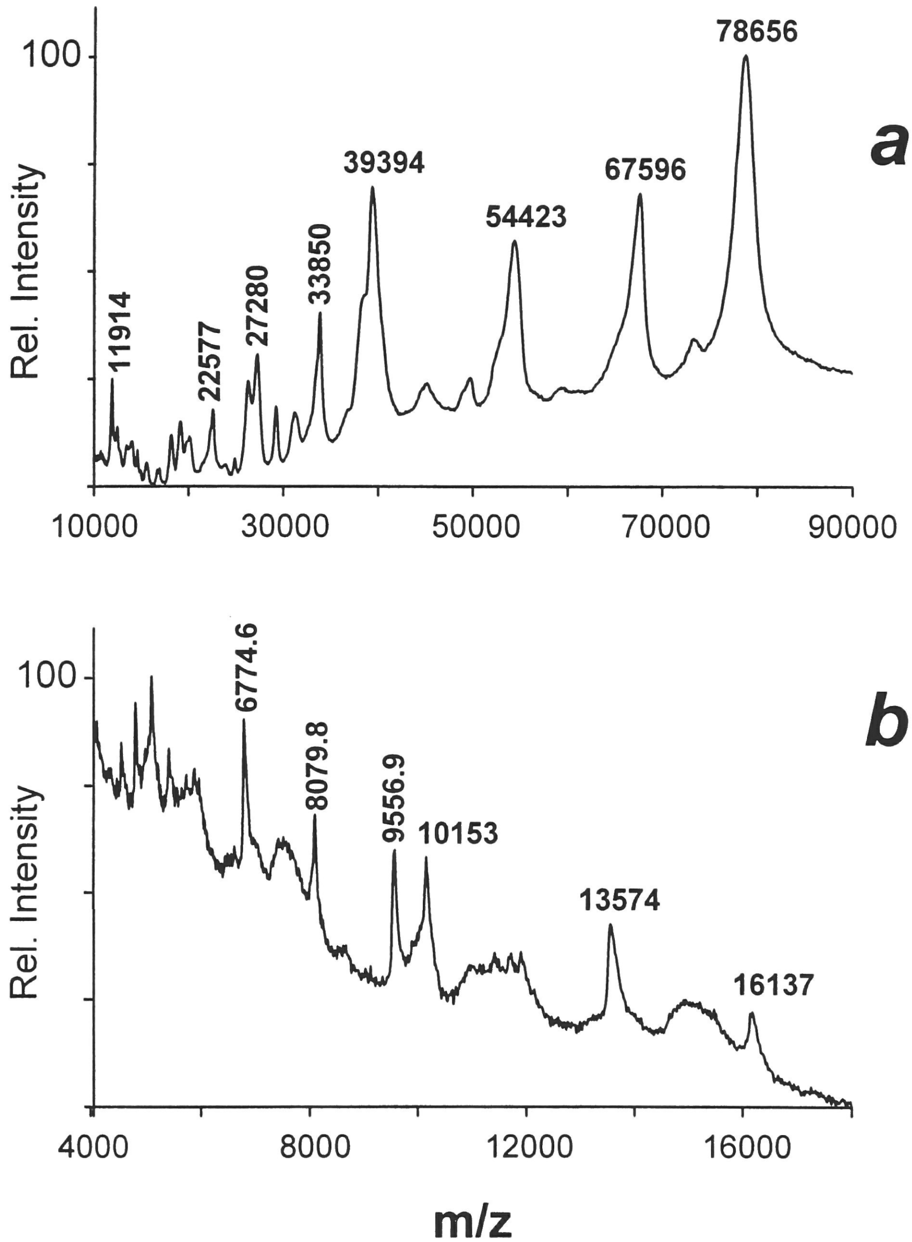
The discovery of the *obese (ob)* gene by Jeffrey Friedman by positional cloning methods has lead to new insights concerning the molecular mechanisms underlying obesity (Zhang et al., 1994). Friedman originally showed that mice homozygous for a mutant form of the *ob* gene exhibited profound heredity obesity and non-insulin-dependent diabetes mellitus. The mutation in these mice was mapped to an Arg-to-stop codon conversion in the middle of the open reading frame of the gene. The wild-type *ob* gene product was predicted to be a 146 residue secreted protein (leptin) with a single disulfide bond (Zhang et al., 1994). (In a later study, I used cyanogen bromide digestion and peptide mapping by MALDI-MS to confirm the presence of the disulfide bridge in bacterially and yeast expressed recombinant forms of leptin; see chapter 5 (Halaas et al., 1995).) The original mutation in *ob* mice was correlated to an absence of *ob* RNA transcripts (Zhang et al., 1994) and to undetectable levels of leptin in blood serum (Halaas et al., 1995). The protein is exclusively expressed in adipose (fat) tissue and at the amino-acid sequence level, the mouse and human forms of leptin are 84% identical, showing little sequence homology to other known mammalian proteins.

Subsequently, Friedman and others demonstrated that daily injections of bacterially expressed recombinant mouse leptin into obese mice, administered over a two week period, caused a marked reduction in body weight, fat, and food

intake and was accompanied by an increase in metabolic and physical activity (Halaas et al., 1995; Pelleymounter et al., 1995; Devos et al., 1996). These remarkable findings were particularly intriguing given that about a third of the American population is afflicted by some form of obesity. However, puzzling data has begun to emerge. For example, the weight-reducing effects of injections of bacterially expressed recombinant leptin into wild-type mice required relatively high levels of leptin (Halaas et al., 1995). The basis for the high dosage requirements were not known. It was also eventually disclosed that contrary to the findings from *ob/ob* mice, a majority of obese humans express *high* levels of leptin in plasma and these levels are highly correlated to the body mass index in both obese and lean human subjects (Maffei et al., 1995; Considine et al., 1996). In addition, no coding sequence polymorphisms in the *ob* gene from a pool of 105 obese human patients were detected (Maffei et al., 1996). Ultimately, clinical human trials for administration of recombinant leptin would be undertaken as treatment for obesity. Thus it became important to address some of these puzzling findings concerning leptin.

One consideration was the protein itself. Using western blotting, leptin from obese and lean humans as well as from bacterially expressed recombinant protein eluted on SDS-PAGE with identical mobilities (apparent molecular mass 16 kDa.) The poor accuracy of SDS-PAGE, however, excluded the possibility of detecting subtle mass differences (such as low molecular mass post-translational modifications). The absence (or presence) of post-translational modifications may give rise to forms of the protein with differing physiological activities.

Application of my gel extraction/MALDI-MS procedure would clarify the question concerning protein modification. I showed earlier that 1-3 pmol of protein loaded onto the gel was sufficient to acquire reasonably good MALDI-MS spectra (see, for example, Figure 3.6c). Obtaining endogenous leptin in pmol quantities, however, proved to be a challenge. The levels of leptin circulating in obese mouse blood serum were estimated (by ELISA) to be in the order of 50 ng/mL (3 pmol/mL) and about four times less in lean mice (Maffei et al., 1995). Similar leptin levels were also found in human blood serum. Jeff Halaas (biomedical fellow/Friedman laboratory) attempted to isolate leptin from mouse serum by immunoprecipitation (IP) using polyclonal antibodies. In the first trial, I ran MALDI-MS directly on the leptin IP from the mouse serum (Figure 3.9a). The spectrum is dominated by unidentifiable high-mass components with little evidence for ~16 kDa proteins. Assuming leptin was present in the IP, high levels of the high-mass components could suppress the leptin MALDI signal. To avoid such suppression problems, the next trial involved separation of the leptin IP by SDS-PAGE and copper staining. The IP gel lanes ran poorly, however, showing uneven streaks of copper staining in the low mass region. The poor performance was attributed to large loading volumes that were needed to attain pmol levels of endogenous leptin. Although no bands appeared for leptin in the IP lane, gel slices were excised by 'bracketing' the IP lane with recombinant leptin lanes. Gel extraction of the bracketed IP lane slice by Extraction Method B yielded the MALDI spectrum shown in Figure 3.9b. The spectrum shows the presence of several components including the intriguing peak at m/z ~16,100.



<u>**Figure 3.9**</u> *MALDI-MS of immunoaffinity isolated human leptin from blood serum.*

(a) Unfractionated elutant from immunoaffinity column.
(b) Elution of SDS-PAGE gel of immunoaffinity precipitation using gel Extraction Method B. The spectra show externally calibrated masses.

(The spectrum was externally calibrated, so the mass accuracy was no better than 0.1% and all attempts to calibrate internally failed.) The conclusions that I drew from the several attempts to isolate native leptin were (1) a protein separation step, such as SDS-PAGE or 2-D electrophoresis, must be performed on the IP isolated fraction and (2) gel extraction should only be applied to reasonably well defined copper stained bands.

Jeff Halaas and I eventually utilized the services of Amgen (Thousand Oaks, CA), the pharmaceutical company that has made a financial commitment to leptin. Researchers at Amgen were close to commencing clinical human trials with recombinant leptin and were very interested in characterizing the endogenous protein. Amgen had the capacity to isolate over ten times the amount of leptin that could be isolated at Rockefeller. The isolation procedures involved two stages of immunoaffinity chromatography performed on 50 mL of serum that was pooled from three moderately obese human females. Subsequently, the protein was eluted from a monoclonal antibody-affinity resin and subjected to SDS-PAGE and zinc staining. The principal band, well separated from contaminating and irrelevant proteins, showed the same mobility as recombinant human leptin (apparent relative molecular mass 16 kDa). The endogenous band was estimated to contain 250 ng (15 pmol) of leptin.

The gel bands were excised by Amgen and sent to me. I destained the gels and performed Extraction Method B. The resulting MALDI spectra are shown in Figure 3.10. When myoglobin was included as an internal calibrant, the molecular mass for endogenous leptin was measured to be 16,026 ± 9, a

value identical, within experimental error, to the molecular mass calculated from the human *ob* cDNA sequence minus the putative 21 residue N-terminal signal sequence, 16,024. For recombinant human leptin, the measured mass was 16,178 ± 10. The expected mass of recombinant protein was 16,155. The mass discrepancy for the recombinant protein was perplexing. It may be attributed to partial methionine oxidation (+16 Da), giving rise to a heterogeneous mixture impossible to resolve by low-resolution MALDI-MS. However, there was little indication of methionine oxidation in a *soluble* sample of recombinant leptin (i.e., from a stock solution of the protein) when analyzed by ESI-MS, a method which has can easily resolve oxidation peaks of a 16 kDa protein.

I further compared the proteins by performing in-gel digests with trypsin followed by MALDI-MS peptide mapping using Extraction Method B. The results of the digests are shown schematically in Figure 3.11. The peptide maps from both proteins essentially superimpose, except for the N-terminal start methionine residue present only in the recombinant protein. Cyanogen bromide digest (data not shown) confirmed the presence of a C-terminal disulfide bridge in the endogenous protein. Taken together, the MALDI-MS results unambiguously showed that endogenous human leptin from obese humans has the wild-type sequence with the predicted posttranslational processing (Cohen et al., 1996). The characterization was completed in less than two days. As a follow up, it would be interesting to characterize leptin isolated from lean humans. However, such a study would require a scale-up of the IP protocols because of the 4-5 fold lower levels of leptin in lean vs. obese subjects.

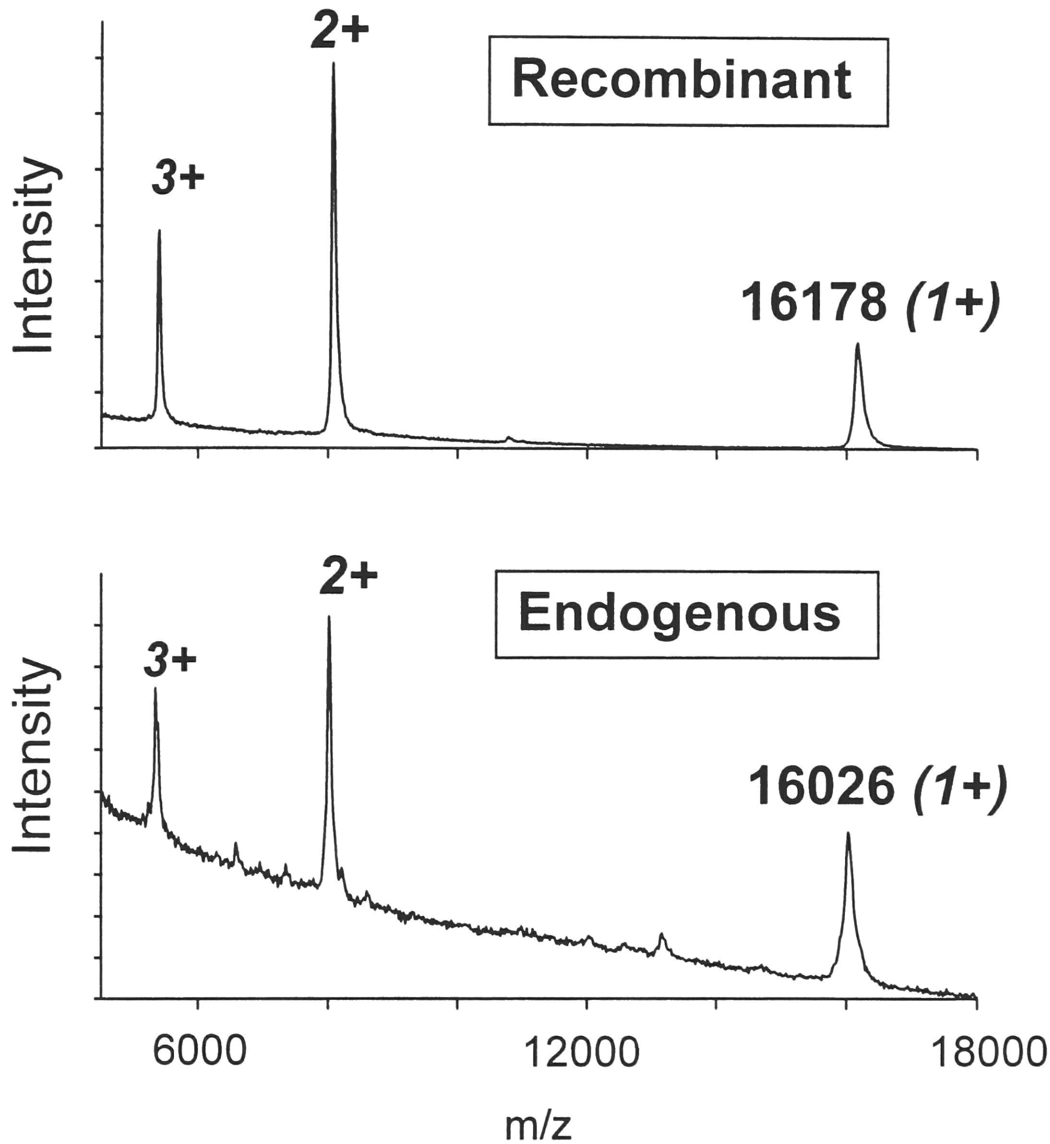


Figure 3.10 *MALDI-MS of human leptins*

MALDI-MS of recombinant ***(top)***, and endogenous ***(bottom)*** human leptin eluted from SDS-PAGE gels using gel Extraction Method B . The labeled masses were obtained from spectra that included an interal calibrant. See text for details.

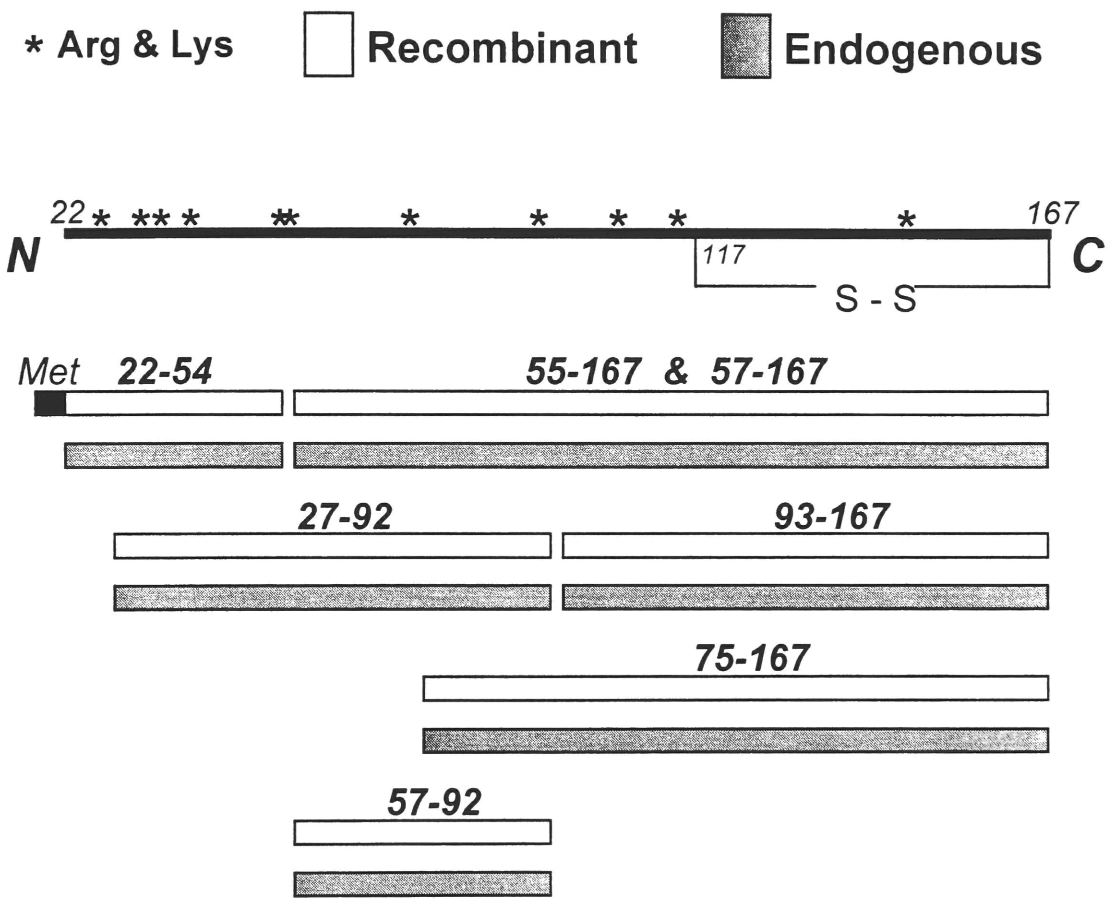


Figure 3.11 *Trypsin digestion of human leptins.*

Comparison of peptide maps of in-gel trypsin digest of recombinant and endogenous human leptins eluted from SDS-PAGE gels using Extraction Method B and MALDI-MS.

3.7 Conclusions

In addition to the characterization of endogenous proteins, I believe that the gel extraction/MALDI-MS analysis procedure will be useful in the study of disease states. Often, diseases result in dramatic changes in the expression levels of certain proteins as monitored by 2-D gel electrophoresis. These proteins can also appear on a 2-D gel in a complex manner, sometimes as multiple 'bands' (a ladder of bands) in either the IEF or the SDS-PAGE dimensions. I have recently applied the gel extraction/MALDI-MS procedure to analyze a ladder of bands from a 2-D gel of hepatocyte proteins. The tissue was from human infants suffering from neuronal ceroid-lipofuscinoses, a fatal degenerative disease. The biologist involved with this project (Bill Jordan) had spent over a year attempting to characterize the gel spots using non-MS methods. Within 4 hours, gel extraction/MALDI-MS of these gel spots revealed a molecular mass of the major protein and accurate (0.1%) estimations of the mass differences between the proteins within the ladder.

Chapter 4 — Proteolysis as a Probe for Protein Structure: Max and Related Proteins

This chapter discusses the development of a biochemical method to probe the structure of proteins using proteolysis and mass spectrometry. The DNA-binding protein Max serves as a model protein. Application of the method to phosphorylated Max and three proteins related to Max are presented.

4.1 Materials

4.1.1 Proteins

All HLH proteins were from Adrian Ferré-D'Amaré (Burley Lab/Rockefeller University). They were obtained by overexpression in *Escherichia Coli* and purified as described previously (Ferré-D'Amaré et al., 1993; Ferré-D'Amaré, 1994). Sequences of the HLH constructs are given in Figure 4.1. **Max:** Two recombinant forms of Max were studied. (1) The basic-helix-loop-helix-leucine zipper (b/HLH/Z) Max construct (Max 22-113, 92 residues, calc. mass 10,826 Da) (Amati et al., 1993). Full length Max is 160 residues). (2) A C-terminal-truncated construct (MaxΔC) that spanned the b/HLH/Z region (Max 22-113) with the additional 11 N-terminal residues (Max 2-12) resulting in a 103 residue construct, 12,045 Da. **Upstream stimulatory factor (USF):** The USF construct sequence spanned the b/HLH/Z domain (USF 197-310 + N-terminal Met,

Max
b/HLH/Z Max (Max 22-113)
92 residues, 10,826 Da

21 *30* *40*

aAD*KRAHHNALERKRR*DHIK
DSFHSLRDSVP**SLQGEKAS**R
AQILDKATEYIQY*MRRKNHT*
HQQDIDDLKRQNALLEQQVR
*ALE*KARSSAQLQT

MaxΔC
Max 2-12/22-113 (13-21 omitted)
103 residues, 12,045 Da

1 *10* *20*

mSDNDDIEVESDeeqprfqs
aAD*KRAHHNALERKRR*DHIK
DSFHSLRDSVP**SLQGEKAS**R
AQILDKATEYIQYM*RRKNHT*
HQQDIDDLKRQNALLEQQVR
*ALE*KARSSAQLQT

Upstream stimulatory factor (USF)
b/HLH/Z USF
(USF 197-310 + N-term. Met. C229S & C248S)
115 residues, 13,423 Da

191 *200* *210*

aprttMDE*KRRAQHNEVERR*
*RR*DKINNWIVQLSKIIP**DSS**
MESTKSGQSKGGILSKASDY
IQE*LRQSNHRLSEELQGLDQ*
*LQLDNDVLRQQV*EDLKNKNL
LLRAQLRHHGLEVVIKNDSN

Sterol regulatory binding protein 1
b/HLH/Z SREBP1
(SREBP1 319-407 + N-term. Gly, C404S)
90 residues, 10,074 Da

311 *320* *330*

skapasaGQSR*GEKRTAHNA*
*IEKRYRS*SINDKIIELKDLV
V**GTEAKLN**KSAVLRKAIDYI
RF*LQHSNQKLKQENLSLRTA*
*VHKSKSLKDLV*SASGSG

Hairy
b/HLH Hairy
(Hairy 30-96, C50S)
90 residues, 10,074 Da

21 *30* *40*

aqlketplkSD*RRSNKPIME*
*KRRR*ARINNSLNELKTLIL**D**
ATKKDPARHSKLEKADILEK
TVKHLQELQRQQAAMQ

Figure 4.1: *HLH constructs*

Numbering systems are based on the wild-type sequences. Calculated average molecular masses are shown. b/HLH/Z regions are dot underlined. Basic region is italicized; loop residues are in bold; leucine zipper is italicized. Acidic N-terminal of MaxΔC is doubly underlined. Double underlined residues in USF & SREBP1 are not in wild-type sequences. Lower case letters indicate wild-type residues that are omitted in the constructs.

with Cys-to-Ser mutations C229S and C248S; 115 residues, 13,423 Da). Full length USF is 310 residues. **Hairy:** *Drosophila* Hairy protein spanned the b/HLH region (Hairy 30-96, C50S, 67 residues, 7890 Da). Full length Hairy is 337 residues. **Sterol regulatory element binding protein 1 (SREBP1)**: SREBP1 spanned the b/HLH/Z domain (SREBP1 319-407+N-terminal Gly, C404S, 90 residues, 10,074 Da). Full length SREBP1 is 661 residues. All of the proteins were normally stored in 10 % glycerol, 100 mM KCl, 10 mM HEPES-KOH (pH 7.5) buffer and kept frozen at -65 or -20°C until used. Concentrations of the proteins were: b/HLH/Z Max, 3.6 mg/mL (333 μM); MaxΔC, 1.5 mg/mL (123 μM); b/HLH/Z USF, 0.7 mg/mL (52 μM); b/HLH/Z SREBP1, 3.4 mg/mL (337 μM); Hairy, 0.5 mg/mL (63 μM).

4.1.2 DNA Oligonucleotides

DNA oligonucleotides (oliogos.) were prepared synthetically in the Burley laboratory (Ferré-D'Amaré et al., 1993; Ferré-D'Amaré, 1994), chromatographically purified, annealed and stored in 100 mM KCl, 1 mM $MgCl_2$, and kept at -20°C until used. A double-stranded, palindromic 14-mer oligo. with the sequence 5'-AGGT<u>CACGTG</u>ACCT-3' (1.5 mM single strand concentration; calculated mass, 4264 Da) was used for experiments involving Max, MaxΔC, USF, and Hairy. This oligo. contains the 6 bp recognition E-box sequence (underlined in the sequence above) that is specific for high affinity DNA-binding (sDNA). Other E-box containing oligos. were the 18-mer

5'-GTAGGTCACGTGACCTAC-3' palindrome (1.25 mM ss. conc.; 5500 Da), and the 16-mer 5'-TAGGCCACGTGACCGG-3' non-palindrome (0.63 mM of each strand, 4908 Da (top strand) and 4858 Da (bottom strand with the complementary sequence)). The sterol regulatory element (SRE) non-palindromic 14-mer DNA oligo. had the sequence 5'-GTATCACCCCACTT-3' (0.6 mM of each strand; 4159 and 4369 Da). The 10 bp SRE is underlined. The double-stranded 14-mer with the sequence 5'-AGCTATAAAAGGGC-3' (TATA DNA) (1.5 mM each strand; 4322 & 4205 Da) served as the non-specific DNA (nsDNA). Another nsDNA oligo. was NOT I linker (New England BioLabs) with the palindromic sequence 5'-TTGCGGCCGCAA-3' (200 μM; 3,647 Da).

4.1.3 Other Materials

Casein kinase II (CKII; 20 μL, 10K units) was purchased in recombinant form from New England BioLabs. The sample of CKII was divided into 0.7 μL aliquots into separate 0.65 mL microtubes and stored at -70°C until needed. Dividing the sample avoids repetitive freeze-thaw cycles. The CKII concentration was given as 0.7 mg/mL (~5 pmol/μL). The buffer for performing the CKII phosphorylation was also supplied (10X) by New England BioLabs. The buffer (1X) consisted of 20 mM Tris (pH 7.5), 50 mM KCl, and 10 mM $MgCl_2$. ATP was from Sigma. All other materials —proteases, buffers, MALDI matrix solutions— are described in the Materials section of chapter 2.

4.1.4 Proteases —Preparation, Storage and Specificities

Endo Glu-C, Arg-C, Asp-N, and Lys-C were kept as lyophilized powder at 4°C for up to a year. To a 50 μg of sample of protease (typical amount purchased from Promega or Boehringer Mannheim) was added 100 μL of distilled water. The solution was distributed in 10 μL aliquots into 0.65 mL microtubes. When possible, the aliquoted samples were fast-frozen in liquid nitrogen. They were immediately lyophilized in a Speed-Vac (no heat) and stored under desiccant at 4°C. Trypsin and chymotrypsin were treated the same way but using 1 mM HCl instead of water to reconstitute the enzyme. I found that once resuspended in solution, the enzymes remained active for a few days when stored at 4°C. Subtilisin preparation: 1-2 mg of subtilisin powder was dissolved into buffer to make a 2 μg/μL solution. The subtilisin buffer was 810 μL of 100 mM ammonium acetate (pH 7) + 210 μL of 500 mM $CaCl_2$. The 2 μg/μL subtilisin solution was then diluted with the buffer solution 1:20 (v/v) to make a 100 ng/μL working solution of subtilisin which was stored at 4°C. Subtilisin remains active for many days in these solutions. For critical work, it's best to prepare a fresh solution of proteases. Read the inserts that usually accompany the enzymes when they are shipped by the manufacturer. Two practical and excellent references for proteolysis are Hames & Rickwood, 1989 and Keil, 1992. A good source for theory and applications of proteolytic enzymology can be found in Mihalyi, 1978.

Activities & specificities of proteases: I found that all of the proteases used in my studies showed good enzymatic activities when used in solutions of pH 6 to 8. Here is list some of the 'gross' details of enzyme specificity. More subtle details of specificity depend on many factors that are frequently substrate dependent. Endo Glu-C (V8) targets primarily glutamate residues, regardless of the buffer used, for most of the proteins that I've studied. Aspartic residues are secondary targets of V8 that cleave only after many hours or days of incubation. In a protein dependent manner, Endo Asp-N targets aspartate residues, but has some moderate activity towards glutamate. Lys-C virtually always targets lysines; trypsin targets exclusively lysine and arginine residues; Arg-C will hydrolyze at arginine residues, although this enzyme (from either Promega or Sigma) is not a good as Lys-C in its 'thoroughness' and tends to show a lot of non-specific cleavages. Chymotrypsin targets tyrosine and tryptophan and less avidly, phenylalanine. Over a longer period (several hours to days) chymotrypsin begins to cleave at non-aromatic residues. Finally, subtilisin has the least degree of specificity, cleaving after many types of residues.

4.2 Methods

4.2.1 MaxΔC Phosphorylation

MaxΔC was phosphorylated (when needed) with casein kinase II (CKII) according to the procedures given by New England BioLabs. Normally 15-30 μL

of 15-30 μM MaxΔC, prepared in NE BioLabs buffer (see materials), was mixed with ATP (to yield 200 μM ATP) plus an aliquot (~0.7 μL) of CKII. Phosphorylation was performed at 30°C and was usually complete in 1-2 hr.

4.2.2 Proteolytic Digestions

The stock solution of Max (or other HLH proteins) were thawed to room temperature and diluted with a 50 mM ammonium phosphate pH 6.0 buffer that included 1 mM $Mg(OAc)_2$. A pH of 6 was adequate for high enzymatic activity towards any of the proteins (without DNA) for all six proteases used in this study. Incidentally, pH 5.5 was used to prepare crystals in the original X-ray crystallographic study of the Max-DNA complex (Ferré-D'Amaré et al., 1993). The final digest solution usually consisted of 30 μL of 30 μM protein (monomer concentration). Smaller solution volumes as well as lower protein concentrations can be used when less protein is available. For each proteolytic digest experiment at least three 30 μL preparations were made: (-) DNA, (+) sDNA and (+) nsDNA. Duplex DNA was added to the solution in a ~1.5-fold molar excess over HLH dimer. The solutions were allowed to incubate at room temperature for ten minutes prior to the initiation of digestion. Digestion was started by the addition of an aliquot of an aqueous solution of a specific protease. The ratio of protease:protein ranged from 1:20 to 1:40 (w/w). The samples were hand shaken BUT NOT VORTEXED. All digests were performed at 25°C, although subtilisin and trypsin digests can be performed at lower temperatures. The time

course MALDI-MS peptide analyses were carried out in the following manner. One half μL aliquots of the digest solution were withdrawn and mixed with 15 μL of MALDI matrix solution (full details in Chapter 2) at the times 1, 2, 5, 15 min.; 1, 2, 4, 8 hr.; and 1, 2 day intervals following the start of digestion. The acidity of the matrix solution (pH < 3) completely quenched any further digestion. MALDI-MS was performed as described in Chapter 2.

A DOS computer program (PEPTIDE, written by David Fenyö) greatly facilitated the proteolytic mapping assignments. Currently the utilities "ProteinInfo" and "PAWS" in the PROWL environment (URL http://chait-sgi.rockefeller.edu) are extremely useful for peptide mapping assignments and searching for protein information. The Insight II program (Biosym Technologies, San Diego, CA), run on an Indigo 2 workstation (Silicon Graphics, Mountain View, CA), was used to create Figure 4.12. Coordinates of the b/HLH/Z Max-DNA X-ray crystallographic complex were supplied by Adrian Ferré-D'Amaré.

4.3 Introduction — Max

Max is a small transcription factor (Prendergast et al., 1991; Blackwood and Eisenman, 1991) important for the regulation of cell development, proliferation, differentiation, and apoptosis. Max is a member of the basic/helix-loop-helix/zipper (b/HLH/Z) family of DNA-binding proteins, a subset of the HLH family of proteins (Murre et al., 1989) (>100 proteins) found in a broad range of

species from *Drosophila* to human (Littlewood TD and Evan, 1994). Generally, the HLH proteins form homodimers and heterodimer complexes with each other that are competent for DNA binding. The DNA recognition sites for high affinity binding of most HLH proteins are termed E-boxes, consisting of the six-base pair palindromic sequences 5'-CACGTG-3' or 5'-CAGCTG-3'.

In vivo, Max homodimerizes with itself as well as heterodimerizes with the proteins Mad and Mxi1 (Ayer et al., 1993; Zervos et al., 1993). Max is also the obligate in vivo dimerization partner of the oncoprotein c-Myc (Blackwood et al., 1992; Amati et al., 1993). *c-myc* is a highly studied oncogene that has been implicated in many types of human cancers (Cole, 1986; Marcu et al., 1992). The oncogenic activity of *myc* has been shown to arise from processes such DNA amplification, chromosomal translocation, and point mutations, events that lead to the disruption of tight transcriptional controls over the production and *trans*-activation activity of the *myc* gene product, c-Myc (Marcu et al., 1992; Rabbitts, 1994). There is growing evidence for the existence of a transcriptional suppression network consisting of Myc, Max, Mad, and Mxi1 (Bernards, 1995). In this network, c-Myc transactivation occurs through the Myc-Max heterodimer binding to E-box consensus sites (CACGTG) in Myc-responsive promoters (Blackwood et al., 1992; Kato et al., 1992; Amati et al., 1993). Max homodimers also bind to these sites. However, since Max lacks a transactivation domain, Max homodimers do not activate transcription, but rather repress E-box containing promoters through competition with Myc-Max heterodimers and possibly other HLH transcription factors (Kretzner et al., 1992; Prendergast et al.,

```
                       -----BASIC----|---HELIX I----|-------LOOP--|
MAX      ( 22) — ADKRAHHNALERKRRDHIKDSFHSLRDSVP------SLQGEKAS
cMYC     (345) — NVKRRTHNVLERQRRNELKRSFFALRDQIP-----ELENNEKAP
MAD      ( 58) — SSSRSTHNEMEKNRRAHLRLCLEKLKGLVP-----LGPESSRHT
USF      (198) — EKRRAQHNEVERRRRDKINNWIVQLSKIIP--DCSMESTKSGQS
SREBP1   (322) — GEKRTAHNAIEKRYRSSINDKIIELKDLVV-------GTEAKLN
HAIRY    ( 30) — SDRRSNKPIMEKRRRARINNSLNELKTLIL---DATKKDPARHS
ID1      ( 73) — RLPALLDEQQVNVLLYDMNGCYSRLKELVP-----TLPQNRKVS
```

```
          --HELIX II----|----------ZIPPER-----------|
MAX      -RAQILDKATEYIQYMRRKNHTHQQDIDDLKRQNALLEQQVRALEKARSSAQLQT..(160)
cMYC     -KVVILKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA(439)
MAD      -TLSLLTKAKLHIKKLEDCDRKAVHQIDQLQREQRHLKRQLEKLGIERIRMDSIG..(221)
USF      -KGGILSKACDYIQELRQSNHRLSEELQGLDQLQLDNDVLRQQVEDLKNKNLLLR..(310)
SREBP1   -KSAVLRKAIDYIRFLQHSNQKLKQENLSLRTAVHKSKSLKDLVSACGSGGNTDV..(661)
HAIRY    -KLEKADILEKTVKHLQELQRQQAAMQQAADPKIVNKFKAGFADCVNEVSRFPGI..(337)
ID1      -KVEILQHVIDYIRDLQLELNSESEVGTTGGRGLPVRAPLSTLNGEISALAAEAA..(176)
```

FIGURE 4.2 *Sequence alignment of HLH proteins*

Sequence alignment of the DNA binding domains of mammalian Max with b/HLH/Z proteins c-Myc, Mad, USF, and SREBP1; b/HLH protein Hairy; and HLH protein Id. Boundaries of the basic, helix I, loop, helix II, and zipper regions are denoted above the sequences. Conserved (homologous) residues are highlighted in bold and underlined for Max. The five heptad repeat residues of the leucine zipper are doubly underlined. The helix-breaking proline residue in the basic region of Hairy is underlined. Id does not have a basic region. The N- and C-termini residues are numbered (in brackets) according to the complete sequence of each protein. Except for c-Myc, the sequences continue to the C-terminus residue of each protein, whose number is italicized.

1992; Amati et al., 1993). Both Mad and Mxi1 also lack transactivation domains and their heterodimerization with Max antagonizes the transformation effects of c-Myc (Bernards, 1995). More recently, the Sin3 proteins were shown to interact with Mad and strongly suppress c-Myc activation when recruited to form Max-Mad-Sin3 DNA-bound ternary complexes (Ayer et al., 1995; Schreiber-Agus et al., 1995).

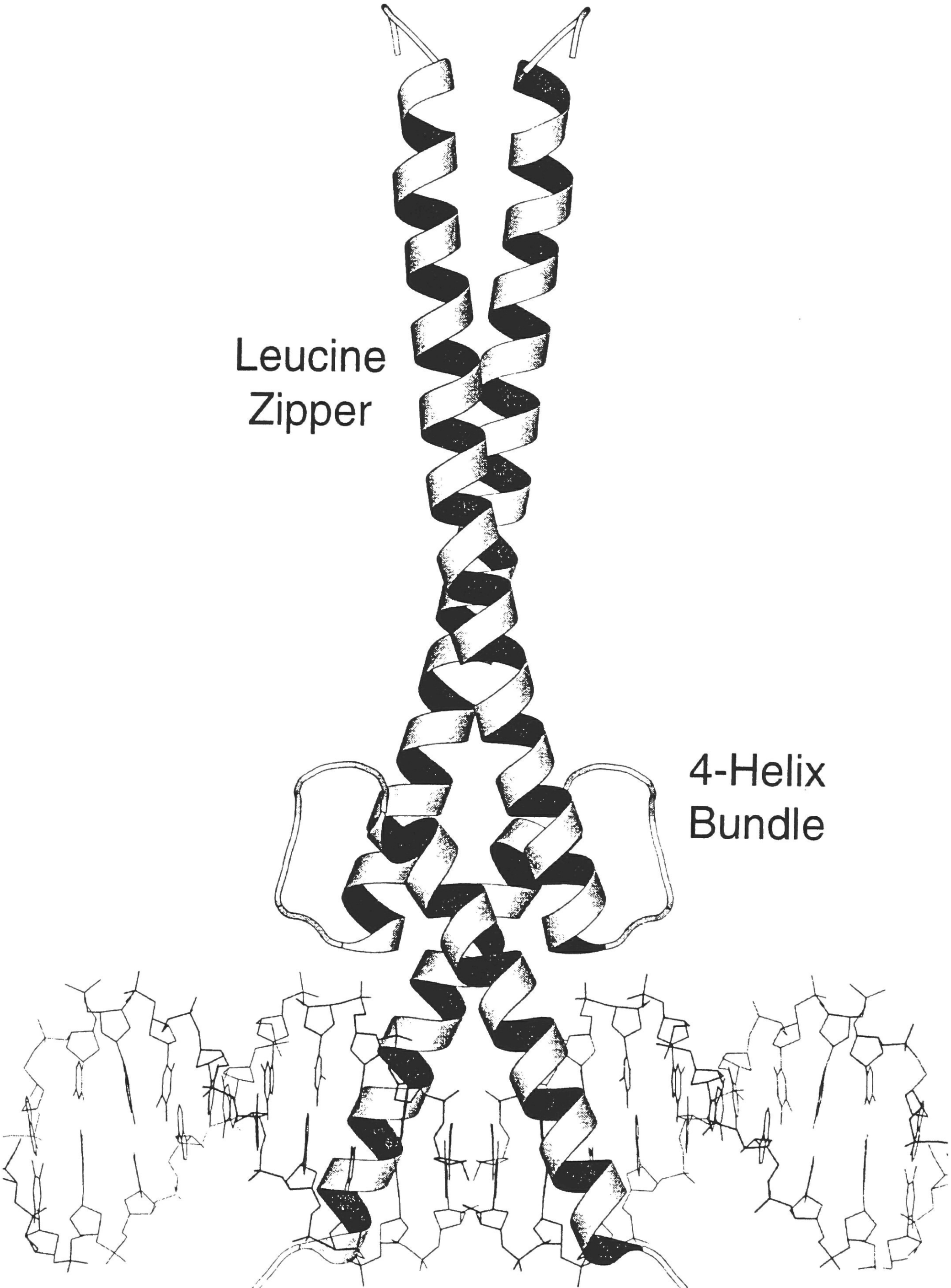
Sequence alignment of the DNA-binding regions of various HLH proteins is shown in Figure 4.2. The figure includes the b/HLH/Z proteins Max, c-Myc, Mad, the upstream stimulatory factor (USF), the sterol regulatory element binding protein 1 (SREBP1); the b/HLH protein Hairy, and the HLH protein Id. USF (Sawadogo and Roeder, 1985) is a b/HLH/Z transcription factor that forms homodimers; there is evidence for tetramerization of USF, a process that may mediate DNA looping (Ferré-D'Amaré et al., 1994). SREBP1 is a *trans*-activator for the low density lipoprotein (LDL) gene (Yokoyama et al., 1993; Wang et al., 1994). Unlike most other b/HLH and b/HLH/Z proteins, SREBP1 only weakly recognizes the canonical 6 bp E-box motifs. Instead, SREBP1 binds to a 10 bp DNA sequence called the sterol regulatory element (SRE). The b/HLH protein Hairy lacks a leucine zipper and has a helix-breaking proline residue within its basic region. Finally, Id (inhibitor of DNA binding) shows no sequence homology with other HLH proteins within the basic N-terminal region. Id is believed to dimerize with specific b/HLH targets through the HLH dimerization interface,

effectively inhibiting DNA binding and the transcriptional activity of the b/HLH dimerization partner in a dominant negative manner (Benezra et al., 1990).

Fig 4.2 shows that the HLH proteins share high sequence homologies in regions that allow them to interact with each other as well as to bind DNA (Baxevanis and Vinson, 1993). Max was the first HLH protein whose three-dimensional structure was determined and clearly revealed the nature of the dimerization and DNA-binding interactions (Ferré-D'Amaré et al., 1993). The crystal structure, solved at 2.9 Å by X-ray crystallography, showed that the Max b/HLH/Z domain binds to its cognate DNA as a symmetric homodimer. Figure 4.3 shows a ribbon diagram of the b/HLH/Z Max-DNA complex from the X-ray crystallographic study (Ferré-D'Amaré et al., 1993). There are four features deduced from the structure of the Max-DNA complex that are of consequence for the present study. These are that (1) Max binds the major groove of E-box DNA as a symmetric homodimer with its N-terminal alpha-helical basic regions, grabbing the DNA in a 'scissors grip' fashion; (2) the HLH region folds into a parallel, left-handed, four-helix bundle which is apposed to the DNA; (3) a parallel coiled coil (i.e., a 'leucine zipper') extends immediately C-terminal to the four-helix bundle; and (4) C-terminal to the zipper, the protein becomes random coil. The structures of three other homodimeric b/HLH proteins —all bound to their cognate DNA— have been determined by X-ray crystallography [b/HLH USF (Ferré-D'Amaré et al., 1994), E47 (Ellenberger et al., 1994), and MyoD (Ma et al, 1994)], providing further structural information about the HLH proteins. In

Figure 4.3 *The b/HLH/Z Max-DNA complex*

A ribbon diagram of the X-ray crystallographic structure of the b/HLH/Z Max-DNA complex (Ferré-D'Amaré et al., 1993). The peptide backbones tracing the two Max monomers are outlined as ribbons. The N-termini (residue 22) are at the bottom, the C-termini (residue 107) are at the top. The coordinates of the last six residues (108-113) were not available because the X-ray crystallographic electron density map indicated that these residues were random coil. The 4-helix bundle and leucine zipper tertiary structural elements are labeled. The duplex DNA containing the E-box recognition site is outlined across the bottom of the figure.


Leucine
Zipper
4-Helix
Bundle

addition to the crystallographic analyses, studies using genetic (i.e., mutational) and biochemical techniques have provided insight into structural features that control the dimerization, DNA-binding specificities and functions of Max and other HLH proteins (reviewed in Littlewood and Evan, 1994).

To further elucidate the biochemical mode of action of Max and other members of the biologically important helix-loop-helix family, it is necessary to understand the relationship between the structures and functions of these proteins in solution. However, little is known about the structure of the HLH proteins in solution, although NMR spectroscopy has been used to study a disulfide-linked dimer of a MyoD-bHLH peptide, which is inactive in DNA binding (Starovasnik et al., 1992), and electron paramagnetic resonance spectroscopy has been used to distinguish between alternate folding topologies of a MyoD bHLH peptide (Anthony-Cahill et al., 1992). (Incidentally, no X-ray crystallographic information is available on the structure of Max in the *absence* of DNA due to a failure of crystallization of unbound protein.) Although multidimensional NMR spectroscopy and mutational analyses are of pivotal utility for solution studies of proteins, it remains desirable to develop additional complementary biochemical approaches that can be applied with ease and speed.

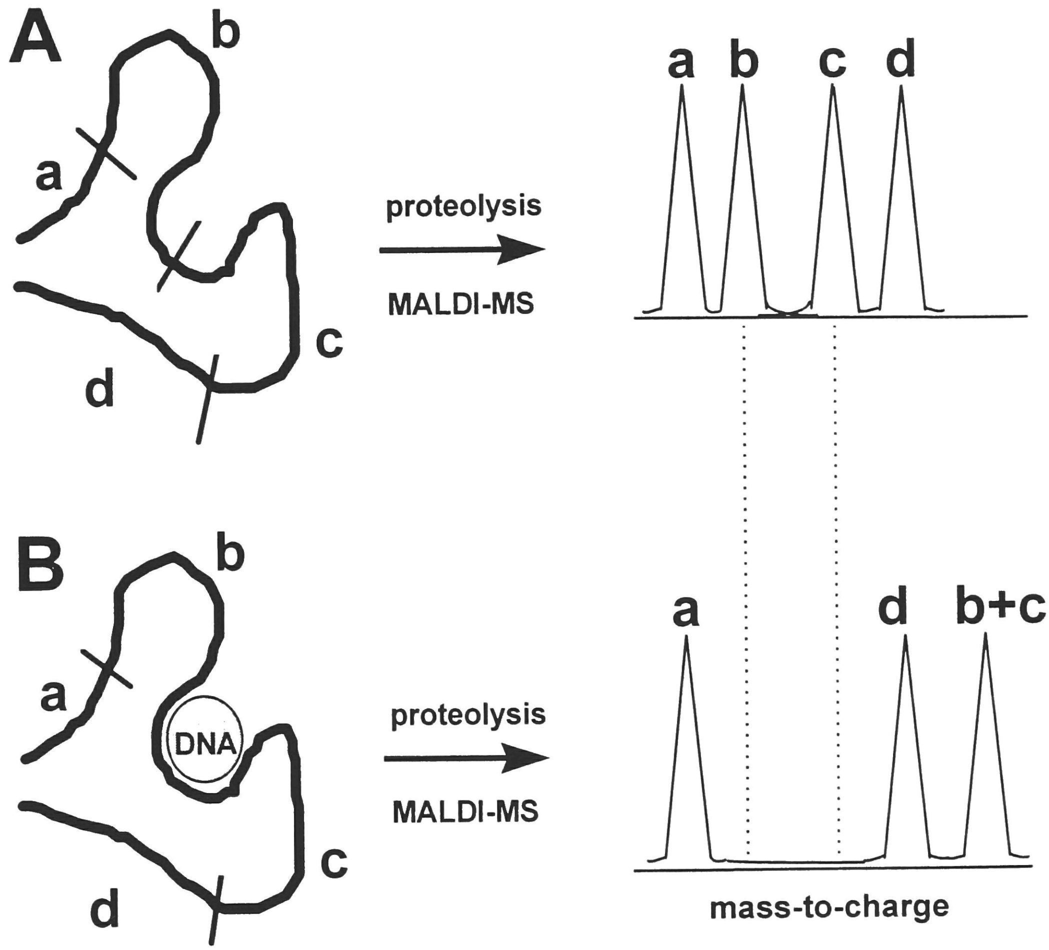
In this present study, I developed an approach that combines enzymatic proteolysis with mass spectrometry to investigate the solution behavior of the Max protein. The method involves performing time-course limited proteolytic

digestions of the protein in both the absence and presence of DNA (Fig. 4.4). Evaluation of the resultant proteolytic fragmentation patterns along with the rates of digestion provides information about cleavage site accessibility and flexibility of the protein in solution. Fundamental to this approach is the notion of protection against proteolysis. Protection is conferred on regions of the protein that are buried, are within a rigid structure, or are involved in protein-protein or protein-DNA interactions. In contrast, regions of the protein that are solvent accessible and flexible or unstructured will be less protected and therefore susceptible to proteolytic cleavage (Fruton, 1975; Fontana, 1989; Hubbard et al., 1994). An essential requirement of the method is a facile and accurate identification of the proteolytic products. MALDI-MS serves this analytical need well because of its superior ability for measuring the molecular masses of mixtures of peptides and proteins (discussed in Chapter 2). Because of the speed and mass accuracies associated with MALDI-MS, the progress of a large number of enzymatic digestions and surveying a wide variety of solution conditions can be achieved practically 'on the fly'.

This study of Max was partly retrospective because the b/HLH/Z Max-DNA X-ray crystallographic structure (Ferré-D'Amaré et al., 1993) was available for a comparative analysis. In the proteolysis study I addressed several specific questions concerning b/HLH/Z Max. What are the structural characteristics of Max in the absence of DNA? How does solution ionic strength affect the structure of the protein? How does the presence of DNA alter the protein? How

Figure 4.4 *Proteolytic protection assay*

An illustration of the mass spectrometric proteolytic protection assay. The figure compares mass spectrometric peptide maps obtained by enzymatic proteolysis of a hypothetical DNA-binding protein, in the absence and presence of DNA. A thick black line traces the polypeptide backbone of the protein containing three proteolytic cleavage sites. **A:** In the absence of DNA, the protein is proteolytically cleaved into four peptide fragments (a, b, c, and d). The mass spectrum of the resulting digest shows four peaks with masses that can be determined with sufficient accuracy to unambiguously identify the four peptide fragments. **B:** In the presence of DNA, one of the three potential sites of proteolysis becomes protected following DNA binding. Under these conditions, the protein is cleaved into three fragments (a, d, and b+c) that can be identified in the mass spectrum. Because the site between b and c is protected against proteolysis, a single peak, corresponding to the combined fragment b+c, appears at higher mass. Comparison of the mass spectra provides information regarding the DNA-binding region. The scheme illustrated is a simplification of the actual requirements for cleavage. In addition to accessibility, there is also the necessity that the cleavage sites be located within flexible segments of the protein, in order to optimize interaction with the extended active site of the protease (Fruton, 1975; Hubbard et al., 1994).


A
a
b
c
d
proteolysis
MALDI-MS
a b c d
B
a
b
DNA
c
d
proteolysis
MALDI-MS
a
d
b+c
mass-to-charge

does the solution structure of the Max-DNA complex compare to the structure deduced from X-ray crystallography?

4.4 Results

Max protein (with and without DNA) was subjected to time-course measurements of proteolysis employing six different endoproteases: Glu-C (V8), Lys-C, Asp-N, trypsin, chymotrypsin and subtilisin. MALDI-MS was used to follow the digestions and accurately map the sites of cleavage in the protein.

4.4.1 V8 Proteolysis of b/HLH/Z Max in the Absence of DNA

Figure 4.5 shows the time-course of a V8 digest of Max in the absence of DNA. V8 protease targets primarily glutamate residues, of which b/HLH/Z Max has five (see Figure 4.1). The digestions were performed at 25°C in 50 mM ammonium phosphate buffer (pH 6) and at a low salt concentration (15 mM KCl). The masses of the proteolytic fragments were determined with sufficient accuracy to enable their unambiguous identification. In the absence of DNA, V8 protease digests Max very rapidly (Figure 4.5a-c). After just two minutes, considerable cleavage is observed from all five glutamate residues (Figure 4.5b). By one hour, the digestion is nearly complete (Figure 4.5c). These findings demonstrate that, in the absence of DNA and at low salt, b/HLH/Z Max adopts a

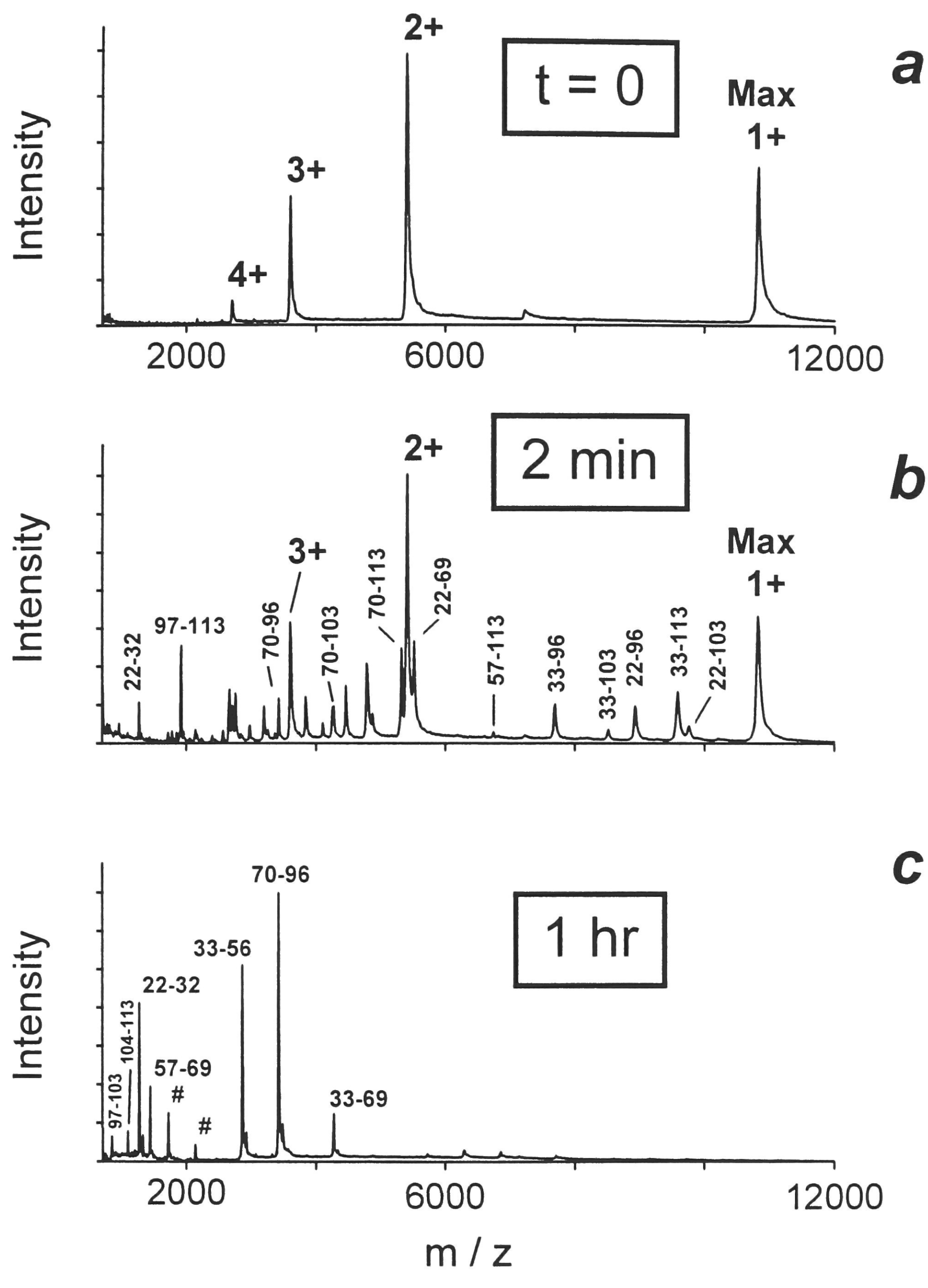


Figure 4.5 *V8 digest of b/HLH/Z Max in the absence of DNA*

MALDI mass spectra of the products of a time-course V8 proteolysis of b/HLH/Z Max in the absence of DNA. Digests were performed in 15mM KCl at pH 6 buffer and 25°C. Three time points are shown. **a:** 0 s (before addition of V8); **b:** 2 min of digestion; and **c:** 1 hr of digestion. Singly charged V8 fragment ions are labeled with their sequences. Multiply charged fragment peaks are marked with the pound symbol, #. Peaks labeled 1+ arise form the singly protonated monomeric Max protein. Peaks labeled 2+-4+ arise from multiply protonated monomeric Max resulting from the MALDI process.

form that is highly susceptible to V8 proteolysis, correlating to a rather open and flexible structure.

4.4.2 V8 Proteolysis in the Presence of Max-specific DNA

The presence of a 50% excess molar quantity (over protein) of a 14-mer Max-specific DNA oligonucleotide (sDNA; see Materials) leads to a dramatic decrease in the proteolytic susceptibility of the protein. Thus, after a two minute period, Max exhibited no V8 digestion (compare Figure 4.6b to Figure 4.5b). After one hour, only a hint of digestion has occurred and this from the C-terminal (fragment 22-103, Figure 4.6c). Even following 48 hours of digestion, a significant fraction of Max remains undigested (Figure 4.7). The fragmentation that does occur by 48 hours falls into two categories. The first resembles the complete digestion observed in the absence of DNA. The second arises from highly selective partial proteolysis of Max to yield two C-terminally truncated fragments, 22-103 and 22-96. These two fragments appear sequentially, first 22-103 (after one hour, Figure 4.6c) followed by 22-96 (after several hours, Figure 4.7). The long persistence of intact protein and fragments 22-103 and 22-96 indicates that in the presence of sDNA, Max adopts a form that is highly protected against V8 proteolysis. In addition, fragments 22-96 and 22-103 retain a high level of protection against further digestion, suggesting that these

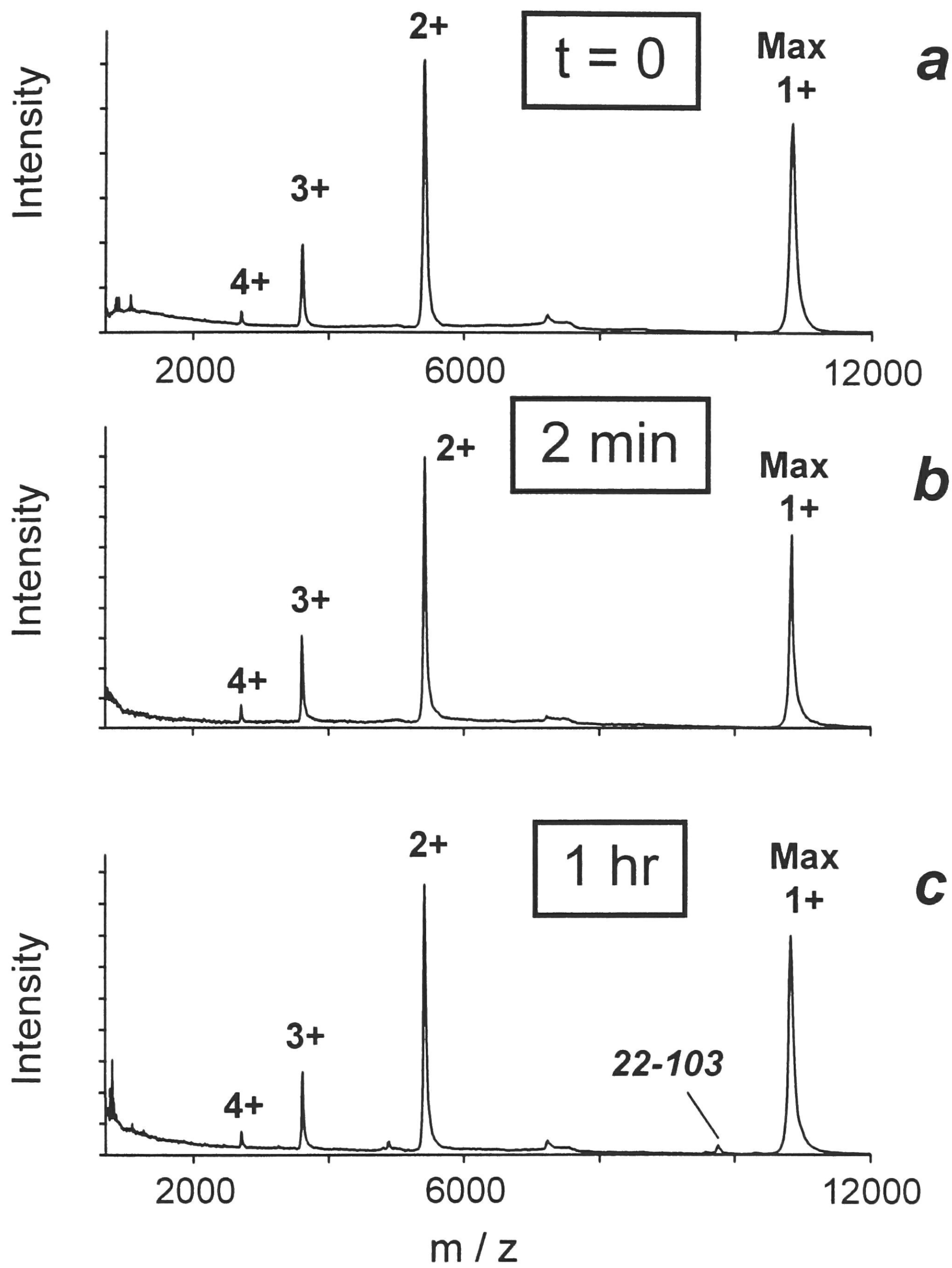
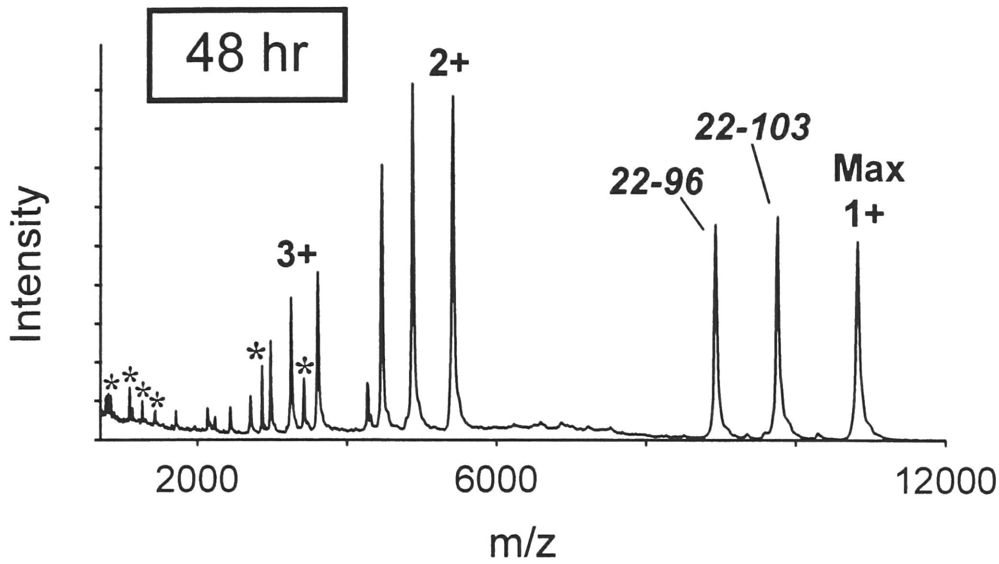


Figure 4.6 *V8 digest of b/HLH/Z Max in the presence of E-box DNA*

MALDI mass spectra of the products of a time-course V8 proteolysis of b/HLH/Z Max in the presence of E-Box DNA. Digests were performed in 15mM KCl at pH 6 buffer and 25°C. Three time points are shown. **a:** 0 s (before addition of V8); **b:** 2 min of digestion; and **c:** 1 hr of digestion. See Figure 4.5 caption for full description of all labels and symbols.



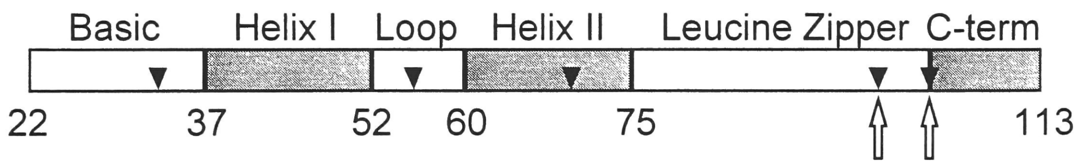


Figure 4.7 *V8 digest (48 hr) of b/HLH/Z Max + E-box DNA.*

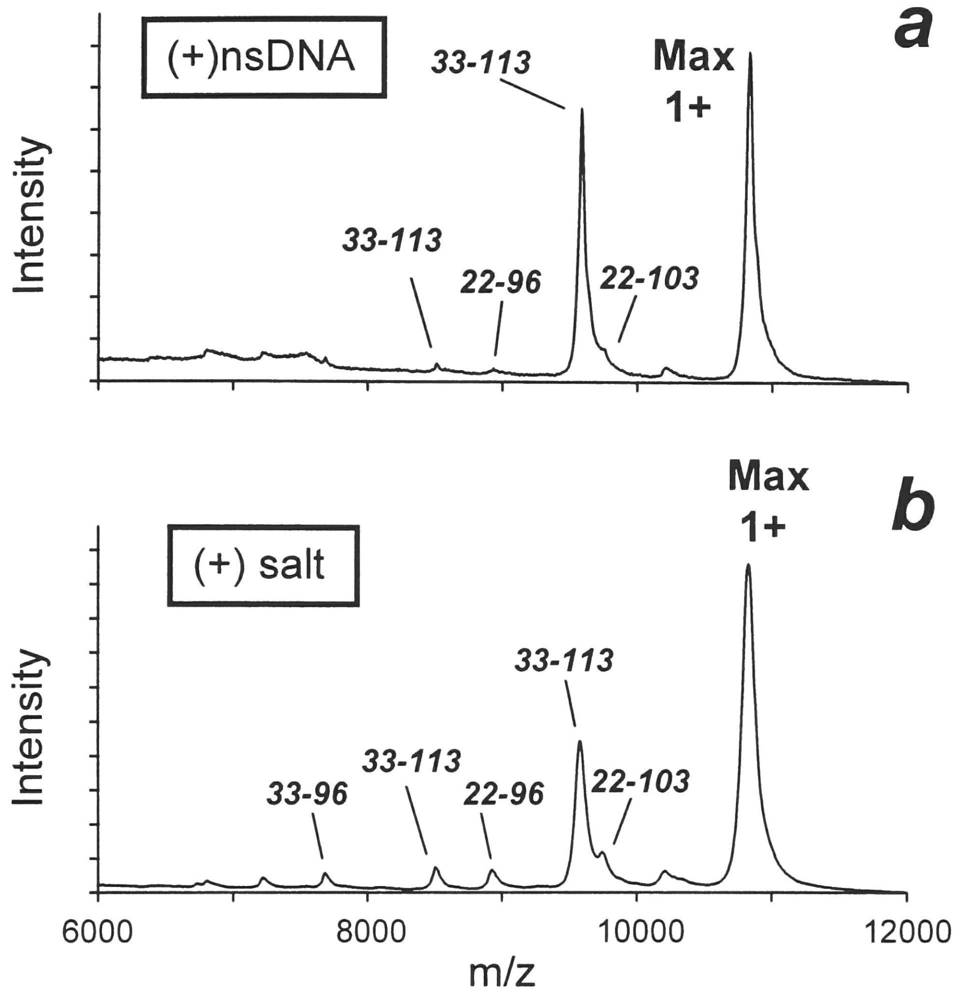
TOP: MALDI mass spectrum of a 48 hr V8 digest of Max in the presence of Max-specific (E-box) DNA. Two peaks, 22-96 and 22-103 are the singly protonated V8 digest fragments of the protein. Peaks labeled with an asterisk represent complete V8 digest fragments.

BOTTOM: Linear diagram of the sequence of b/HLH/Z Max showing the two V8 cleavage sites in the DNA-bound form of the protein that are accessible to protease (open arrows pointing to Glu96 and Glu103) out of the five Glu residues accessible to V8 in the unbound form of Max (solid arrowheads).

fragments remain tightly bound to DNA. Results of the 48 hour *partial* V8 proteolysis are summarized in the diagram at the bottom of Figure 4.7

4.4.3 V8 Proteolysis in the Presence of Non-specific DNA

Figure 4.8a shows a five minute V8 digest of Max performed in the presence of a 14-mer non-specific DNA (nsDNA), which does not contain the E-box Max recognition site (TATA DNA; see Materials). From Figures 4.8a and 4.5b, it is seen that the rate of digestion of Max has slowed in the presence of nsDNA compared to that in the absence of DNA. The rate of digestion, however, is considerably greater than that observed in the presence of sDNA (compare Figures 4.8a and 4.6c). Figure 4.8a shows that after five minutes of digestion of Max in the presence of nsDNA, two principal components dominate. The first is undigested Max and the second is the fragment 33-113, which arises from a cleavage in the N-terminal region of the protein. A minor amount of cleavage from the C-terminal region is also seen (fragments 33-103, 22-96, and 22-103). This pattern of fragmentation shows that in the presence of nsDNA, susceptibility to V8 proteolysis follows the order N-terminal region >> leucine zipper > four-helix bundle. Following rapid cleavage of the N-terminal of the protein, digestion of the remaining V8 sites is slower, requiring over four hours for completion (data not shown). These results indicate that the presence of nsDNA reduces the susceptibility of the helix-loop-helix-zipper regions of Max to V8 proteolysis, possibly through non-specific DNA interactions. In an attempt to



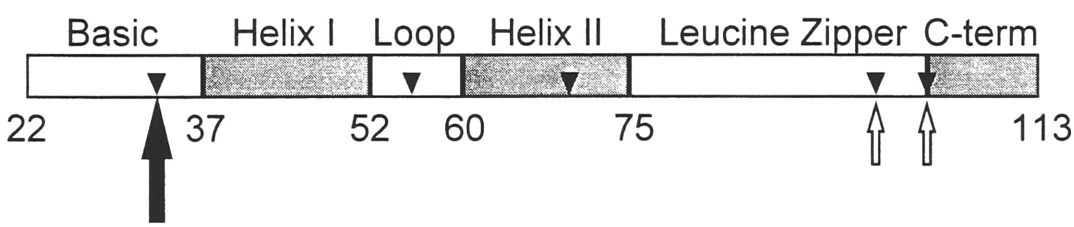


Figure 4.8 *V8 digest of b/HLH/Z Max with nsDNA or salt.*

TOP: a, MALDI mass spectrum of a 5-min V8 digest of b/HLH/Z Max performed in the presence of nsDNA (TATA DNA; see Materials). Digest was carried out at 15 mM KCl. Cleavage at the N-terminal of Max gives rise to the peptide 33-113. (For simplicity, only the high-mass portion of the spectrum is shown); and **b,** MALDI mass spectrum of a 5-min V8 digest of b/HLH/Z Max in the absence of DNA but at 150 mM KCl. Note the similarity of two the spectra. **BOTTOM:** Linear diagram summarizes the V8 results. Solid arrowhead indicate Glu residues. Dark arrow signifies most rapid cleavage (min); open arrows signify slower cleavage (min-hrs).

diminish such non-specific interactions, a V8 digest of Max (in the presence of nsDNA) was carried out at 150 mM KCl (data not shown). The rates and patterns of V8 cleavage of Max under these conditions were similar to those obtained at 15 mM KCl. It should be recognized, however, that an elevated salt level *by itself* promotes a tightening of the dimeric structure of the protein (see below.)

4.4.4 Proteolysis and the Effects of Ionic Strength

Increasing the ionic strength had notable effects on the V8 proteolysis of Max in the absence of DNA. Figure 4.8b shows a five minute V8 digest of Max in 50 mM buffer and 150 mM KCl. Compared to the digestion carried out under low salt conditions (15 mM KCl), the elevated ionic strength leads to a modest reduction in the rates of proteolysis and a large change in the patterns of fragmentation (compare Figure 4.8b and 4.5b). At elevated ionic strength, digestion occurs primarily from the N-terminal and to a lesser extent from the C-terminal regions of Max (Figure 4.8b). Even after one hour of digestion, there remains a portion of high-mass peptides, fragments 33-96, 33-103 and 33-113 (data not shown). These findings indicate that physiological salt levels lead to a considerable protection against proteolysis of the four-helix bundle. Two other points are noteworthy. First, the pattern of proteolytic cleavage of Max, measured at physiological salt levels (Figure 4.8b), is remarkably similar to the

pattern obtained in the presence of nsDNA and low salt (Figure 4.8a). Second, moderate ionic strengths (150 mM KCl) have little effect on either the rates or patterns of proteolysis of Max in the presence of sDNA (data not shown).

4.4.5 Proteolysis of the Leucine Zipper

A sequential proteolytic removal of residues of the leucine zipper was shown to occur during the V8 digestion of Max (Figs. 4.6a-c and 4.7). Similar and even more striking effects on the leucine zipper were observed with the use of trypsin and subtilisin. Subtilisin has the broadest specificity of all the endoproteases used in this study. After just one minute, in the absence of DNA, Max undergoes extensive digestion by subtilisin, leaving very little starting protein intact (Figure 4.9a). By one hour, only small peptide fragments (<2500 Da) remain (data not shown). This outcome changes dramatically in the presence of sDNA. At one minute of digestion, a highly selective cleavage has occurred between residues 108 and 109 on a portion of the protein, although a substantial proportion of Max (>50%) remains undigested (Figure 4.9b). After four hours, digestion continues to occur exclusively from the C-terminal region in a sequential manner (Figure 4.10a). Over the next two days, the fragmentation progresses sequentially through the leucine zipper, in the N-terminal direction until just three fragments dominate, 22-77, 22-83 and 22-84 (Figure 4.10c). The progression of cleavage abruptly stalls at residue 77, near the base of the

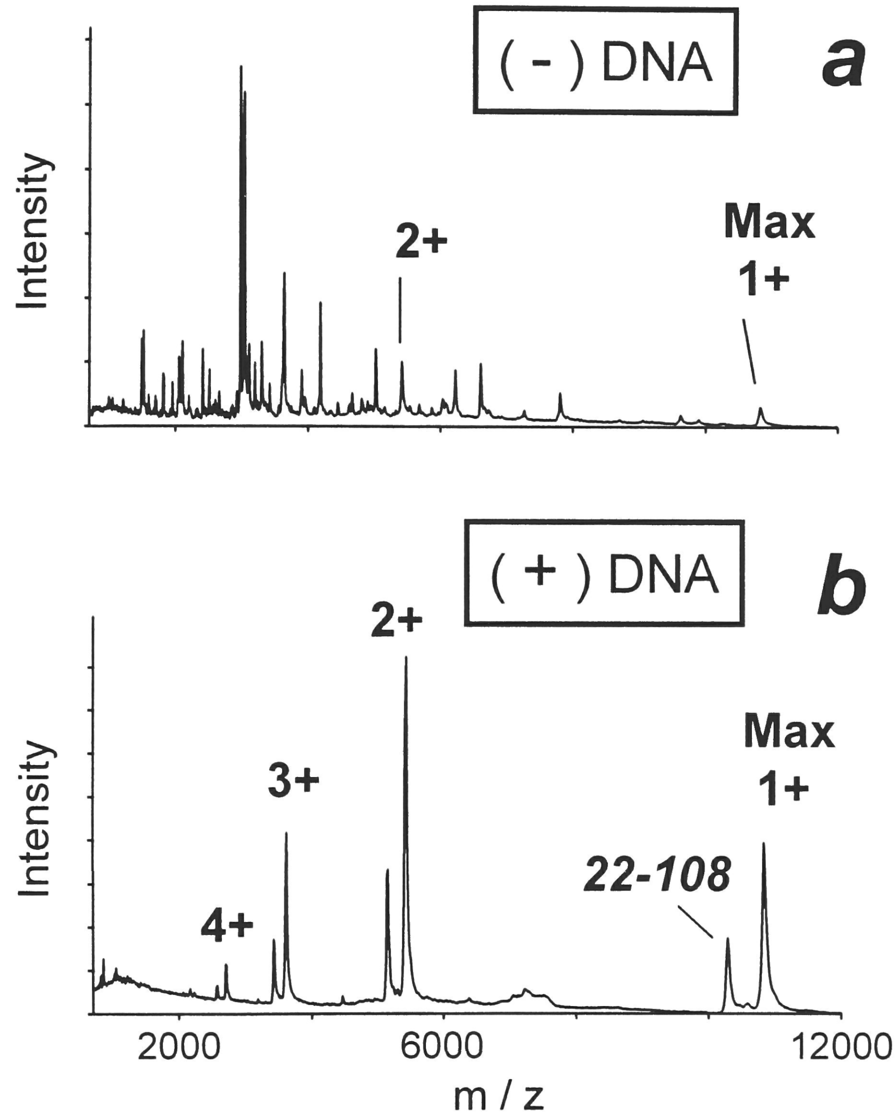
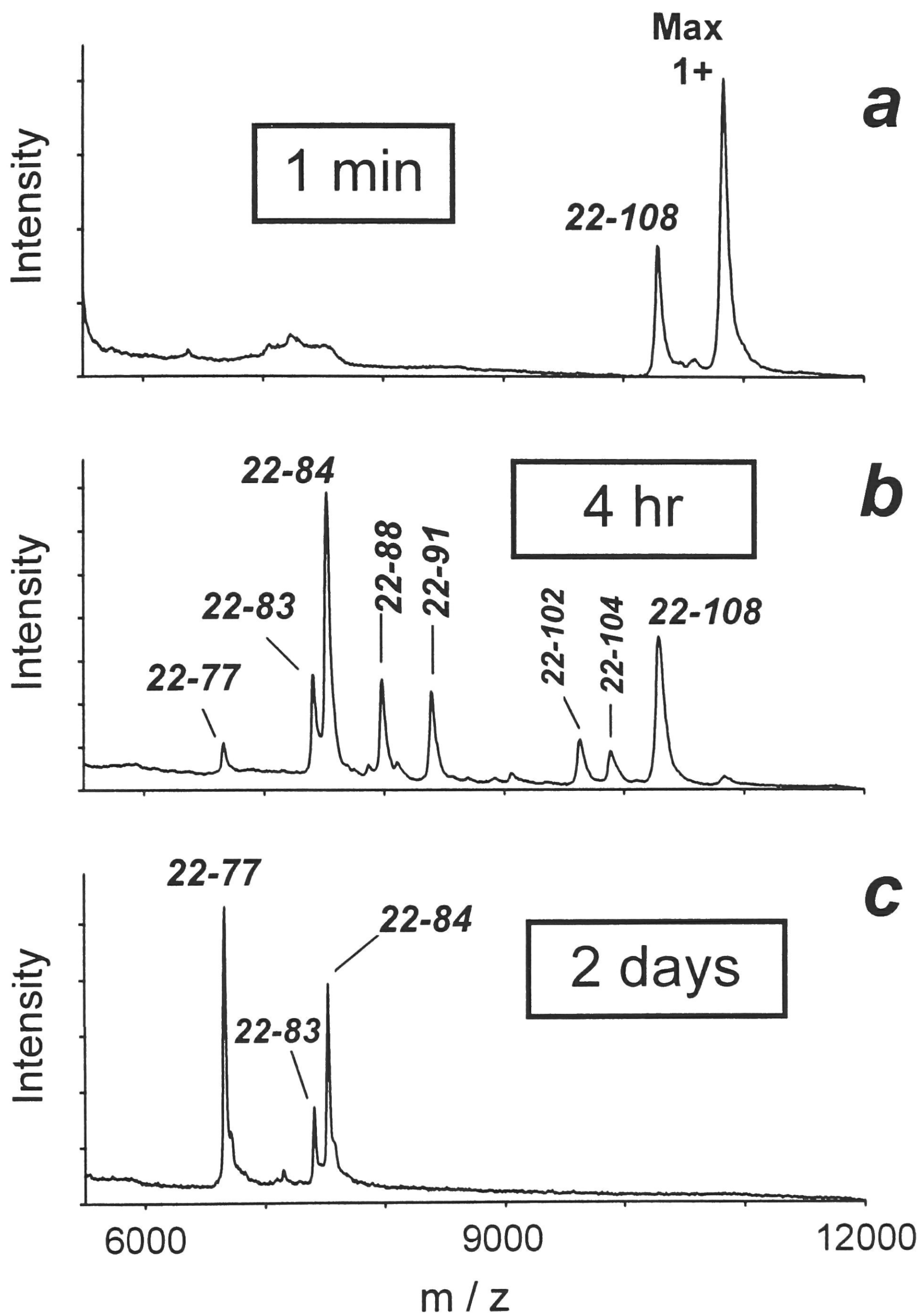


Figure 4.9 *Subtilisin digest (1 min) of b/HLH/Z Max (-/+) DNA.*

MALDI mass spectra of the subtilisin proteolysis of b/HLH/Z Max.
a: The 1-min subtilisin digest in the absence of DNA. Extensive cleavage has occurred. The large number fragment peaks are not labeled. **b:** The 1-min subtilisin proteolysis of Max in the presence of E-box DNA. Rapid cleavage at residue 108 is observed resulting in the fragment 22-108.



<u>**Figure 4.10**</u> *Subtilisin digest of b /HLH/Z Max (+) E-box DNA.*

MALDI mass spectra of the subtilisin proteolysis of b/HLH/Z Max in the presence of E-box DNA. **a, b, c:** The 1-min, 4-hr, and 2-day subtilisin proteolysis. Only the relevant high mass regions are shown. The spectrum in **a** is shown complete in Figure 4.9b. Rapid cleavage at residue 108 is followed by a slow, sequential cleavage through the leucine zipper. Cleavage stalls at residue 77.

leucine zipper and just above the four-helix bundle. This 'unzipping and stalling' pattern also appears in the V8 digest results (Figures 4.6 and 4.7) and the trypsin results (data not shown). For the trypsin digestion, the progression of cleavage stalls at residue 75.

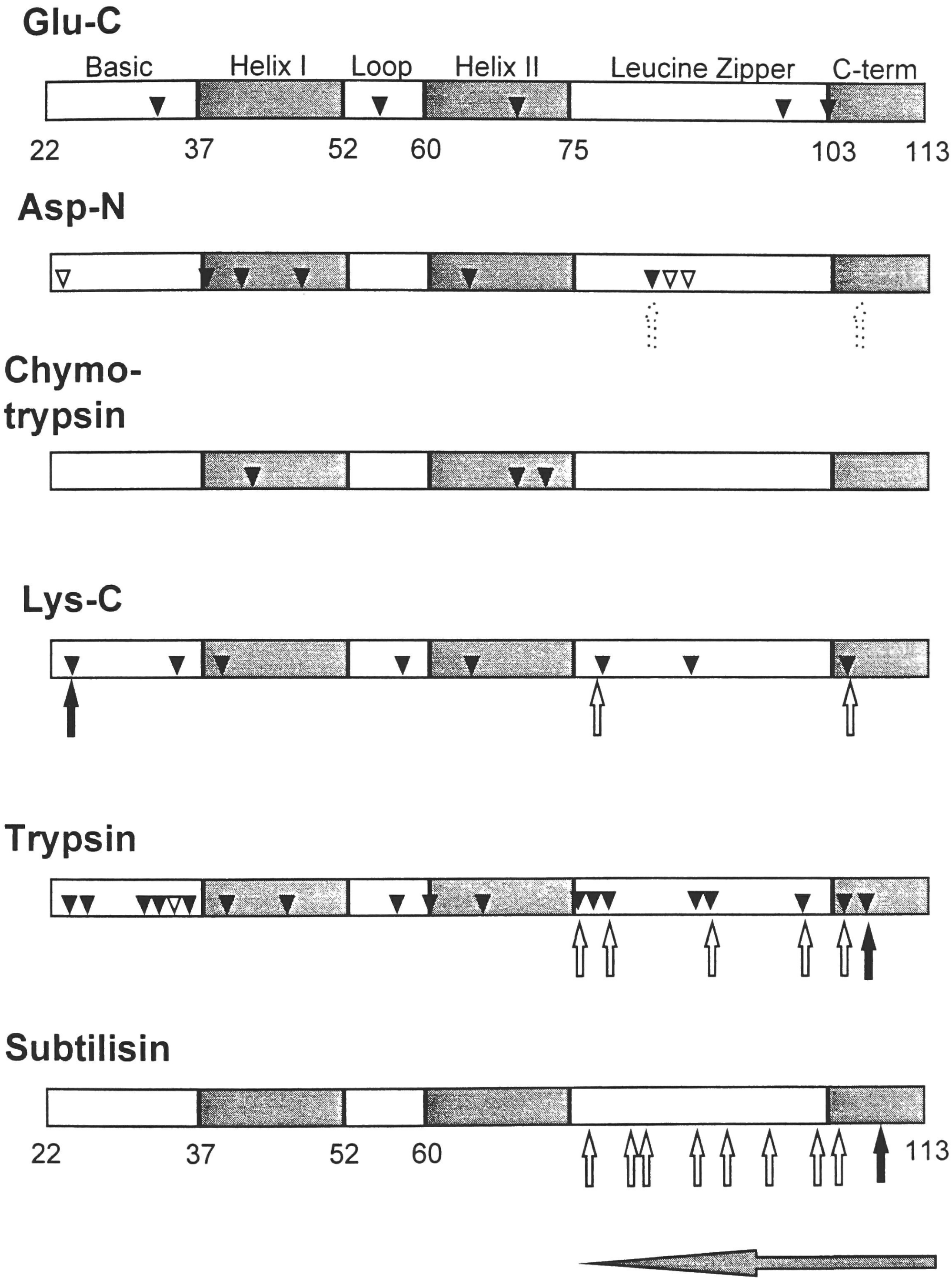
4.4.6 Additional Proteases

In a manner similar to the above-described experiments with V8, subtilisin and trypsin, Max was also extensively studied with endo Asp-N, Lys-C, and chymotrypsin. Several coherent themes emerge by combining the results obtained using all six proteases. First, the cleavage of Max, in the absence of DNA and at low ionic strengths, was virtually complete after two hours of digestion by any of the six proteases. Elevated ionic strengths slowed the V8 proteolysis considerably (Figure 4.8b) but had relatively less effect in slowing the digestions of Max with the five other proteases. A second theme was that in the presence of sDNA the rates of digestion of Max were greatly lowered —as previously described for the V8 and subtilisin results. Over a two day period, all of the proteases studied showed some degree of complete as well as partial digestion of the Max-DNA complex. The extent of complete digestion varied—V8, Lys-C and chymotrypsin showed the greatest whereas Asp-N showed the least.

The data involving partial digestion are the most informative in characterizing the Max-DNA structure. Results for six of the proteases are

Figure 4.11 *Summary of proteolysis of the b/HLH/Z Max-DNA complex*

Results summarizing the partial proteolysis of b/HLH/Z Max in the presence of DNA by the endoproteases Glu-C, Asp-N, chymotrypsin, Lys-C, trypsin, and subtilisin. Small arrowheads inside the diagrams point to the respective potential proteolytic cleavage sites. The cleavage sites can be divided into two types: those that are cleaved by protease in the absence of DNA (indicated by the solid arrowheads) and those that are not cleaved in the absence of DNA (indicated by open arrowheads). The large arrows outside the diagrams point to sites of proteolysis of the Max-DNA complex, as measured by MALDI-MS, that *result in fragments which maintain DNA-binding*. Three levels of proteolysis are represented: a solid arrow signifies a rapid and complete cleavage (minutes to hours), an open arrow signifies a slower and nearly complete cleavage (hours to 2 days), and an arrow drawn in broken outline indicates a very slow and incomplete cleavage (2 days or more). Proteolysis by Asp-N included the non-specific cleavage between Ala105 and Arg106 which did not occur in the absence of DNA. Proteolysis by subtilisin occurred sequentially over a two day period through the C-terminal in the direction given by the horizontal arrow.



Glu-C
Basic
Helix I
Loop
Helix II
Leucine Zipper
C-term
22
37
52
60
75
103
113
Asp-N
Chymo-
trypsin
Lys-C
Trypsin
Subtilisin
22
37
52
60
113

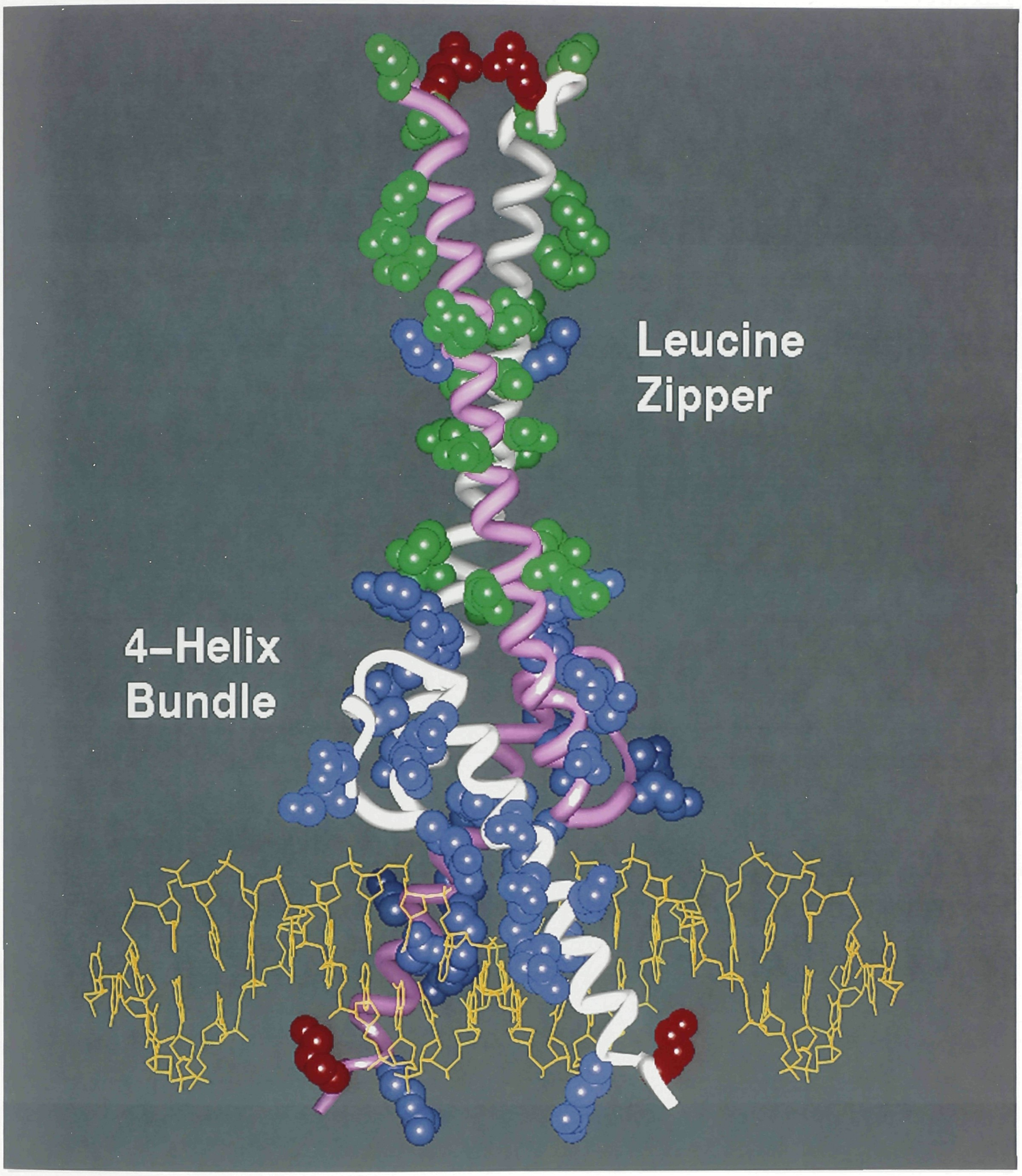
summarized in Figure 4.11. The diagrams depicted in these figures show the observed sites of cleavage which lead to fragments *that maintain DNA binding.* The Lys-C and chymotrypsin results can serve as examples for elucidation. Over a period of three days, Lys-C proteolysis of the Max-DNA complex leads to complete as well as partial digestion (data not shown). Partial proteolysis results in three high-mass peptides: 25-77, 25-104, and 25-113. These fragments show significant resistance to digestion, suggesting that they remain bound to DNA. These findings are summarized on the Lys-C diagram of Figure 4.11: the solid arrow at the N-terminal signifies complete cleavage (within 12 hours) at Lys 24; the open arrows at the leucine zipper and C-terminal regions signify incomplete cleavage (after one day) at Lys 77 and 104. In contrast, the chymotrypsin digest of the Max-DNA complex did not produce any long-lived partial proteolytic fragments. After one day of chymotrypsin treatment of Max, only undigested protein and a modest amount of completely digested Max was observed (data not shown). Thus, no arrows are drawn in the chymotrypsin diagram depicted in Figure 4.11.

4.4.7 The Proteolytic Protection map

The current findings can be presented in the context of the X-ray crystal structure of b/HLH/Z Max-DNA. Figure 4.12 shows a colored version of the ribbon diagram (Ferré-D'Amaré et al., 1993) shown earlier (Figure 4.3) superimposed with results from the Max proteolysis experiments. Amino acid residues that are targets of cleavage (by the six endoproteases used in the

Figure 4.12 *Proteolytic protection map of the b/HLH/Z Max-DNA complex*

The proteolytic protection map is a depiction of the combined results of analyses by MALDI-MS of a large number of proteolytic digestions performed on Max (in the presence of Max-specific DNA) with the endoproteases V8, Lys-C, Asp-N, chymotrypsin, trypsin and subtilisin. Using a color-coded measure of protection against proteolysis, the figure was created by superimposing the MALDI-MS proteolysis results onto a ribbon diagram of the X-ray crystal structure of the Max-DNA complex (Ferré-D'Amaré et al., 1993). The peptide backbone tracing the two Max monomers are outlined as white and pink ribbons. The N-termini (residue 22) are at the bottom, the C-termini (residue 107) are at the top. The coordinates of the last six residues (108-113) were not available because the X-ray crystallographic electron density map indicated that these residues were random coil (they would be colored red in the diagram). Amino acid residues that are proteolytic targets have their sidechains highlighted as colored spheres in a space-filling representation. (All highlighted sites were shown to rapidly cleave in the absence of DNA). The three colors reflect a relative measure of protection against partial proteolysis at the highlighted residue. Blue represents the sidechains that dramatically resist proteolysis over a 2-4 day period. Green indicates less protection where complete or partial proteolysis at these sites has occurred over this time. Red represents the lowest level of protection, where complete cleavage at these sites occurred in less than one day. Protection from proteolytic cleavage at the 'blue' sites together with cleavage at either a 'green' or 'red' site results in a truncated form of the protein which remains bound to DNA.


Leucine
Zipper
4–Helix
Bundle

study) are depicted as colored spheres in a space-filling representation; for clarity, other residues are omitted (see figure caption for complete description.) The colors correspond to the three levels of protection against *partial* proteolysis of the Max-DNA complex that allows for the maintenance of DNA binding. Blue represents the greatest degree of protection, where no partial proteolysis was observed over two to four days of digestion. Green indicates a lesser degree of protection, where complete or partial proteolysis at these sites has occurred over this time. Red represents the least amount of protection, where complete cleavage at these residues has occurred in less than one day.

4.5 DISCUSSION

4.5.1 Max in the Absence of DNA

Under conditions of low salt (50 mM buffer, 15 mM KCl, 1 mM $MgCl_2$), Max exhibits little resistance to digestion by six different endoproteases. The rates of digestion of Max are similar to those usually observed for denatured proteins or peptides. With the use of V8 protease, for example, the digestion occurs rapidly and indiscriminately at all five glutamate residues of Max and is nearly complete in one hour (Figure 4.5a-c). These findings demonstrate that at low ionic strengths, Max has an open, flexible structure, one that has insufficient higher-order folding and/or rigidity to resist rapid proteolytic digestion. Although

these results do not allow us to conclude that the protein is in an oligomeric form other than a monomer, they suggest that any self-association is weak at low ionic strengths.

Elevation of the ionic strength has pronounced effects on the V8 digestion of Max (Figure 4.8b), although the effects with the other proteases were less dramatic. These distinctions may be related to the differences in the activities and the modes of action of the proteases. At 150 mM KCl the rate of V8 proteolysis of Max drops by almost an order of magnitude compared to observations under low ionic strengths. Two factors may account for this reduction of proteolytic rates. Elevated chloride ion has previously been observed to inhibit V8 activity (Sorensen et al., 1991) and is predicted to give an approximate two-fold decrease in V8 activity in going from 15 mM to 150 mM KCl. The second factor is based on the notion that elevated ionic strengths may induce a structural change of the protein; evidence for this is suggested by the observed changes in the pattern of cleavage by V8 at elevated salt levels. The cleavage pattern indicates three regions of proteolytic protection that decrease in the order four-helix bundle > leucine zipper >> N-terminal region. By extrapolating from the X-ray crystal structure of the Max-DNA complex, this order of proteolytic protection appears to correlate to the extent of hydrophobic interactions found within each region. The greatest number of hydrophobic interactions are found within the four-helix bundle, less in the leucine zipper and fewest in the N-terminal, which, in the absence of DNA, is expected to be a random coil (Ferré-D'Amaré et al., 1993). These correlations of proteolytic

protection with hydrophobic interactions indicate that physiological salt levels favor a dimeric form of Max. Because dimerization occurs mainly through hydrophobic interactions found within the four-helix bundle and leucine zipper domains, an elevation of the salt level is predicted to stabilize the dimer.

4.5.2 Max in the Presence of DNA

My findings are summarized as follows. (1) In the presence of DNA, both complete and partial proteolysis of Max were observed, the extent of which was protease dependent. The fragmentation patterns of partial proteolysis were the most informative for the studies. (2) In the presence of DNA, striking changes were observed in the rates and patterns of proteolysis of Max. The effective rates of proteolysis were greatly reduced —by up to a hundred fold— compared to those seen in the absence of DNA. The steep drop in rates indicates that in the presence of DNA, Max undergoes structural changes which correlate to a significant reduction of solvent accessibility and/or flexibility of the protein. (3) With each of the proteases used, the patterns of proteolysis were dependent on the nature of DNA present in solution. In the presence of sDNA, the N-terminal of Max remains highly protected against proteolysis, whereas in the presence of nsDNA the N-terminal is rapidly cleaved. This behavior is shown, for example, in the V8 digestion of Max (compare Figures 4.6a-c and 4.7 with 4.8a). Similar findings are also observed with the Lys-C and trypsin digestions of Max in the presence of nsDNA (data not shown). I can infer from the proteolytic protection

assay that Max binds at its N-terminal to E-box DNA with high specificity and affinity, strengthening the X-ray crystallographic findings and earlier biochemical studies (Ferré-D'Amaré et al., 1993 and references cited therein). Conversely, in the presence of nsDNA, the N-terminal remains highly susceptible to proteolysis, an indication that it is flexible and accessible to proteolytic attack This finding is in agreement with circular dichroism measurements which suggest that the N-terminal of b/HLH/Z proteins is a random coil in the presence of nsDNA (Ferré-D'Amaré, 1994; Ferré-D'Amaré et al., 1994). (4) The distinction between the patterns of proteolysis of Max in the presence of sDNA and nsDNA supports an 'induced fit' model for DNA binding (Spolar and Record, Jr. 1994). For Max the induced fit model predicts that site-specific DNA binding would proceed by a random-coil-to-helix folding transition of the N-terminal of the protein. The folding transition occurs only in the presence of DNA containing the cognate E-box recognition site, as evidenced by the rapid cleavage of the N-terminal of Max in the presence of nsDNA. (5) The results of the proteolytic protection assay suggest that the binding of Max to E-box DNA is accompanied by an extensive rigidification of the protein throughout its entire length. Proteolytic susceptibility is greatest only towards the extreme N- and C-termini (summarized in Figures 4.11 and 4.12). The MALDI-MS peptide mapping results indicate that there are three distinct regions of proteolytic protection in the presence of Max-specific DNA. These are the N-terminal, the central and the C-terminal regions.

4.5.2.1 *The N-terminal Region*

The N-terminal DNA-binding region (residues 24-36) consistently exhibited the greatest protection from proteolysis in the presence of sDNA. This observation is in agreement with the finding that the N-terminal adopts a rigid helical conformation that makes a large number of specific contacts with the E-box consensus site (Ferré-D'Amaré et al., 1993). The proteolytic protection assay also addresses a specific structural question concerning Lys 24. X-ray crystallography of the Max-DNA complex indicated that because of crystal packing, the ε-NH_2 of Lys 24 was interacting with a *neighboring* DNA molecule (Ferré-D'Amaré et al., 1993). However, it could not be determined whether Lys 24 was making contact with its *bound* DNA in the absence of crystal packing. The proteolysis results show that Lys 24 is rapidly cleaved by Endo Lys-C. The resulting shortened protein (i.e., fragment 25-113) remains highly resistant to further Lys-C cleavage, indicating that DNA binding is maintained. Thus, the results have shown that Lys 24 is not required to *preserve* DNA binding. I did not, however, determine whether Lys 24 was needed to *establish* DNA binding.

4.5.2.2 *The Central Region*

The second level of proteolytic protection is the region of Max that correlates to the four-helix bundle. The results are consistent with the notion that the four-helix bundle is a highly compact, globular domain. The three conserved

aromatic residues of the bundle (Phe 43, Tyr 70, Tyr 74) were highly protected against chymotrypsin cleavage, reflecting the inaccessibility of these residues to proteolytic attack. All of the Lys, Arg, Glu and Asp residues of Helix I and II were also highly protected against proteolytic cleavage. The eight-residue loop of the bundle showed a remarkable resistance to cleavage at Glu 56 (by V8) and Lys 57 (by Endo Lys-C and trypsin), and especially towards cleavage by subtilisin, a highly active protease with broad specificity. These results suggest that despite its solvent accessibility, the loop maintains a rigid structure in solution, in agreement with the X-ray crystallographic finding that the peptide backbone of the loop is well defined in the electron density map of the Max-DNA complex (Ferré-D'Amaré et al., 1993).

4.5.2.3 The C-terminal Region

The lowest level of proteolytic protection of the Max-DNA complex was associated with the C-terminal region. Three conclusions can be deduced from the observed rates and patterns of proteolytic cleavage of this region. First, the complex showed a high susceptibility to cleavage from residues 107-113 (by subtilisin and trypsin), demonstrating the presence of a stretch of flexible polypeptide at the C-terminal. This result is consistent with the X-ray crystallographic finding that the C-terminal six residues were not visible in the electron density map of the Max-DNA complex (Ferré-D'Amaré et al., 1993). Second, the observed proteolytic 'unzipping' of the complex is characteristic of a leucine zipper motif. [A similar pattern of proteolysis was observed to occur in

another b/HLH/Z protein, SREBP1 (sterol regulatory element binding protein 1); discussed later]. The sequential and gradual cleavage of the zipper can be understood in terms of its structure. The zipper consists of a coiled-coil arrangement of two amphipathic helices, resulting in a rigid structure that is stabilized by both hydrophobic and polar interactions between the two helices (Landschulz et al., 1988; O'Shea et al., 1989). The most accessible part of the zipper is its C-terminus where the two helices can more readily unfold and yield to proteolysis. Based on previous studies of leucine zipper domains, a strong polar interaction is predicted to occur between Lys 89 of one helix and Asp 84 of the other helix (and vice versa) (Baxevanis and Vinson, 1993). The results show a high proteolytic protection of Lys 89 (against Lys-C and trypsin) and Asp 84 (against Asp-N), findings that may reflect the predicted interhelical electrostatic interaction between these two sidechains. The 'stalling' of proteolysis at residue 77 (by subtilisin) or at residue 75 (by trypsin) demonstrates that the region N-terminal to the zipper (the four-helix bundle) greatly resists proteolysis. This is the first report of an 'unzipping and stalling' pattern of proteolysis of a leucine zipper. A third conclusion regarding the C-terminal region concerns the role of the leucine zipper on stabilization of the Max-DNA interaction. Partial or complete removal of residues of the zipper by proteolysis does not eliminate DNA binding. This result indicates that the four-helix bundle is a highly stable structure that is sufficient to maintain DNA binding without the leucine zipper and supports the notion that the leucine zipper may be more important in defining dimerization specificity than it is in stabilizing the Max-DNA complex (Reddy et

al., 1992). Finally, a note about the effects of higher-order oligomerization of Max dimers. Tetramerization of other b/HLH/Z proteins has been observed and is believed to occur through an interaction between the leucine zipper domains of b/HLH/Z dimers (Ferré-D'Amaré et al., 1994) and references cited therein). The proteolysis results are inconclusive concerning the tetramerization of Max.

4.5.3 Evaluating the Proteolytic Protection Assay

Several factors that impact upon the proteolytic protection assay are discussed below. These include specific vs. non-specific protein-DNA interactions, dynamics of the Max-DNA complex, and the use of MALDI-MS in proteolytic peptide mapping.

4.5.3.1 Specific vs. Non-specific Protein-DNA Interactions

The results of this study indicate that the interactions between Max and sDNA lead to a significant protection against proteolysis of the Max dimer. Interactions between protein and DNA are usually categorized as either specific or non-specific (Record, Jr. et al., 1991). In the case of Max, there are highly specific interactions between the DNA-binding region of the protein and the oligonucleotide bases lining the major groove of the E-box recognition sequence (Ferré-D'Amaré et al., 1993). The dramatic increase in proteolytic protection of Max upon binding to DNA is an indication that specific protein-DNA interactions lead to a rigidification of the dimeric structure of the protein. There can also be weaker, non-specific interactions that occur between positively charged groups

of Max and negatively charged phosphate groups of DNA. These non-specific protein-DNA interactions could shield the protein against proteolysis, complicating the simple correlation of proteolytic protection with structure. Since non-specific protein-DNA interactions are known to weaken with increasing cation concentrations (Record, Jr. et al., 1991), several proteolysis experiments were performed at elevated ionic strengths. However, earlier experiments in this study have shown that raising the ionic strengths also leads to a tightening of the Max dimerization, masking any reduction in shielding that arises from non-specific protein-DNA interactions. A compelling reason to suggest that the shielding does not greatly influence the patterns of proteolytic protection derives from experiments involving the proteolysis of Max carried out in the presence of nsDNA. Despite the high potential for shielding to occur between nsDNA and the highly basic N-terminal, this region of Max is rapidly and preferentially cleaved by either V8 (refer to Figure 4.8a), Lys-C or trypsin. These findings suggest that non-specific protein-DNA interactions do not inhibit proteolytic activity in the basic N-terminal region, and by inference most likely across the rest of the protein.

4.5.3.2 *Dynamics of the Max-DNA Complex*

If Max-DNA were a static complex, proteolytic fragments could only arise from digestion of the intact complex and there would be no ambiguity as to the origin of the fragments. However, protein-DNA complexes are known to be in dynamic equilibrium with unbound protein and DNA, so that proteolysis can

occur from both the bound and unbound protein —leading to ambiguities in the interpretation of the data. This difficulty was minimized by maintaining an initial 50% molar excess of unbound DNA over Max-DNA complex in all of the experiments. An excess of unbound DNA ensured that virtually all of the Max molecules are complexed to DNA. (A simple estimation of the % protein bound is given in Figure 4.13). At the same time, a 13-fold molar excess of total DNA over protease was maintained. This ensured that upon dissociation of the complex, Max is considerably more likely to encounter and reassociate with unbound DNA than it is to encounter protease. The success of this approach was confirmed by the long apparent lifetime (days) of the Max-DNA complex observed in the proteolysis experiments.

4.5.3.3 *MALDI-MS is Well Suited for the Proteolytic Protection Assay*

This study has demonstrated that MALDI-MS has adequate speed and mass accuracy to analyze hundreds of complex digest mixtures 'on the fly'. A comparable task handled by methods such as SDS-PAGE, liquid chromatography and peptide sequencing would be tedious and prohibitively slow. As an alternative to the MALDI-MS technique, proteolytic digestions can be analyzed by electrospray ionization mass spectrometry. For example, limited proteolysis combined with ESI-MS was successfully applied to the analysis of structural changes in various forms of calmodulins upon calcium binding (Brockerhoff et al., 1992). A recent proteolysis study undertaken in the Chait lab of the HIV-1 protein Nef complexed to SH3 protein also demonstrated the

<u>Under equilibrium conditions</u>

$$(Max)_2 \; + \; sDNA\text{ (duplex)} \rightleftharpoons \{(Max)_2 \bullet sDNA\text{ (duplex)}\} \qquad \textbf{(4.1)}$$

$$K_D = \frac{[(Max)_2][sDNA)]}{[(Max)_2 \bullet sDNA]} \qquad \textbf{(4.2)}$$

An estimate of the percentage of Max dimer bound to DNA at equilibrium can be calculated using the experimental initial concentrations of Max dimer and DNA (15 μM and 22.5 μM, respectively) and assuming a K_D of approximately one nanomolar (Sha et al., 1995).

<u>Initial Concentrations</u>	<u>Equilibrium Concentrations</u>
$(Max)_2 = 1.50 \times 10^{-5}$ M	$(Max)_2 = 1.50 \times 10^{-5}$ M - B
$sDNA$ (duplex) = 2.25×10^{-5} M	$sDNA$ (duplex) = 2.25×10^{-5} M - B
$K_D \cong 1 \times 10^{-9}$ M	B = *conc. of the Max-DNA complex*

From Eq. 2, $$K_D \cong 10^{-9}\text{ M} = \frac{\left(1.50 \times 10^{-5} - B\right)\left(2.25 \times 10^{-5} - B\right)}{B}$$

Solving for *B*, $$B = \left[(Max)_2\right]_{Bound} = 1.4998 \times 10^{-5}\text{ M}$$

$$\%\ Bound = \frac{\left[(Max)_2\right]_{Bound}}{\left[(Max)_2\right]_{Initial}} = \frac{1.4998 \times 10^{-5}\text{M}}{1.500 \times 10^{-5}\text{M}} \times 100\ \% = \underline{99.987\ \%}$$

<u>Figure 4.13</u> *Calculation of percent Max dimer bound to DNA.*

Calculation of the percentage of Max dimer bound to Max-specific DNA (sDNA) under solution equilibrium conditions. The association between Max dimer and sDNA is given by equation 4.1. The corresponding dissociation constant (K_D) relating equilibrium concentrations is given by equation 4.2. The calculation shows that under equilibrium conditions a very small percentage (0.013 %) of Max dimer is not bound to DNA.

feasibility of utilizing ESI-MS (Lee et al., 1996). However, the electrospray technique requires that samples for peptide mapping be free of involatile buffers and salts, necessitating thorough desalting. In addition, successful analysis of complex mixtures by electrospray often requires chromatographic separation of the mixture. In contrast, MALDI-MS is highly tolerant of buffers, salts and other biochemical additives and complex mixtures can be directly analyzed from digestion solutions.

Finally, a note regarding the quantitative aspects of the mass spectral data. The proteolytic protection assay is based on correlating proteolytic fragments to protein structure. The identities of the proteolytic fragments were determined from their masses. The progress of the proteolytic digestions was determined by a semi-quantitative evaluation of the relative heights of the fragment ion peaks. Although accurate quantitation of peptides and proteins by MALDI-MS can be difficult, dependence on absolute quantitation was lessened by relying on a comparative analysis of the relative rates that proteolytic fragments appear and disappear (as mass spectral ion peaks) during the course of a digestion.

4.6 Conclusions of the Max Study

A simple biochemical method that combines enzymatic proteolysis and MALDI mass spectrometry was developed to probe aspects of the solution structures of DNA-binding proteins (Cohen et al, 1995). The method is based on

the notion that structural information can be inferred from a determination of protection against enzymatic proteolysis (as governed by solvent accessibility and protein flexibility). The procedure was applied to a study of the transcription factor Max. In the absence of DNA and at low ionic strengths, Max is rapidly digested by each of six endoproteases selected for the study, results consistent with an open and flexible structure of the protein. At physiological salt levels, the rates of digestion are moderately slowed. This reduction in rates together with the patterns of cleavage are consistent with homodimerization of Max through a predominantly hydrophobic interface. In the presence of Max-specific DNA, the protein becomes dramatically protected against proteolysis, exhibiting up to a hundred-fold reduction in cleavage rates, compared to the rates observed in the absence of DNA. The fragmentation patterns of proteolysis correlate well with those expected of a stable dimer bound to DNA at its basic N-terminals, in good agreement with the Max-DNA structure previously determined by X-ray crystallography. Protection is greatest in the N-terminal and helix-loop-helix regions of the protein, corresponding to the DNA-binding region and the globular four-helix bundle domain, respectively. An intermediate level of protection is found within the leucine zipper domain. The leucine zipper is distinguished by undergoing a sequential proteolytic cleavage that starts from the C-terminal of the zipper and terminates at the protease-resistant four-helix bundle. Proteolysis also indicates a high affinity of Max for its target DNA which remains high even when the leucine zipper is proteolytically removed. Proteolytic protection is least near the N- and C-termini indicating that these are the most flexible and

accessible regions of the protein. The present biochemical results confirm and extend the X-ray crystallographic findings of the Max-DNA complex.

In addition to the study of the helix-loop-helix protein Max and its complex with DNA, the present method appears well suited for many other structural biological applications. These include the precise characterization of the boundaries of rigid protein domains (important for NMR and X-ray crystallography; see next chapter), the determination of binding specificity of DNA-binding proteins, and the characterization of conformational changes induced in proteins by covalent modifications such as phosphorylation (see next section).

4.7 Proteolysis to Structurally Probe Other HLH Proteins

4.7.1 SREBP1, USF, and Hairy

SREBP1, USF, and Hairy were structurally probed using proteolysis and mass spectrometry in a similar fashion as was done for Max. Each of the three proteins possess interesting features. SREBP1 has a non-palindromic 10 bp DNA recognition sequence (the SRE) instead of the canonical 6 bp E-box sequence (see Materials section). The SREBP1 loop length is seven residues, compared with the Max loop length of eight. For USF, the hydrophobic heptad repeat of the leucine zipper goes 'out-of register' in the middle of the zipper; there is evidence that DNA-bound USF dimers form tetramers that may promote DNA looping (Ferré-D'Amaré et al., 1994); also, USF has a longer loop than Max (12 vs. 8 residues). Finally, Hairy, a small b/HLH protein, has a helix-breaking proline in the DNA-binding basic region. Hairy has an eleven residue loop. My goal was to proteolytically probe the structures of these three HLH proteins. The results are given below.

4.7.1.1 SREBP1-Sterol Regulatory Element Binding Protein 1

The b/HLH/Z SREBP1-SRE DNA complex yielded to subtilisin and endo Lys-C digestions in a fashion resembling b/HLH/Z Max proteolysis using these enzymes. I will present and discuss the subtilisin results; identical arguments can be made for the trypsin digestions. With subtilisin, proteolysis of the b/HLH/Z SREBP1-SRE DNA complex, occurred down the leucine zipper (N-terminal direction), initially resulting in the rapid removal of the first seven C-terminal residues (by ~1 hr), then in a more gradual, sequential removal of residues of the leucine zipper (in the N-terminal direction) so that by 1 day, the digest stalled near the bottom of the zipper at residues 380 and 384. This puts the maximum number of residues between the 'stalled' C-terminal and the top of the four helix bundle (residue 374) at ten. For Max the number of residues was also ten. Thus, inasmuch as 'gross' structural features can be inferred from proteolysis, b/HLH/Z SREBP1 and Max (bound to DNA) are very similar (i.e., in terms of rigidity and solvent accessibility), especially in the HLH and leucine zipper regions. The 'lifetimes' of the two complexes in solution (based on the MALDI-MS signal of the subtilisin digest fragments) are also comparable. However, any structural differences between the two protein-DNA complexes in their DNA-binding regions cannot be distinguished by proteolysis. Thus although the DNA recognition elements differ in both sequence and length (SRE vs. E-box), the proteolysis results strongly suggest that the b/HLH/Z forms of SREBP1 and Max (bound to DNA) are structurally similar. (An early report of the crystal structure of b/HLH/Z SREBP1-SRE-

DNA complex (Antonio Parraga/Burely) indicates that the SREBP1-SRE X-ray structure closes resembles that of the b/HLH/Z Max-E-Box DNA complex.)

4.7.1.2 USF-Upstream Stimulatory Factor

The MALDI-MS mapping of the subtilisin digestion of b/HLH/Z USF revealed rapid (5 min) cleavage of ~3-4 residues from the C-terminal of USF, similar to the findings for Max (Figure 4.10a) and SREBP1, discussed previously. However, for extended subtilisin digest times (>1 hr) the MALDI-MS signal intensity from high mass fragments for the USF-DNA complex rapidly diminished. By four hours of digest there were several groups of peaks in the m/z 6000-8000 range, but they were too weak to obtain an accurate mass for identification. I interpreted these findings as follows. USF protein yields to subtilisin digestion more quickly than either Max or SREBP1. One reason for the greater susceptibility to digestion may be related to the long loop in USF. USF has four more residues in its loop than Max and five more than SREBP1 (see Figure 4.2). Unlike Max or SREBP1, the long loop in USF may be 'extended' enough for suitable subtilisin attack, whereas the shorter loops in Max and SREBP1 are too rigid and/or inaccessible to the subtilisin active site.

4.7.1.3 Hairy

The b/HLH protein Hairy was much more susceptible to proteolysis than was b/HLH/Z Max (when bound to DNA). For example, endo Glu-C (V8)

digest of either Hairy or Max in the absence of DNA was essentially complete in less than 15 minutes. In the presence of 14 mer E-box DNA, the digestion rate of Hairy dropped by a factor of ~10, while for Max the reduction was estimated to be a factor of ~100. The greater susceptibility of Hairy (bound to DNA) to proteolysis compared to Max may arise from several factors. The equilibrium binding constant (K_D) for b/HLH proteins is in the order of 10^{-6} M whereas for b/HLH/Z proteins it is ~10^{-9} M (Sha et al., 1995). Using these estimations for K_D, the percent bound calculated by Eq. 4.2 (see Figure 4.13) is 90% for Hairy and 99.99% for Max. Thus, there would be a greater percentage of b/HLH than b/HLH/Z proteins in the unbound state (i.e., more exposure to proteolytic degradation). In addition, Hairy has a longer loop than Max (11 vs. 8 residues), a feature that may also contribute additional susceptibility of Hairy to proteolysis.

The b/HLH Hairy proteolysis results raises an interesting issue since they run counter to the earlier findings with b/HLH/Z Max. Slow proteolytic removal (48 hr) of the leucine zipper of Max (see Max study) resulted in a stable, proteolytically resistant 'b/HLH' form of Max (e.g., Max 12-77 in Figure 4.10c). It is not clear why 'b/HLH' Max has what appears to be a much longer lifetime than b/HLH Hairy. b/HLH Max may be intrinsically more stable than b/HLH Hairy (smaller K_D). The longer loop in Hairy (11 residues) may also play a role in increasing the susceptibility of proteolytic attack compared to Max which has a shorter loop (8 residues).

4.8 Proteolysis to Structurally Probe Phosphorylated Max

Protein phosphorylation is a powerful means of exerting regulatory control over eukaryotic cellular events. The ubiquitous phenomenon of phosphorylation is observed from signal transduction to cell cycle mechanisms. Nuclear proteins are frequently targets for phosphorylation by specific cellular kinases. One example is the AP1 family of transcription factors, which includes the proto-oncoproteins Fos and Jun. These two proteins are known to form homo- and heterodimers with each other. Elevated expression of Fos and Jun in stimulated cells has been implicated in the activation of transcription of growth-promoting genes. Fos and Jun have multiple phosphorylation sites, each of which has been shown exert distinct control over the DNA-binding and *trans*-activation activities of the two proteins (Abate et al., 1993).

Max, like Fos and Jun, is a DNA binding protein whose activity depends on oligomerization. Inspection of the N-terminal sequence of Max (Figure 4.14) reveals consensus sites for casein kinase II (CKII) phosphorylation (Meisner and Czech, 1991; Berberich and Cole, 1992; Prendergast et al., 1992). The 4-residue consensus sequence for CKII phosphorylation (of serine or threonine) is S/TxxE/D, where x is any amino acid (Meisner and Czech,

1 5 10 15 20

Max......M **S** D N D D I E V E **S** D E E Q P R F Q **S** A A D K....

2 5 10 (22)

MaxΔC........**S** D N D D I E V E **S** D A D K....

331 335 340 345 350

Myc.....R V L R Q I S N N R K C T S P R S **S** D T E E N V....

529 535 540 545 550

E12......T **S** P D E D E D D L L P P E Q K A E R E K E R R....

Figure 4.14 *Casein kinase II consensus sites in HLH proteins.*

Casein kinase II consensus sites for serine phosphorylation in regions N-terminally adjacent to the b/HLH/Z domains of Max, MaxΔC and Myc and to the b/HLH domain of E12. The consensus sequence for CKII is S/TxxE/D. Serine residues that fit the CKII consensus are enlarged and shaded in bold. The N-termini of the basic DNA-binding regions are doubly underlined. The 9-residue alternatively spliced insert sequence in Max is singly underlined. The form of Max investigated in this study is designated MaxΔC (Max 2-12/22-113) which does not include the insert.

1991). Figure 4.14 shows the amino acid sequences immediately adjacent to the N-terminal of the basic regions of the b/HLH/Z proteins Max, MaxΔC and Myc, and the b/HLH protein E12. Here, Max refers to the long form (i.e., includes the nine-amino acid alternatively spliced insert, underlined in Figure 4.14) that has three possible CKII sites. The shorter form, termed MaxΔC (excludes the insert) shows 2 CKII sites. Myc and E12 have one CKII site each.

Interestingly, the CKII sites are flanked by a rather high number of acidic residues in Max, MaxΔC, and E12 but not Myc. In the Max N-terminal sequence, 9 out of 23 (9/23$\approx$0.4) residues are acidic; for MaxΔC the ratio is 7/13 (0.5), for E12, 10/23 (0.4) and for Myc, 3/23 (0.1). Upon phosphorylation by CKII, the ratios increase further because of the introduction of an acidic phosphate group. The presence of the acidic residues may, in part, assist kinase targeting, binding and phosphorylation. The acidic residues may also play a structural role affecting DNA binding of the HLH dimer complex. An earlier study reported that E12 preferentially binds DNA as a heterodimer with other b/HLH proteins, such as E47 and MyoD, rather than with itself as an E12 homodimer (Sun and Baltimore, 1991). In addition, unlike E12, E47 was found to bind DNA as a homodimer. E47 has a lower number of acidic residues in the region adjacent to the b/HLH domain compared to E12 (7 out of 23 residues). The study showed that the 'acidic region' of E12 (residues 529-550) conferred heterodimerization specificity. Swapping the two

N-terminal regions in E12 and E47 potentiated homodimerization of E12. It was concluded that the high number of acidic residues adjacent to the (then putative) DNA binding domain negatively influence DNA binding of E12 homodimers (Sun and Baltimore, 1991).

A comparable argument was offered for Max, where mutation of the two CKII serine sites in Max∆C (Ser2 & 11) to non-polar residues enhanced DNA binding (Prendergast et al., 1992) relative to the wild-type protein. In this study, a model was proposed suggesting that phosphorylation may be weakening DNA binding through electrostatic repulsion between phosphate residues on the protein and the DNA backbone. It was also argued that DNA-binding could be compromised by an interaction between the highly negatively charged acidic region and the positively charged DNA-binding (basic) region (Sun and Baltimore, 1991; Prendergast et al., 1992).

The high resolution X-ray study of b/HLH/Z Max (Max 22-113)-DNA complex provided a clue into the structure of the N-terminal acidic region. Initial crystallization trials of Max were with Max∆C (Max 1-113, minus the 9-residue insert). No crystals of Max∆C were obtained. Dynamic light scattering experiments indicated that solutions of Max∆C were polydisperse, a finding that was attributed to extensive aggregation of the protein (Ferré-D'Amaré and Burley, 1997). Dynamic light scattering determines whether the protein is in solution as oligomers (i.e., monomer, dimer, etc.) of the same size (monodisperse) or as randomly aggregated species (polydisperse). Removal

of the N-terminal acidic region yielded a construct of the protein (i.e., b/HLH/Z Max, residues 22-113) that was monodisperse and ultimately crystallizable. The dynamic light scattering results strongly suggest that the acidic N-terminal of Max (bound to DNA) is unstructured (i.e., a random coil), a condition that would favor aggregation at elevated protein concentrations. Thus, it is probable that no high resolution structure of MaxΔC (with its acidic N-terminal region) will be attained, either by X-ray crystallography or NMR.

In lieu of a high-resolution structure, I decided to apply the proteolysis/mass spectrometry protocols presented earlier in this chapter to probe MaxΔC. One of the questions that I addressed was whether proteolysis can probe how phosphorylation of MaxΔC effects its structure and DNA binding.

4.8.1 Results

Phosphorylation of MaxΔC by CKII was straightforward, taking ~2 hr to complete. MALDI-MS was useful for monitoring the progress of the phosphorylation. The molecular mass of the intact protein following treatment by CKII increased by 160 Da from the unmodified MaxΔC consistent with the addition of two phosphate groups to the protein. Peptide mapping using both MALDI- and ESI-MS confirmed that Ser2 and Ser11 were completely phosphorylated (data not shown). (Hereinafter, the doubly phosphorylated protein will be referred to as MaxΔC(+)P, whereas unphosphorylated protein

will be referred to as MaxΔC.) It is worth noting that N-terminal peptides derived from the N-terminal of MaxΔC are very acidic (see Fig. 4.14). Acidic peptides are notoriously difficult to analyze by MALDI (Juhasz and Biemann, 1994), even by negative-ion MALDI-MS. I found that acidic peptides that contained at least two strong basic residues (i.e., the V8 fragment MaxΔC2-32) were observed in both positive- and negative-ion MALDI using 4HCCA or DHB matrices dissolved in water/acetonitrile (data not shown).

4.8.1.1 Proteolysis of MaxΔC and MaxΔC(+)P

Four proteases were employed to probe MaxΔC: endo Lys-C, Glu-C (V8), Asp-N, and subtilisin. (1) Proteolysis performed in the absence of DNA showed that MaxΔC or MaxΔC(+)P was rapidly cleaved by all four enzymes, in a similar fashion to b/HLH/Z Max (see Fig. 4.5) (data not shown). (2) In the presence of non-specific DNA, the basic region of MaxΔC was quickly (< 1 hr) cleaved by any of the four proteases, whereas MaxΔC(+)P was cleaved rapidly by three of the four enzymes (see Figure 4.15). (3) In the presence of E-box DNA, the digest rates and patterns dramatically changed. Also, distinct differences in the patterns and rates of proteolysis between MaxΔC and MaxΔC(+)P were observed for some of the enzymes. A summary of these results is shown in Figure 4.16 and discussed below for each enzyme.

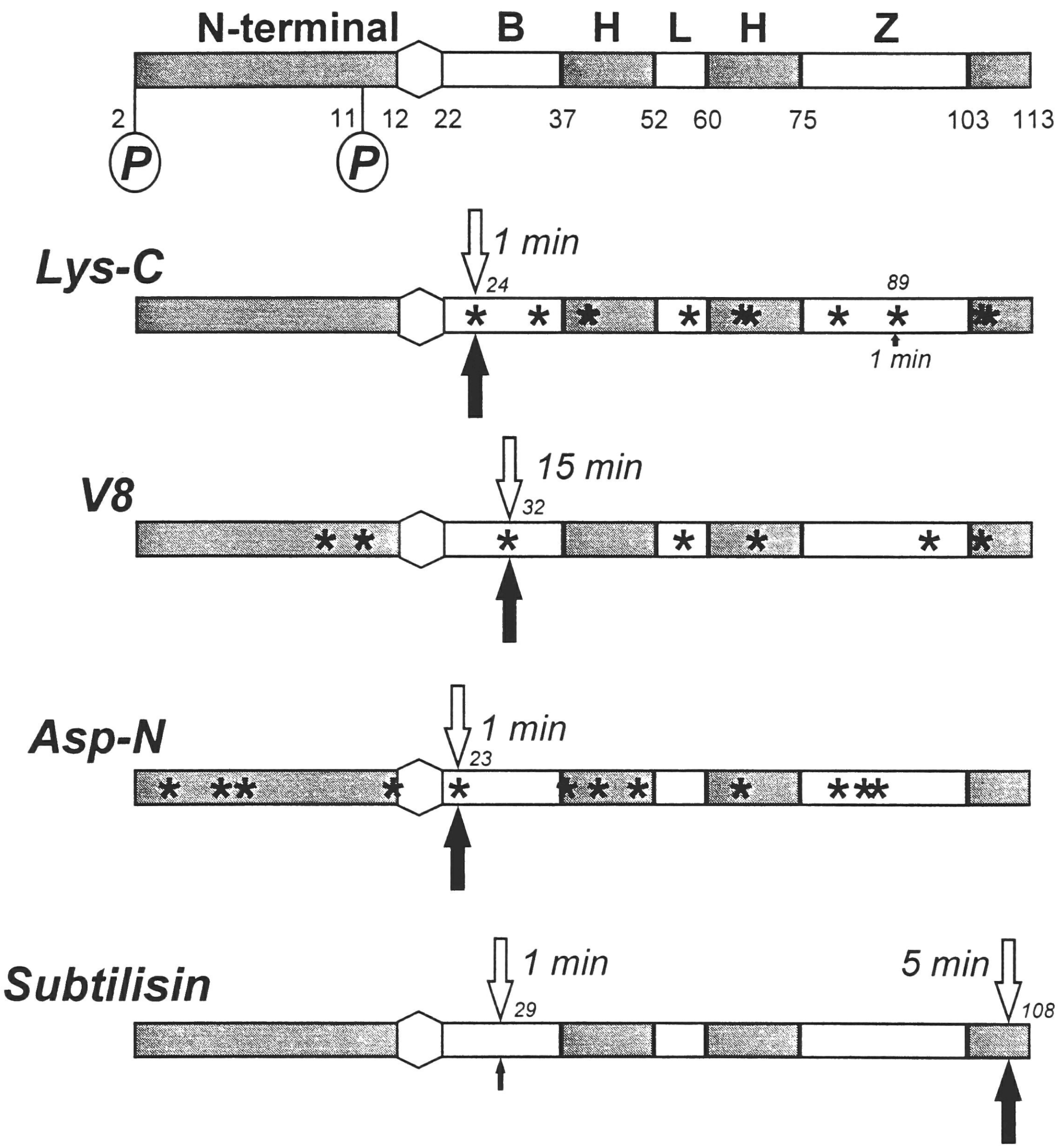


Figure 4.15 *MaxΔC + nsDNA proteolysis summary.*

TOP: Linear representation of MaxΔC (Max 2-12/22-113, not to scale). N-terminal length is exaggerated. CKII phosphorylation sites (Ser 2 & 11) are marked (P). **BOTTOM 4 DIAGRAMS:** Summary of MALDI-MS analyses of Lys-C, V8, Asp-N, and subtilisin digests in the presence of non-specific DNA. Vertical arrows designate sites (numbered) that are partially or completely cleaved in MaxΔC (hollow arrows drawn above each diagram) and MaxΔC(+)P (solid arrows drawn below each diagram). Large arrows denote extensive cleavage whereas smaller arrows indicate less as measured by MALDI-MS at the given times of digestion [times are same for both MaxΔC and MaxΔC(+)P]. Stars placed in the sequence indicate protease target residues for Lys-C, V8 & Asp-N proteases.

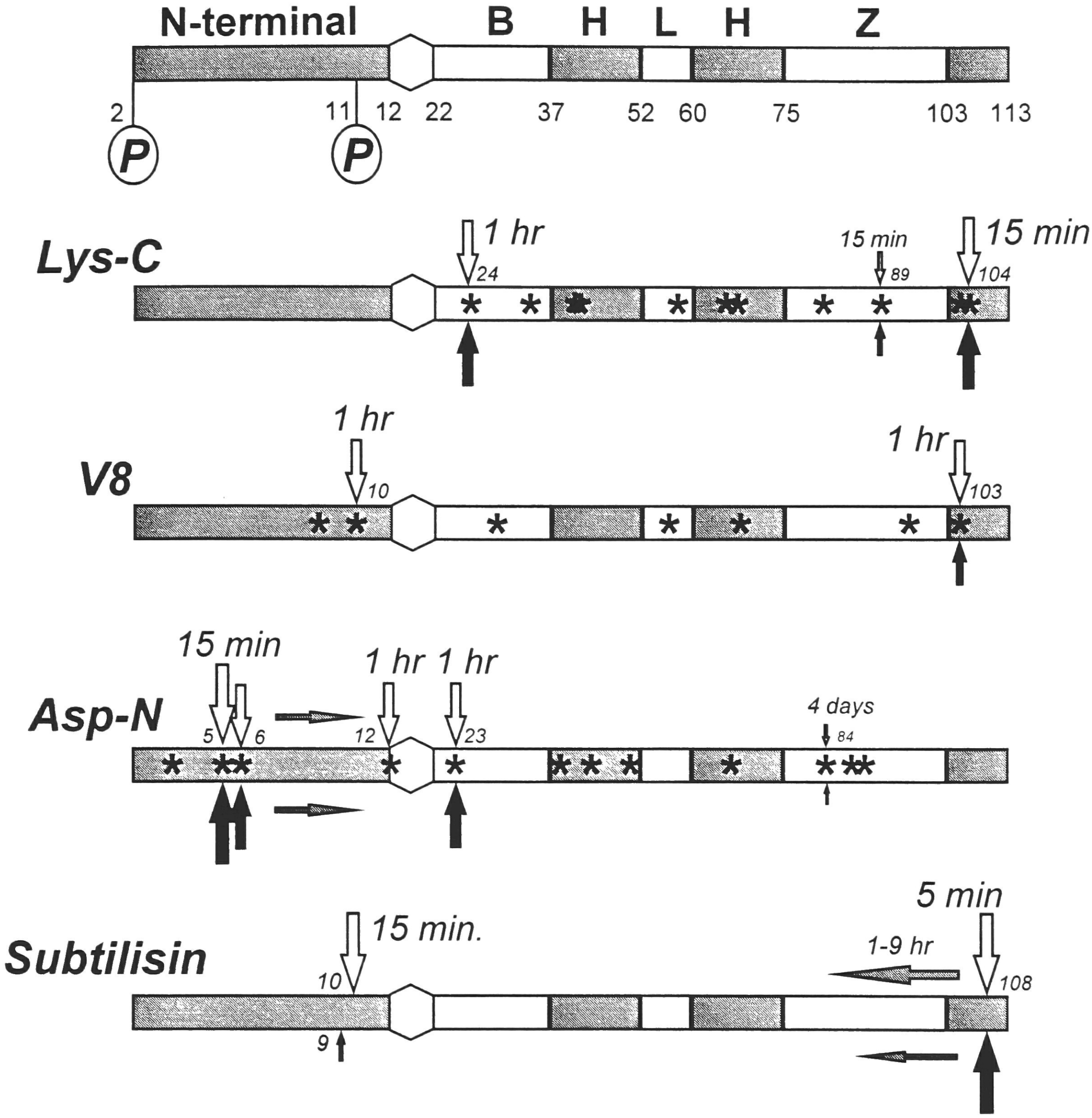


Figure 4.16 *MaxΔC + E-Box DNA proteolysis summary.*

TOP: Linear representation of MaxΔC (Max 2-12/22-113, not to scale). N-terminal length is exaggerated. Casein Kinase II phosphorylation sites (Ser 2 & 11) are marked (P).

BOTTOM 4 DIAGRAMS: Summary of MALDI-MS analyses of Lys-C, V8, Asp-N, and subtilisin digests of MaxΔC and MaxΔC(+)P performed in the presence of E-box (Max-specific) DNA. Horizontal arrows denote direction of digestion from either the N- or C-termini. See Figure 4.15 caption for full description of all other symbols & terms.

4.8.1.2 Lys-C Digestion of MaxΔC

In the presence of nsDNA, both MaxΔC and MaxΔC(+)P were rapidly (1 min) cleaved at Lys24. In the presence of E-box DNA, there was no significant difference between the patterns and rates of Lys-C digest of MaxΔC or MaxΔC(+)P. Cleavage by Lys-C initially occurred mainly at Lys104 (15 min) then gradually at Lys24 (1 hr), yielding the fragments 2-104 and 25-104. Interestingly, this pattern of digestion is reversed for the Lys-C proteolysis of b/HLH/Z Max —cleavage at Lys24 was the initial dominant site for Lys-C attack, whereas cleavage at Lys104 occurred much more slowly (data not shown).

4.8.1.3 Glu-C (V8) Digestion of MaxΔC

V8 digests of MaxΔC and MaxΔC(+)P in the presence of nsDNA were dominated by cleavage at Glu32, a conserved residue in the DNA-binding region. In comparison, V8 digestion performed in the presence of E-box DNA showed significant differences in the patterns and rates of proteolysis between MaxΔC and MaxΔC(+)P. Cleavages at Glu10 and Glu103 were prevalent for MaxΔC, yielding the fragments 2-103 and 11-113 (1hr) and 11-103 (3 hr). By contrast, although cleavage at Glu103 was present, cleavage at Glu10 was conspicuously absent for MaxΔC(+P). Glu10 is adjacent to phosphoserine11 in MaxΔC(+)P.

Careful inspection of the V8 digest data for the MaxΔC(+)P-DNA complex revealed proteolytic fragmentation at Glu32 (15 min). The resulting peptide fragment, Max 32-113, gave a weak but persistent MALDI-MS signal over a hour digest period. (Unphosphorylated Max did not show any significant cleavage at Glu32.) At 24 hr of V8 digest of MaxΔC, the N-terminal was entirely cleaved, and the C-terminal was partially cleaved, leaving the high-mass fragments 11-103 and 11-96. This situation was very similar to the 2 day V8 digest of b/HLH/Z Max (see Figure 4.7). In contrast, at 24 hr of V8 digest of MaxΔC(+)P, *the protein was completely digested* (no MALDI-MS peaks >6 kDa were observed).

4.7.1.4 Asp-N Digestion of MaxΔC

As in the Lys-C proteolysis, Asp-N cleavage of MaxΔC and MaxΔC(+)P essentially paralleled each other in both the patterns and rates of digestion. With either form of Max and in the presence of non-specific DNA, cleavage occurred rapidly (minutes) at Asp23 (Figure 4.15). In the presence of E-box DNA (Figure 4.16), proteolytic cleavage was much slower (hours) and sequential, starting at residue Asp5 and going through the N-terminal in a 'PacMan'-like fashion from Asp5 to 6 to 12 and 'stalling' at Asp23. By one day of Asp-N digestion, the fragment 24-103 dominated for MaxΔC and MaxΔC(+)P. Even after 4 days of incubation with Asp-N, the fragment 24-103 remained intense and there were the first hints of cleavage at Asp 84. The

only difference worth highlighting between the Asp-N digests of MaxΔC and MaxΔC(+)P was that phosphorylation of Ser11 greatly reduced Asp-N hydrolysis of the phosphoSer11-Asp12 bond.

4.7.1.5 Subtilisin Digestion of MaxΔC

Phosphorylation of MaxΔC had significant impact on the subtilisin proteolysis, both in the patterns and rates of digestion. In the presence of non-specific DNA (Figure 4.15) subtilisin digestion of MaxΔC occurred rapidly (minutes), first at Asn29 in the basic DNA-binding region, then at Ser108 in the C-terminal to leave the fragment 30-108, a polypeptide that is incapable of binding DNA. By one hour, the digestion of the whole protein was complete. For MaxΔC(+)P, digestion of the N-terminal was substantially less than for unphosphorylated Max. By one hour the dominant fragment was 2-108. It is unclear how phosphorylation of Ser2 and Ser11 would interfere with subtilisin cleavage of Asn29 (the target for unphosphorylated protein), particularly if the protein is not bound to DNA (i.e., nsDNA). Even at one hour, the 2-98 fragment remained, and complete digestion required >4 hr.

In the presence of E-box DNA, subtilisin digestion of MaxΔC resulted in the fragment 11-108 (1 hr), which essentially is the b/HLH/Z domain of Max (Figure 4.16). Like b/HLH/Z Max, any further subtilisin activity appeared as a sequential cleavage of the C-terminal residues (1-9 hr). In contrast, digestion of MaxΔC(+)P left the fragment 2-108 (1 hr), and no evidence for cleavage at

Glu10. Because Glu10 is adjacent to phosphorylated Ser11, subtilisin does not cleave peptide bonds adjacent to p1hosphorylated serines, as was observed for V8 and Asp-N. Over a longer period (1-4 hr), the phosphorylated N-terminal of MaxΔC slowly yielded to subtilisin cleavage at Glu8. By 9 hr phosphorylated Max was completely digested.

4.8.2 Discussion

There are several features that the proteolysis data reveals about MaxΔC and the effects of N-terminal phosphorylation.

4.8.2.1 Differences Between MaxΔC vs. b/HLH/Z Max

The heightened protection of Lys24 against Lys-C digest of MaxΔC may reflect an increase in the rigidity of the long polypeptide chain (Max 2-12/22-24, 14 residues) emerging, in the N-terminal direction, from the major groove of the E-box recognition sequence in MaxΔC vs. the much shorter chain (Max 22-24, 3 residues) for b/HLH/Z Max. Because Lys24 in b/HLH/Z Max is situated just three amino acids away from the N-terminal residue of b/HLH/Z Max (i.e., Ala22), the peptide chain around Lys24 may be more flexible and accessible by Lys-C in b/HLH/Z Max than in MaxΔC. In comparing Lys-C proteolysis of MaxΔC and MaxΔC(+)P, phosphorylation of MaxΔC did not

change the pattern of Lys-C digestion, probably because cleavage at Lys24, in removing the acidic N-terminal, resulted in identical forms of the protein (Max 25-104).

4.8.2.2 The N-terminal Region

In the presence of E-box DNA, additional protection against proteolysis of the N-terminal was observed with Asp-N protease in both MaxΔC and MaxΔC(+)P (Figure 4.16), compared to the digests performed in the absence of DNA. The gradual sequential cleavage through the N-terminal up to Asp23 (immediately adjacent to the DNA) suggests that the N-terminal acidic region is perhaps conformationally restricted or inflexible and/or the protease is sterically hindered against direct cleavage at Asp23 (as was the case for Asp-N cleavage in the presence of nsDNA; see Figure 4.15).

4.8.2.3 Effects of Substrate Phosphorylation on Protease Activity.

For three of the proteases it was shown that phosphorylation severely reduced (Asp-N) or completely abolished (V8 and subtilisin) hydrolysis of peptide bonds that shared the phosphorylated residue. It appears that the additional negative charge and/or bulk of the phosphate group significantly interferes with enzyme-substrate binding.

Phosphorylation inhibition of proteolysis was an intriguing finding. It is interesting to ask whether such inhibition could be used in nature for in vivo

control of protein processing For example, could a specific substrate be targeted for proteolytic attack following substrate dephosphorylation (by a phosphatase)? In this scenario, phosphorylation would act to inhibit protein processing. Although little has been published concerning the effects of phosphorylation and proteolysis, I found one paper that investigated the change in specificity of trypsin and a trypsin-like protease responding to substrate phosphorylation (Benore-Parsons et al., 1989). The study employed reverse-phase HPLC and radiolabeled peptides to perform the analysis. Using the heptapeptide substrate Leu-Arg-Arg-Ala-Ser-Leu-Gly, the authors showed that phosphorylation of Ser5 in the peptide (by pyruvate kinase) altered the kinetics and patterns of proteolysis. Both enzymes cleave between Arg3 and Ala4 on the unphosphorylated peptide. Phosphorylation of Ser5 dramatically reduced the rate of cleavage by trypsin (by almost 50-fold) and it redirected the site of proteolysis by the trypsin-like enzyme from Arg3 to Arg2. The authors discussed a possible influence of substrate phosphorylation on selectively influencing cellular protein processing and turnover events (Benore-Parsons et al., 1989).

4.8.2.4 Lifetime of the MaxΔC-DNA complex.

Some of the proteolysis data suggest that phosphorylation of MaxΔC reduces the 'lifetime' of the MaxΔC(+)P-DNA complex. The V8 digest is considered first. Phosphorylation of Ser11 prevented V8 cleavage at Glu 10

(Figure 4.16). The N-terminal of MaxΔC(+)P was thus left 'intact'. Discounting other loss mechanisms, the 'disappearance' of MaxΔC(+)P from the digest solution can only occur by proteolysis in the DNA-binding or HLH regions of the protein. The study of the b/HLH/Z Max-DNA complex (presented earlier in this chapter) clearly indicated that access to either the DNA-binding or HLH regions by protease requires the dissociation of the protein-DNA complex. For MaxΔC(+)P, indirect evidence for such dissociation is seen in the V8 digest of the MaxΔC(+)P-DNA complex. V8 fragmentation at Glu32 (in the DNA-binding region), observed for MaxΔC(+)P, but not MaxΔC, suggests that the protein-DNA complex is dissociating, permitting V8 hydrolysis at Glu32. Additional evidence for dissociation of MaxΔC(+)P-DNA is provided by the 24 hr V8 digest data. For MaxΔC, at 24 hr of V8 digest, the N-terminal was entirely removed and the C-terminal was partially cleaved, leaving the high-mass fragments 11-103 and 11-96. This situation essentially mirrors the 2 day V8 digest of b/HLH/Z Max (see Figure 4.7). In contrast, at 24 hr of V8 digest of MaxΔC(+)P, *the protein was completely digested.* The subtilisin results also support the notion that Max phosphorylation increases the dissociation of the protein-DNA complex. At 9 hours of subtilisin digest of MaxΔC, the b/HLH/Z protease-resistant domain remained. In contrast, after 9 hours of V8 hydrolysis, the MaxΔC(+)P-DNA complex *was completely digested.* From this data, I hypothesize that the MaxΔC(+)P-DNA complex has a shorter 'lifetime' than the MaxΔC-DNA complex.

The effects of phosphorylation on the association and dissociation rates of Max was shown in a recent study by Lüscher (Bousset et al., 1993). Employing electrophoretic mobility shift assay gels, the authors demonstrate the effects of N-terminal phosphorylation. Using the gel shift assays, they showed that phosphorylation of MaxΔC completely abolished DNA binding. This finding was in general agreement with other studies which also suggested that phosphorylation of Max by CKII substantially reduced or abolished DNA binding (Berberich and Cole, 1992; Prendergast et al., 1992; Bousset et al., 1993; Koskinen et al., 1994). From competitive binding assays, Lüscher reached their principal findings of the study —that phosphorylation of Max by CKII increased both the 'on' (association) and 'off' (dissociation) rates of DNA binding. The changes in the rates were found to be similar. They correlated these results to the fact that steady state binding was unaffected by phosphorylation. Similar changes in k_{on} and k_{off} (the association and dissociation rate constants of the Max-DNA complex) would in effect 'cancel out' leaving K_D (the binding constant, $=k_{off}/k_{on}$) essentially unchanged.

The main result of Lüscher —that the on and off rates of MaxΔC increase upon phosphorylation— is in agreement with the proteolysis results described above. The proteolysis data suggested a reduced lifetime of the MaxΔC(+)P-DNA complex compared to the MaxΔC-DNA complex. However, it remains unclear why the MaxΔC(+)P-DNA complex is not observed by gel

shift assays given that the proteolysis data strongly suggests that phosphorylated Max binds DNA and forms a proteolytic-resistant complex. The reasons for not observing the gel-shift complex may be related to the DNA-protein binding kinetics (i.e., the on and off rates) of the complex through the gel. A high dissociation rate may drastically reduce the number of intact complexes in the gel so that few or none are detected.

4.8.3 Conclusions

The HLH family of DNA-binding proteins has served as a good model to study protein structure by limited proteolysis and mass spectrometry. Used carefully, proteolysis can usefully probe protein structure. Important to the success of discerning protein structure by proteolysis is the use of MALDI-MS, where a large number of digestions, performed under various conditions, can be examined with high mass accuracy, ease and speed.

A particularly attractive use of the methods described in this chapter would be the exploration of the effects of substrate length and conformational flexibility on protease activity. A great deal of work has been done in the past 30 years on protease specificity, but not nearly as much on the understanding other factors such as substrate polypeptide flexibility. This lifetime of effort, by the way, was achieved almost exclusively using non-MS methods. I believe MS will open up many avenues for further exploration.

The proteolysis study of Max has already revealed some intriguing findings regarding the proteases themselves. For example, it was shown for

the b/HLH/Z Max-DNA complex, that Lys-C cleaved at Lys24 (the third residue from the N-terminal) but that trypsin (or subtilisin) did not cleave at this residue and nor did Asp-N after Asp23 (the second residue from the N-terminal). However, Asp-N successfully cleaved at Asp23 in the MaxΔC-DNA complex, probably because in MaxΔC, Asp23 is positioned 14 residues from the N-terminal. Thus Asp-N, subtilisin and trypsin seem to require more than two N-terminal residues flanking the target amino acid whereas Lys-C can cleave very short substrates. Systematic studies of digestions combined with the potential use of synthetic polypeptide substrates would shed light on the substrate length specificities and conformational flexibility requirements as well as information regarding the size of the protease active site. Availability of high resolution structures of substrate proteins and the proteases themselves will be very useful to correlate and interpret the experimental data.

Finally, the ability to rapidly and precisely define protein folding domains is a distinct benefit derived from the proteolysis/mass spectrometric study of Max. This application —domain elucidation— will be presented in the next (and last) chapter.

Chapter 5 — Domain Elucidation by Mass Spectrometry

The success of the proteolysis/mass spectrometry study of Max (discussed in Chapter 4) established a direct and powerful application —protein domain elucidation (Cohen, 1996). This chapter describes the utility of domain elucidation in the structural investigation of three proteins —Mef 2C, leptin, and TAFs 42/62.

5.1 Materials

5.1.1 Proteins

Sequences of recombinant proteins are given in Figure 5.1. **Mef 2C:** Mef 2C was from Lei Zhong (Stephen K. Burley Lab, Rockefeller University). The protein construct consisted of Mef 2C 2-105 + the 8-residue C-terminal His-Tag LEHHHHHH (112 residues; calc. mass 13,134 Da). The protein concentration was 300 μM in ~100 mM KCl, HEPES buffer, DTT & 10 % glycerol. **Leptin:** Various forms of recombinant leptin (Ob1) were from Ketan Gajiwala (Burley Lab) and Jeffrey Halaas (Jeffrey Friedman Lab, Rockefeller). The leptins included human —hOb1y (wild type seq. 22-167, 146 residues, 16,024 Da, with disulfide) expressed in yeast; hOb1a (Amgen) (22-167 + N-term. Met, 147 residues, 16,156 Da, with intact disulfide linkage) expressed in *E. Coli*; hOb1b (22-167 + 4-residue N-terminal GSHM from HIS Tag, 150 residues, 16,437 Da,

with disulfide); ΔhOb1 (22-53/62-167 (54-61 deletion), 142 residues, 15,555 Da, with disulfide); and mouse leptin —mOb1a (Amgen preparation) (22-167, 147 residues, 16,136 Da, with disulfide) and mob1b (22-167 + 4-residue N-terminal GSHM from HIS Tag, 150 residues, 16,417 Da, with disulfide) expressed from *E.Coli*. The leptin samples were usually in KCl and buffer solution (no glycerol) and stored at -70°C.

TAFS 42/62: Various constructs were obtained from Xiaoling Xie (Burley Lab). *Drosophila* $TAF_{II}42$ (p42) "original construct" (wild type seq. 2-100, 99 residues, 11,297 Da; and a C55S mutant, 11,281 Da); "new construct 1" (NC1) (2-95, 94 residues, C55S, 10,732 Da), NC2 (11-95 + N-term. Met, 86 residues, 9,935 Da and a C55S mutant 9,919 Da.) $dTAF_{II}62$ (p62) 'original construct' (wild type seq. 1-91 + N-term. GlySer from thrombin cleavage, 93 residues, 10,240 Da) and NC1 (1-82 + N-term. Gly, 83 residues, 9,072 Da). The samples were typically 0.6-8 mg/mL (as a heterodimer complex TAF42/62), stored in aqueous solutions of ~20 mM buffer (pH 6), 50 mM KCl, 5 mM $MgCl_2$, 10 mM DTT (for wild-type constructs), 1 mM PMSF (serine protease inhibitor) and 10 % glycerol at - 20°C.

5.1.2 Other materials

Buffers, MALDI matrices and matrix solutions are described in Chapter 2. Centricon desalting devices were from Amicon (Beverly, MA).

Mef-2C
Human Mef-2C (2-105 + 8-residue + C-terminal His tag LEHHHHHH)
112 residues, 13,134 Da

1 *10* *20*
| | |
mGRKKIQITRIMDERNRQVT
FTKRKFGLMKKAYELSVLCD
CEIALIIFNSTNKLFQYAST
DMDKVLLKYTEYNEPHESRT
NSDIVETLRKKGLNGCDSPD
PDADDLEHHHHHH

---

hOb1y
Leptin 22-167+disulfide
146 residues, 16,024 Da

21 *30* *40*
| | |
aVPIQKVQDDTKTLIKTIVT
RINDISHTQSVSSKQKVTGL
DFIPGLHPILTLSKMDQTLA
VYQQILTSMPSRNVIQISND
LENLRDLLHVLAFSKS**C**HLP
WASGLETLDSLGGVLEASGY
STEVVALSRLQGSLQDMLWQ
LDLSPG**C**

hOb1a (Amgen)
Leptin 22-167+N-term Met+disulfide
147 residues, 16,156 Da

21 *30* *40*
| | |
MVPIQKVQDDTKTLIKTIVT
RINDISHTQSVSSKQKVTGL
DFIPGLHPILTLSKMDQTLA
VYQQILTSMPSRNVIQISND
LENLRDLLHVLAFSKS**C**HLP
WASGLETLDSLGGVLEASGY
STEVVALSRLQGSLQDMLWQ
LDLSPG**C**

hOb1b
Leptin 22-167
+N-term. GHSM+disulfide
150 residues, 16,437 Da

(21) *30* *40*
| | |
GHSMVPIQKVQDDTKTLIKT
IVTRINDISHTQSVSSKQKV
TGLDFIPGLHPILTLSKMDQ
TLAVYQQILTSMPSRNVIQI
SNDLENLRDLLHVLAFSKS**C**
HLPWASGLETLDSLGGVLEA
SGYSTEVVALSRLQGSLQDM
LWQLDLSPG**C**

ΔhOb1 (54-61 deleted)
Leptin 22-53/62-167
+ N-terminal STDP +disulfide
142 residues, 15,555 Da

(11) *(20)* *30*
| | |
STDPVPIQKVQDD
TKTLIKTIVTRINDISHTQS
VSSkqkvtgldFIPGLHPIL
TLSKMDQTLAVYQQILTSMP
SRNVIQISNDLENLRDLLHV
LAFSKS**C**HLPWASGLETLDS
LGGVLEASGYSTEVVALSRL
QGSLQDMLWQLDLSPG**C**

<u>mOb1a (Amgen)</u>

mouse leptin 22-167
+ N-terminal Met + disulfide
147 residues, 16,134 Da

(21) *30* *40*
| | |
MVPIQKVQDDTKTLIKTIVT
RINDISHTQSVSAKQRVTGL
DFIPGLHPILSLSKMDQTLA
VYQQVLTSLPSQNVLQIAND
LENLRDLLHLLAFSKS**C**SLP
QTSGLQKPESLDGVLEASLY
STEVVALSRLQGSLQDILQQ
LDVSPE**C**

<u>mOb1b</u>

mouse leptin 22-167
+ N-terminal GSHM + disulfide
147 residues, 16,417 Da

(11) *(20)* *30*
| | |
<u>GSHM</u>VPIQKVQDD
TKTLIKTIVTRINDISHTQS
VSAKQRVTGLDFIPGLHPIL
SLSKMDQTLAVYQQVLTSLP
SQNVLQIANDLENLRDLLHL
LAFSKS**C**SLPQTSGLQKPES
LDGVLEASLYSTEVVALSRL
QGSLQDILQQLDVSPE**C**

<u>TAF42 (original construct)</u>

$dTAF_{II}$ 42 2-100
99 residues, 11,297 Da

1 *10* *20*
| | |
mSAEKSDKAKISAQIKHVPK
DAQVIMSILKELNVQEYEPR
VVNQLLEFTFRYVTCILDDA
KVYANHARKKTIDLDDVRLA
TEVTLDKSFTGPLERHVLAK

<u>TAF42 (original construct) C/S</u>

$dTAF_{II}$ 42 2-100; C55S
99 residues, 11,281 Da

1 *10* *20*
| | |
mSAEKSDKAKISAQIKHVPK
DAQVIMSILKELNVQEYEPR
VVNQLLEFTFRYVT**S**ILDDA
KVYANHARKKTIDLDDVRLA
TEVTLDKSFTGPLERHVLAK

<u>TAF42 NC1</u>

$dTAF_{II}$ 42 2-95 C55S
94 residues, 10,732 Da

1 *10* *20*
| | |
mSAEKSDKAKISAQIKHVPK
DAQVIMSILKELNVQEYEPR
VVNQLLEFTFRYVT**S**ILDDA
KVYANHARKKTIDLDDVRLA
TEVTLDKSFTGPLER

TAF42 NC2
dTAF$_{II}$ 42 11-95 + N-term. Met.
86 residues, 9,935 Da

```
1         10          20
|         |           |
MsaeksdkakISAQIKHVPK
DAQVIMSILKELNVQEYEPR
VVNQLLEFTFRYVTCILDDA
KVYANHARKKTIDLDDVRLA
TEVTLDKSFTGPLER
```

TAF42 NC2 (C/S)
dTAF$_{II}$ 42 11-95 C55S + N-term. Met. 86 residues, 9,919 Da

```
1         10          20
|         |           |
MsaeksdkakISAQIKHVPK
DAQVIMSILKELNVQEYEPR
VVNQLLEFTFRYVTSILDDA
KVYANHARKKTIDLDDVRLA
TEVTLDKSFTGPLER
```

TAF62 (Original Construct)
dTAF$_{II}$ 62 1-91+ N-term GlySer
93 residues, 10,240 Da

```
1         10          20
|         |           |
GSMLYGSSISAESMKVIAES
IGVGSLSDDAAKELAEDVSI
KLKRIVQDAAKFMNHAKRQK
LSVRDIDMSLKVRNVEPQYG
FVAKDFIPFRFAS
```

TAF62 NC1
dTAF$_{II}$ 62 1-82+ N-term Gly
83 residues, 9,072 Da

```
1         10          20
|         |           |
GMLYGSSISAESMKVIAESI
GVGSLSDDAAKELAEDVSIK
LKRIVQDAAKFMNHAKRQKL
SVRDIDMSLKVRNVEPQYGF
VAK
```

Figure 5.1 *Protein constructs for domain elucidation studies.*

Numbering systems are based on the wild-type sequences with insertions and deletions as indicated. Lower case letters designate residues that are omitted in the constructs. Underlined residues are not in the wild-type sequence. Histidine Tag in Mef2C is underlined. Calculated average masses are shown. The pair of Cys residues in leptin (in bold) are disulfide linked. Cys to Ser mutations in TAF42 are bold.

5.2 Methods

5.2.1 Proteolysis and Cyanogen Bromide Digestion

Proteolytic digestion and MALDI-MS peptide mapping were carried out as described in Chapter 2. Cyanogen bromide (CNBr) digests of leptin were performed in either 50-70% formic acid or 0.1 N HCl. A small crystal of CNBr was added to the acidified protein solutions to commence the digestion, performed in the dark and in the hood. Aliquots (~1 μL) were sampled at 1, 2, 5, 15 min and 1 hr. of digestion and added to matrix solution (4HCCA/FWI; see Chapter 2). Longer digest times were not necessary, and when carried out in formic acid resulted in peptide formylation. Alternatively, CNBr digests were performed *directly in the matrix solution* from which samples were taken and deposited on the MALDI probe tip for matrix crystallization (IN THE HOOD!).

5.2.2 Electrospray Ionization Mass Spectrometry (ESI-MS)

ESI-MS was performed on a Finnigan TSQ-700 triple quadrupole mass spectrometer with the assistance of Urooj A. Mirza. ESI-MS on protein samples was carried out in one of three ways—denaturing direct infusion, denaturing on-line desalting, and non-denaturing conditions.

5.2.2.1. Denaturing Conditions/Direct Infusion

Proteins that were in solutions having low involatile salt concentration (i.e., below 1 mM in final the ESI solution) were directly infused into the mass

spectrometer via a syringe pump (Harvard Apparatus). For infusion, proteins were dissolved in water:methanol:acetic acid (1:1:2.5%) to make 20-50 μL of a 10-20 μM protein solution. The infusion flow rate was 3 μL/min. The electrospray conditions were 4.5 kV spray voltage, ~3 μA current, 175-200°C stainless steel transport capillary temperature, ~100 V skimmer voltage and 1.2-1.3 kV multiplier. The mass spectrometer was typically set to scan m/z 600-2000. The raw m/z spectra were converted to mass data using the Finnigan deconvolution software.

5.2.2.2 Denaturing Conditions/On-line Desalting Cartridge

If protein samples contained salts and additives, the samples were loaded onto an on-line protein desalting trap cartridge (Michrom BioResources, Inc., Pleasanton, CA). Initially, the cartridge was flushed clean with 100 μL injections of 1:1 water/ACN or water/methanol followed by flushing with pure water. Protein samples were then loaded onto the cartridge from solutions containing little or no organic solvents. From the same direction as protein loading, the cartridge was flushed with pure water or 0.05 % acetic acid solution. The number of washings depended on the salt level in the original protein sample. In a direction *opposite* to that for protein loading, the protein was eluted off the cartridge into the ESI source with a solution of 70% ACN/30% water (acidified with 2.5% acetic acid) infused via the Harvard syringe pump at a rate of 6μL/min. (All other instrument conditions are as described for direct infusion.) After about

six minutes of elution, protein signal appeared and remained steady for a few minutes for data collection.

5.2.2.3 Non-Denaturing ESI-MS

Non-denaturing mode was used to study non-covalent interactions between TAF42 and TAF62 in solution. Protein samples were extensively treated to remove salts and additives via the Centricon (see below). Traces of salts and additives severely compromised the electrospray process. Low concentrations ($\leq$ 10mM) of a volatile buffer, such as ammonium acetate, are usually tolerable. If a buffer is included it should be of the highest purity. In the study of the TAFs, it was determined that pure water (no buffer) gave the best electrospray results. Appropriately prepared protein solutions were directly infused into the ESI source of the mass spectrometer as described above. Typical concentrations of proteins used in the non-covalent studies ranged from 50-100 μM, infusion flow rate of 3μL/min, capillary temperature of 150-180°C, and a skimmer potential of 130-150 V. The mass spectrometer was normally scanned in the range m/z 2000-3800. Adjustments made in the capillary temperature and the skimmer voltage affected the probability of observing non-covalent interactions. After each adjustment of capillary temperature, 10-15 minutes was allowed to reach thermal equilibrium. Observing non-covalent interactions with ESI-MS is a more an art than science and requires patience and experimenting.

5.2.3 Centricon Desalting

Protein samples for ESI-MS non-covalent studies were in solutions essentially free of involatile salts, buffers and other additives (e.g. glycerol). The Centricon 10 concentrator (10 kDa molecular weight cutoff) (Amicon; see Materials) was used to desalt/clean the TAF42/62 samples. The Amicon instruction booklet details information on handling and assembling the Centricon tubes. The Centricon desalting usually takes a whole day (up to 10 hr). (1) A fresh supply of Centricon tubes was always used (shelf-life is a few months). (2) The Centricon membrane contains glycerol & antibacterial agents that must be thoroughly removed by first hand rinsing with MilliQ water and then centrifuging TWICE with 2 mL of water through the Centricon using the Beckman GS-6R refrigerated Centrifuge, set to 4°C at ~6000 rpm (~5500 g relative centrifugal force) for one hour. A second Centricon tube filled with an equivalent volume of water was always to balance the centrifuge. (3) For desalting, 1 mL of water was first added to the prepared Centricon tube followed by the protein sample (typically 10 nmol of protein sample, thawed from -20°C storage) and then enough water to almost fill the tube (~2 mL total). The tube was centrifuged for ~1 hr at 6000 rpm and 4°C. The tube was refilled with fresh water and the centrifuging/washing cycle was repeated 7 times. After the final wash, the tube was inverted and spun at 2500 rpm for 3 minutes to collect about 40 μL of desalted sample. The samples were either immediately used for ESI-MS or

stored at 4°C. TAF 42/62 was vulnerable to adventitious proteolysis and needed to be used within a day or two after desalting.

5.3 Introduction

Protein crystallographers continually seek improved strategies to grow crystals that will diffract to high resolution. In order to obtain high-resolution structures it is important to have the assurance of working with compact, well defined proteins. This generally accepted notion presumes that restricting the degrees of conformational motion in a protein increases the chances that it will crystallize into well-ordered lattices (Joachimiak and Sigler, 1991). A precisely defined compact 'folding domain' also results in a protein with reduced tendencies towards aggregation.

A hallmark of large, multi-functional proteins is their modularity. The constituent modules are often described as being organized as 'beads on a string'. The modules represent autonomously folded polypeptide structures, frequently referred to as domains, where a domain is generally from 50-150 amino acid residues in length. Protein domains are linked together usually by flexible, unstructured polypeptide chains. Figure 5.2 depicts the situation in cartoon form. A protein with three folding domains is shown. Conformational

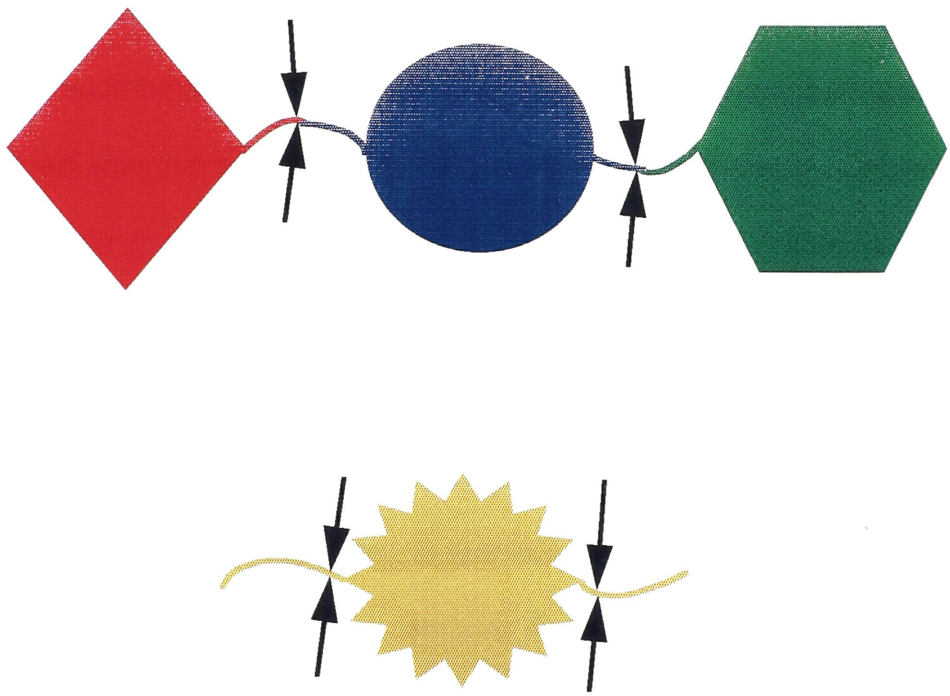


Figure 5.2 *Domain elucidation*

TOP: Hypothetical three-domain protein. Folding domains are shaded red, blue and green. Interdomain linkages are denoted as thin colored lines. Limited proteolysis in solution would cleave within the polypeptide linkages (arrows) releasing the individual folding domains.

BOTTOM: Single protein domain that has polypeptide chains extending beyond compact folding region as a random coil at the N- and C-termini. Limited proteolysis would trim off these flexible chains (arrows) leaving a more compact folding domain.

flexibility about the interdomain linkages would reduce packing efficiency within a crystal of the three-domain protein. Limited proteolysis of the protein in solution would cleave within the interdomain linkages, releasing the individual domains. Alternatively, recombinant forms of individual domains, whose boundaries are not well defined, may have polypeptide extending from the N- and/or C-termini, a situation that may give rise to poor crystal packing or aggregation instead of crystallization. The classical biochemical method to identify protein domains is limited proteolysis combined with SDS-PAGE, HPLC and N-terminal sequencing. However, the poor mass accuracies of SDS-PAGE (no better than 5%) and the expense and time requirements of N-terminal sequencing (days-weeks) is unappealing in the need for speedily attaining a well defined folding domain for high resolution structural studies.

In Chapter 4, limited proteolysis combined with MALDI-MS provided precise details of the domain boundaries of Max and related proteins. The proteolysis study of MaxΔC serves as a good example of domain elucidation using MALDI-MS. Figure 5.3 summarizes the results for the proteolysis of MaxΔC (Max 2-12/22-113). In the absence of DNA, proteolysis of the protein with any of six enzymes occurred rapidly (< 1 hr) and throughout the length of the protein. Upon binding to Max-specific DNA, a protease-resistant folding domain (Max 25-108) become defined by about one hour of digestion. This domain (Max 25-108) was slightly more compact than the original b/HLH/Z

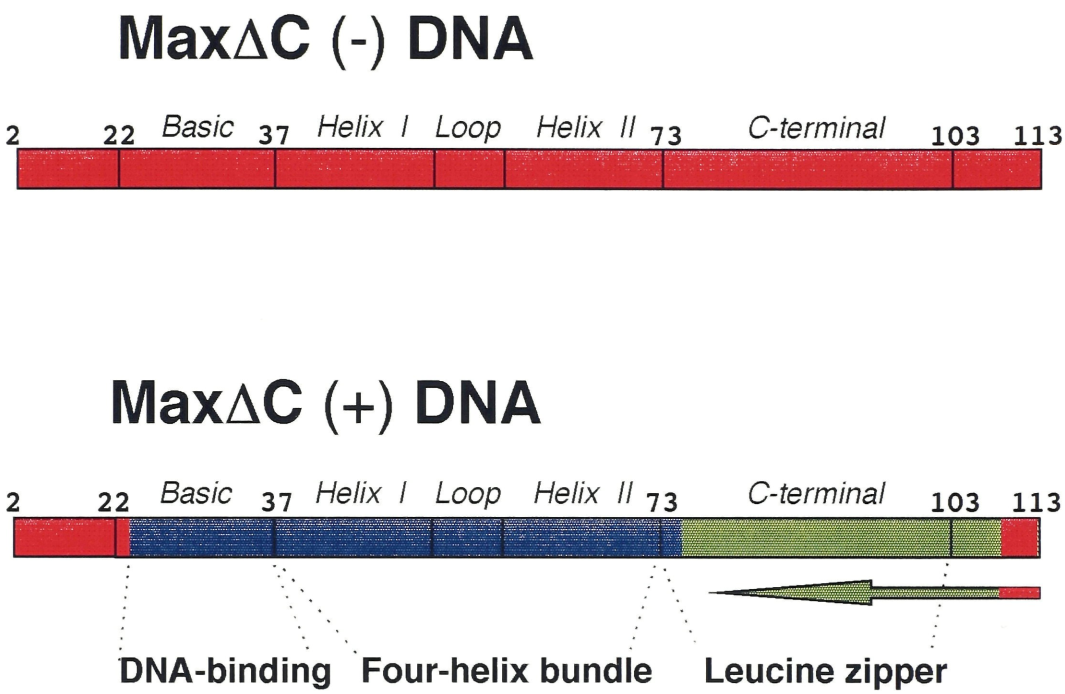


Figure 5.3 *Domain elucidation of MaxΔC*

The protein MaxΔC was probed by proteolysis in the absence **(TOP)**, and presence **(BOTTOM)** of Max-specific DNA. The three colors code for relative rates of proteolysis as measured by MALDI-MS: red is most rapid (minutes-hours), green is intermediate (hours-days), blue are regions where little or no proteolysis occurred. Horizontal arrow denotes direction of gradual, sequential proteolysis observed in the C-terminal region in the presence of DNA. Two protease-resistant folding domains become defined upon binding to DNA: 25-108 (b/HLH/Z Max) and 25-75 (b/HLH Max).

construct (Max 22-113) used in the X-ray crystallographic study of Max. A gradual proteolytic cleavage through the leucine zipper (over a ~48 h period) resulted in a 51-residue protease-resistant structure (Max 25-75). This structure represents a b/HLH (zipperless) form of the protein. Strong resistance to any further proteolysis demonstrated the remarkable globular-like characteristics of the four-helix bundle (Ferré-D'Amaré et al., 1993). In retrospect, knowledge of the more compact, protease-resistant forms of Max may have benefited the Max-DNA crystallization study, especially since the original long form of Max (i.e. MaxΔC) failed to produce crystals (Ferré-D'Amaré and Burley, 1997).

The utility of limited proteolysis and MALDI-MS analysis for domain elucidation can also be seen from a recent investigation of the domain organization of the *E. Coli* RNA polymerase α subunit (Blatter et al., 1994). In this study, SDS-PAGE of a number of different proteolytic digests showed that the α subunit was comprised of two domains connected by a short flexible linker. MALDI-MS was used to determine the exact N- and C-terminal residues of the proteolytically-released domains as well as the length of the linker chain. The authors used this information in developing a model for the role of the α subunit in transcription activation. More recently, limited trypsin digestion, MALDI-MS, and amino acid sequencing were used to determine the domain organization of the *E. Coli* RNA polymerase σ^{70} subunit (Severinova et al., 1996).

The combination of proteolysis and mass spectrometry can provide information about the compact folding domains of proteins with unprecedented

accuracy and speed. Armed with a knowledge of the precise boundaries of protein folding domains, protein constructs can be redesigned that are 'well behaved' for NMR studies and exhibit elevated propensities towards yielding high resolution diffraction-quality crystals for X-ray crystallography. Three proteins with unknown structures will be discussed to demonstrate the utility of the proteolysis/mass spectrometry domain elucidation strategies.

5.4 Results — Domain Elucidation of Proteins of Unknown Structures

5.4.1 Mef-2C

Myocyte enhancer factor-2C (Mef2C) is a eukaryotic tissue-restricted transcription factor that is a member of the MADS box DNA-binding protein family (Martin et al., 1993). The MADS box is a 58-residue domain that is conserved in a diverse range of proteins. In Mef2C the MADS box spans residues 1-58 (see Figure 5.4). The MADS box represents a novel DNA-binding domain whose structure was not available (at least at the time of the thesis study). Previous mutagenesis experiments localized the DNA-binding region of the MADS box to residues 1-30 (Martin et al., 1993). Mef2C also has a second region called the Mef2 domain, positioned immediately C-terminal to the MADS box (residues 60-88, see Figure 5.4) that is believed to confer *trans*-activation activity.

A high-resolution X-ray crystallography investigation was begun on the Mef2C protein (Lei Zhong/Burley). For the study, the initial construct of Mef2C

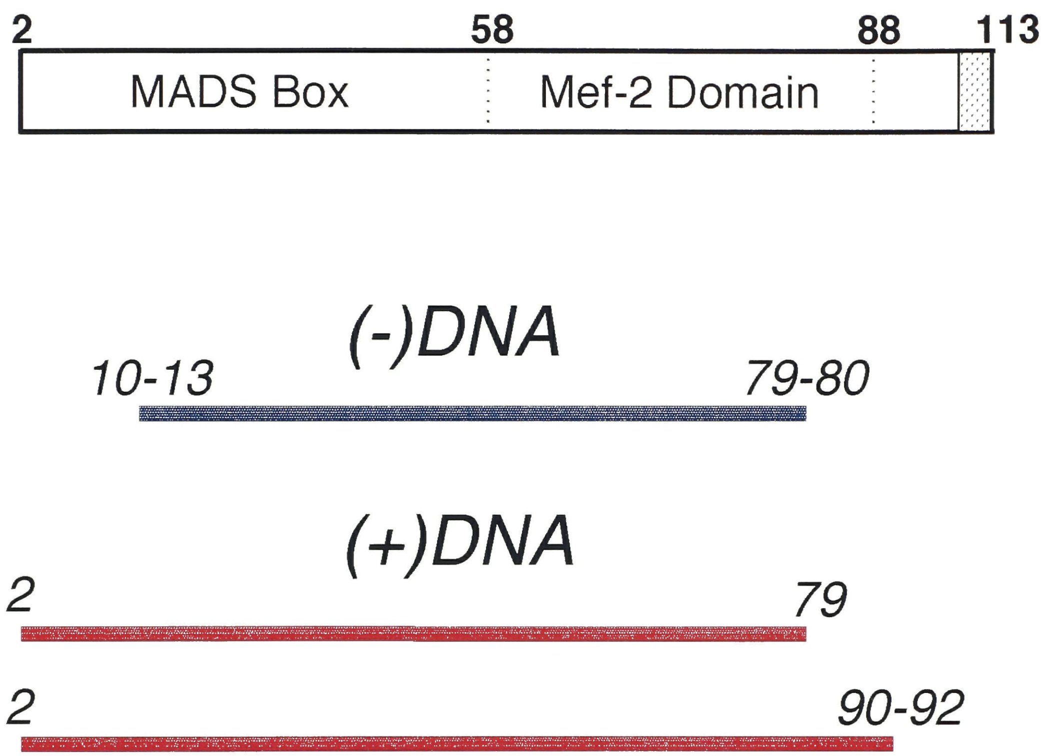


Figure 5.4 *Mef 2C domain elucidation*

TOP: Representation of Mef2C (2-105 + 8-residue His-Tag) containing the DNA-binding MADS box and the MEF2 domains. Stippled region at C-terminal denotes the His-Tag. **BOTTOM:** Summary of the MALDI-MS peptide mapping of the limited proteolysis of Mef2C. The solid colored lines designate proteolytic resistant regions observed in the absence (blue) and presence (red) of DNA.

was based on sequence homology alignment with other proteins of the Mef2 family. A Histidine Tag was included at the C-terminal end of the protein to assist in protein isolation and purification following recombinant expression. Although normally positioned at the N-termini of proteins, the His Tag was placed at the C-terminal to avoid disrupting the highly conserved N-terminal DNA-binding region. Although sequence alignment is a useful guide for designing protein constructs, the application of limited proteolysis and mass spectrometry was used here to more accurately define the folding domain(s) of the Mef2C protein.

Subtilisin and trypsin digests were performed on the Mef2C protein in the absence and presence of DNA. The MALDI-MS peptide mapping results are summarized in Fig. 5.4. The data strongly indicated that residues 1-10 of Mef2C play a critical role in DNA binding since these residues are highly protected from proteolysis in the presence, but not absence of Mef-specific DNA. The proteolysis data also suggested that the C-terminal domain boundary of Mef2C when bound to DNA was at residues 79 or 90-92. Because the entire Mef2 homology domain spans residues 60-88, residue 91 was selected as the domain boundary. Thus the design of a new construct of Mef2C was based on combining the proteolysis data and sequence homology information.

The domain elucidation of Mef2C by MALDI-MS took less than 2 days to complete. Expression of a new construct would require a few weeks to prepare. However, during this time a group in Switzerland released an X-ray structure of a MADS box protein bound to DNA (Pellegrini et al., 1995), essentially scooping

the Mef2C project. It is nevertheless instructive to consider some of the relevant findings of the MADS box X-ray crystallography. The protein in the published X-ray structure was serum response factor (SRF). The crystallographers used a 92-residue construct of SRF that contained the MADS box and solved the structure to 3.2 Å resolution. Interestingly, the SRF-DNA structure showed that the N-terminal 11 residues of the MADS domain made specific contacts with the minor groove of the DNA (Pellegrini et al., 1995). The minor groove interactions provide a rationale for the proteolytic protection of the N-terminal residues that I observed in the Mef2C-DNA complex (Figure 5.4). The crystallographers that solved the SRF-DNA complex also predicted that the N-terminal residues of the MADS box would be unstructured in the absence of DNA. The proteolysis data on Mef2C (without DNA) are consistent with the prediction of an unstructured N-terminal. Finally, I speculate on whether more precisely defined protein boundaries of the SRF construct (i.e., domain elucidation) would have improved the X-ray diffraction data? The SRF construct included an extra nine residues extending *beyond* the N-terminal of the MADS box domain that showed little or no electron density in the X-ray data (Pellegrini et al., 1995). Was inclusion of these extra residues in the SRF construct detrimental to the crystal packing and the diffraction quality? The Mef2C proteolysis data *suggests* that the extra N-terminal residues in SRF were dispensable. It is possible that an optimized construct could have yielded an X-ray structure with resolutions better than that obtained (3.2 Å).

5.4.2 Leptin

The leptin protein was introduced and discussed in Chapter 3, particularly in regards to characterizing the endogenous human form of leptin for post-translational modifications. Additional mass spectrometric research was also undertaken on leptin and is presented here.

Leptin was under intense investigated by several laboratories at Rockefeller. Physiological studies of injections of recombinant leptin into mice were undertaken in the Friedman laboratory. At the same time, a high resolution X-ray structural study was being carried out by Ketan Gajiwala in the Burley laboratory. Thus, there was a strong need to examine recombinant leptin to determine correctness of expression, purity levels, and the oxidation state of the two cysteine residues present in the protein. In addition, domain elucidation by proteolysis/mass spectrometry of leptin would benefit the structural studies.

Over a one year period, about a dozen different batches of leptin (human and mouse forms) expressed from various sources (bacteria and yeast) were analyzed by MS. ESI-MS was used principally to measure the masses of the intact protein (via the desalting cartridge, see Methods) (data not shown). MALDI-MS was also used to measure the protein mass. Although MALDI- and ESI-MS mass accuracies are similar (0.01% at 16 kDa), attaining these accuracies by MALDI-MS requires internal calibration and repetitive measurements. In contrast, ESI-MS consistently provided 'spot-on' masses with single injections of sample.

MALDI-MS was more useful than ESI-MS to monitor the purity levels of recombinant leptin samples. This has been my general experience in dealing with samples containing low levels (few percent) of 'impurities.' From the MALDI analyses of leptin, it became apparent that the protein was unstable in solution. Many of the recombinant leptin samples that I received consistently showed degradation products (see below).

Ultimately, a preparation of recombinant mouse leptin (from Amgen) was found to be exceptionally pure by MALDI-MS (Figure 5.5a). Using the Amgen protein as a control, I performed a simple experiment to monitor the stability of the protein in solution. The results, shown in Figure 5.5, indicate that the protein is susceptible to partial degradation when stored at +4°C over a long period (5 months) (Figure 5.5d). Little or no degradation was seen at lower temperatures (Figure 5.5b,c). Although the MALDI-MS data is not strictly quantitative, I estimated ~1-2% degradation in the sample stored at 4°C. Trace proteases, perhaps left over from the protein isolation procedures may explain the degradation products. The stability experiment suggested that long-term storage of the protein is best at done -70°C (Figure 5.5b.) Such information is useful for the crystallization trials since they are carried out at 4-25°C for periods of weeks to months.

Careful inspection of the MALDI-MS data (Figure 5.5d) indicates that protein 'self-degradation' is occurring in the N-terminal portion of recombinant mouse leptin, specifically in the stretch of residues 51-60, with the dominant cleavage between residues 57/58. Similar results were obtained in preparations

Figure 5.5 *Long-term stability of recombinant mouse leptin in solution.*

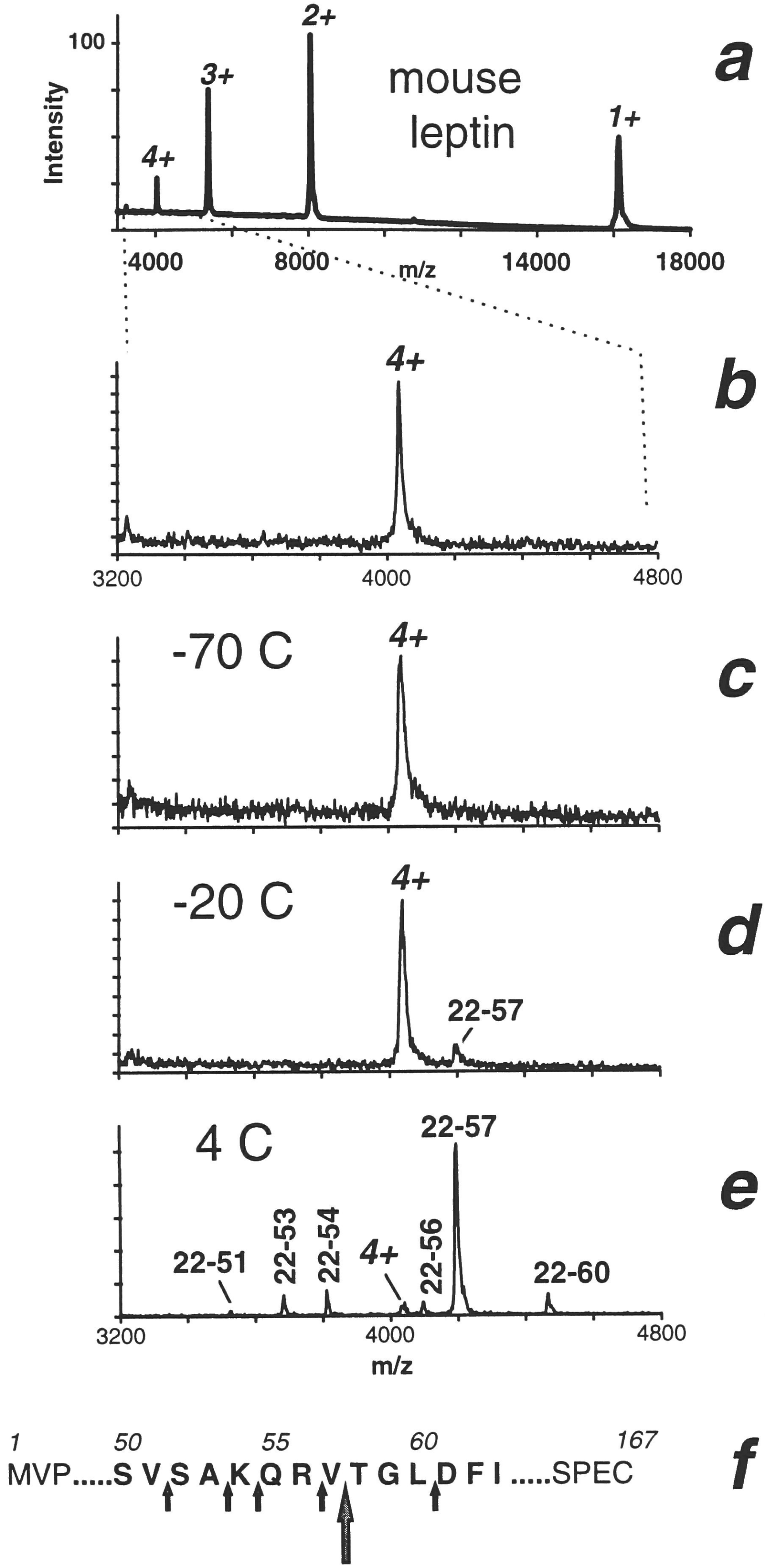
MALDI-MS of recombinant mouse leptin (mOb1/Amgen) that was stored at three different temperatures for five months.

(a) Spectrum obtained from a fresh preparation of protein before storage. The multiple charge states (1+ - 4+) indicate presence of a single component (leptin).

(b) Blow-up of the region m/z 3200-4800 (containing the 4+ ion of intact protein) of the spectrum in panel **a**. Protein degradation products for leptin would appear in this region if they were present, as will be seen in the succeeding panels. For the fresh sample of leptin, there is no indication of protein degradation. The leptin sample was divided and stored at three different temperatures.

(c-e) MALDI mass spectra, collected on the leptin samples after five months of storage at **(c)** -70, **(d)** -20, and **(e)** 4°C. Degradation is most noticeable in the sample that was stored at 4°C. The strong peak arising from degradation at Val57 (i.e. fragment 22-57) observed in the spectrum in panel **(e)** has caused the relative intensity of the +4 ion (from the intact protein) to drop considerably. However, the *signal-to-noise* of the 4+ ion has not changed much from the other spectra **(b-d)**. The mass accuracy of MALDI-MS enables an unambiguous identification the peaks, indicating that the region of mouse leptin containing residues 51-60 is susceptible to degradation.

(f) Summary of mouse leptin 'self' degradation. Partial sequence of the mouse leptin containing the residues susceptible to degradation as determined by MALDI-MS of the recombinant protein (leptin 22-167) shown in panels **a-e**. Arrows point to observed cleavage sites —small arrows designate minor amounts of cleavage, the large arrow designates the site (Val57-Thr58) of greatest degradation.



a
mouse leptin
Intensity
100
4+
3+
2+
1+
4000
8000
m/z
14000
18000
b
4+
3200
4000
4800
c
-70 C
4+
d
-20 C
4+
22-57
e
4 C
22-57
22-51
22-53
22-54
4+
22-56
22-60
3200
4000
4800
m/z
f
1
50
55
60
167
MVP.....S V S A K Q R V T G L D F I .....SPEC

of the human form of leptin, whose residues 50-60 were susceptible to 'self' degradation (Figure 5.6a). To further probe the structure of leptin, subtilisin and trypsin limited digests were performed on the leptin samples. The digests revealed that the most susceptible sites for proteolysis were between residues 54-61 for human leptin, in agreement with the 'self' degradation findings.

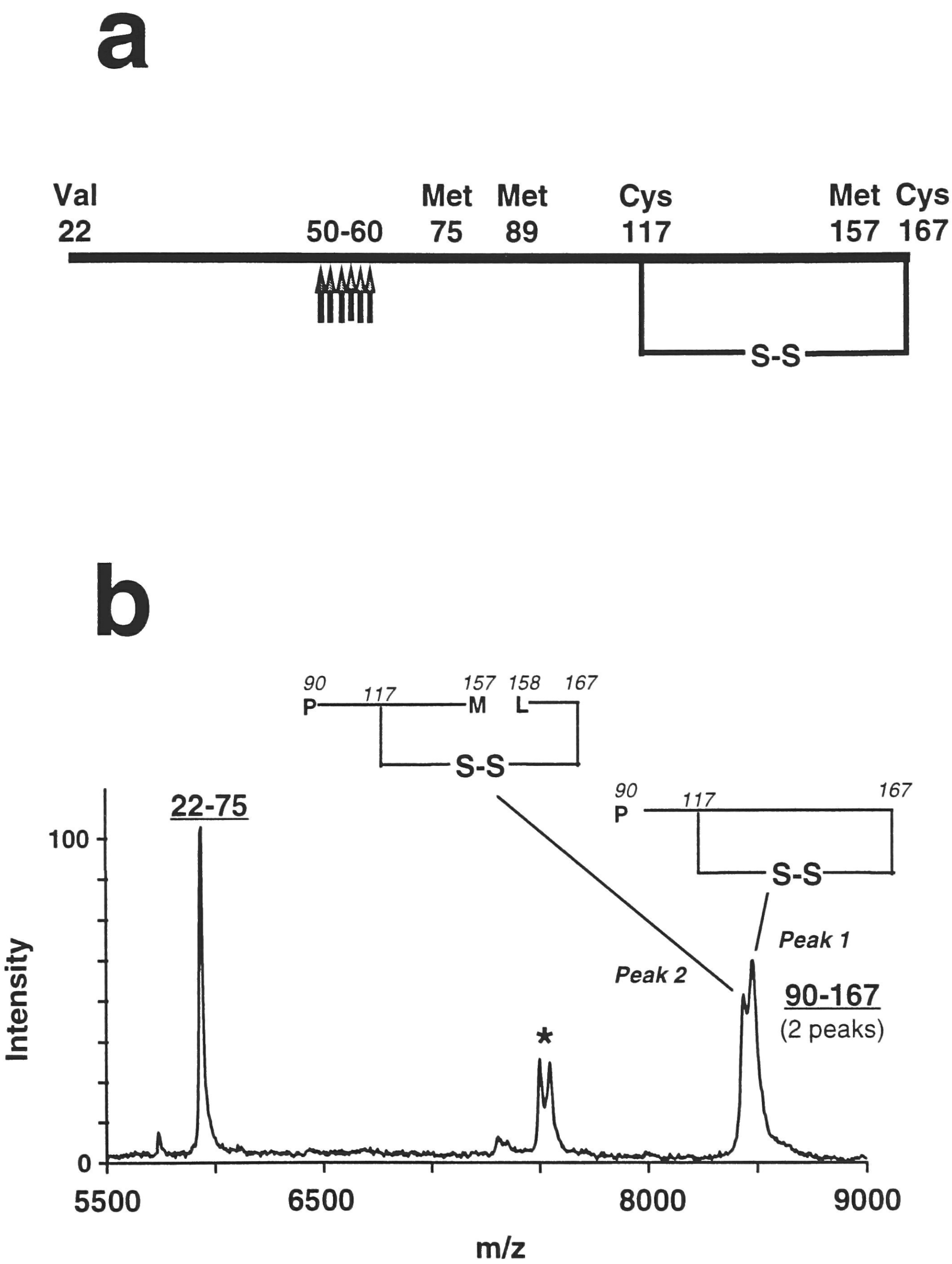
The facile 'self' degradation of leptin suggested that residues 54-61 in human leptin were located within a flexible loop. A deletion mutant of leptin (ΔhOb1) that omitted residues 54-61 was designed. The mutant was substantially more resistant (>100 fold) than the wild-type protein to trypsin proteolysis, especially at the N-terminal (data not shown) —supporting the idea that the region 54-61 is unstructured. Unfortunately, crystallization trials of ΔhOb1 were unsuccessful.

A methylated form of human leptin (obtained through reductive methylation of the free amino groups of the recombinant protein) was also examined by MALDI-MS. Methylation may stabilize the protein and increase the chances for its crystallization. Besides MS, no other methods are available to analyze methylated proteins and accurately map all the sites of modification, which was successfully carried out on methylated leptin (data not shown). Unfortunately, methylation was not the 'magic bullet' that would yield protein crystals. Ketan Gajiwala has spent considerable effort (>1 yr.) on leptin crystallization trials, and has only recently obtained crystals that diffract, albeit to low resolution, by modifying the protein refolding strategies.

Figure 5.6 *Characterization of recombinant human leptin*

(a) Linear representation of recombinant human leptin (hOb1 22-167). Arrows indicate sites of 'self' degradation as measured by MALDI-MS. The three methionine residues of leptin are noted. These are the target residues for CNBr digest. Cyanogen bromide digests indicate the presence of a disulfide linkage between Cys117 and Cys 167 (see below).

(b) MALDI-MS of cyanogen bromide digest of recombinant human leptin, expressed in yeast. The digestion was performed in 0.1 N HCl for one hour. Three of the major peaks can be identified —22-75 and 90-167 (2 peaks)— as CNBr products that map to leptin. The two peaks that correlate to fragment 90-167 are marked peak 1 and 2. CNBr cleavage of leptin at Met89 yields a product that produces peak 1. If there were no disulfide, subsequent cleavage of 90-167 at Met157 would result in 90-157, a peptide that was not observed by MALDI (the second product, 158-167 has no basic groups and was also not observed). If a disulfide linkage were present between Cys117 and Cys167, CNBr cleavage of the 90-167 fragment at Met157 yields the fragment ion (peak 2) whose mass differs from peak 1 by the CNBr hydrolysis of the Met157-Leu158 bond. The pair of peaks labeled with an asterisk could not be identified and were presumed to be an artifact of the CNBr digest.


a
Val
22
50-60
Met
75
Met
89
Cys
117
Met
157
Cys
167
S-S
b
90
117
157
158
167
P
M
L
S-S
22-75
100
0
Intensity
90
117
167
P
S-S
Peak 1
Peak 2
90-167
(2 peaks)
*
5500
6500
8000
9000
m/z

A further use of MS associated with leptin characterization was the identification of the C-terminal intramolecular disulfide bond. Cyanogen bromide digests provided strong evidence for the presence of a disulfide linkage in recombinant human leptin (Figure 5.6b) (Halaas et al., 1995). The key to the disulfide bond identification was the presence of Met157, positioned between the two cysteine residues (Cys117 &167) (See Figure 5.6a). CNBr cleavage produced several products —fragments 22-75 and a pair of peaks that correspond to 90-167 (Figure 5.5b, peaks 1 & 2). CNBr cleavage of leptin at Met 89 leaves fragment 90-167 (peak 1). Subsequent cleavage of the 90-167 peptide at Met157 would result in two fragments —90-157 & 158-167. However, a disulfide between Cys117 and Cys167 leaves the two fragments (90-157 and 158-167) covalently connected through the disulfide linkage, resulting in peak 2. Trypsin digest of the leptin also provide results that support the CNBr findings of the presence of an intramolecular disulfide bond (data not shown).

5.4.3 TAFs 42/62

The TAF proteins are factors associated with the TATA box-binding protein (TBP) (Goodrich and Tjian, 1994; Burley and Roeder, 1996). For RNA polymerase class II promoters, TBP and the TAFs (more specifically TAF_{II}s) represent subunits of the general transcription factor TFIID. TFIID is critical for recruiting proteins in the assembly of the eukaryotic transcriptional machinery and plays a pivotal regulatory role in gene expression. Although, the TAFs are

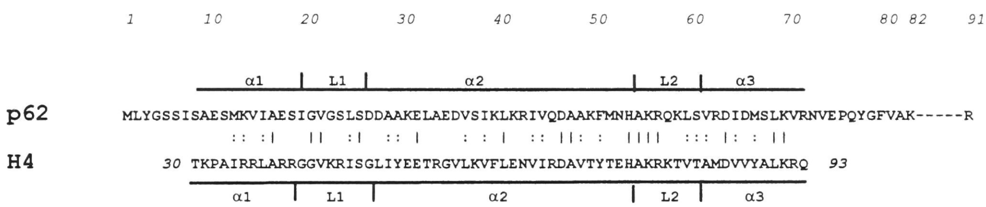


Figure 5.7 *Sequence alignments of dTAF42 & 62 with the histones H3 & H4.*

The histone-like portions of *Drosophila* TAF42 and TAF62 are aligned with chicken histones H3 and H4. Amino acid identities are denoted with **(|)**, and similarities with **(:)**. TAF numbering systems are above the sequences. Secondary structural assigments (α-helices, loops) are based on the TAF42/62 and histone core crystal structures (Xie et al, 1996).

not essential for basal level transcription, they appear to be important for activator-dependent transcription, at least for in vitro assays. The TAFs are believed to mediate interactions between the basal transcriptional machinery (bound to the TATA box through TBP) and transcription and regulatory factors (bound to distal sites on the DNA). A large number of TAFs have been cloned in yeast, *Drosophila* and human (Burley and Roeder, 1996) that display remarkable cross-species sequence homologies for a given TAF protein.

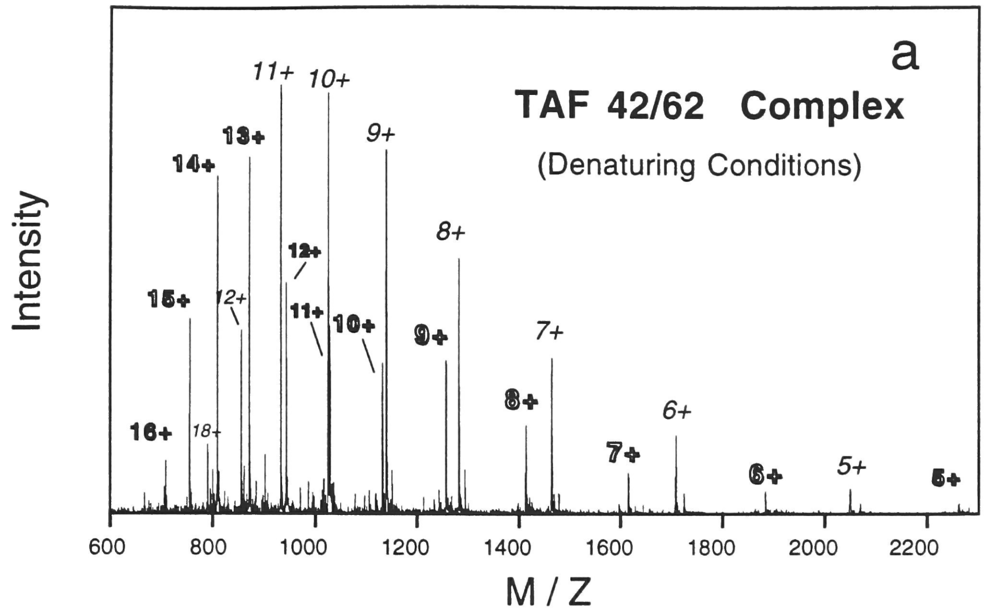
Using biochemical methods, interactions have also been mapped between many of the TAF proteins (Burley and Roeder, 1996). Two of these interactions were mapped to the subunits *Drosophila* $TAF_{II}42$ and $TAF_{II}62$ (Kokubo et al., 1994). Both subunits have homologous yeast and human forms. Immunoprecipitation experiments indicate that TAF42 and TAF62 interact strongly (Kokubo et al., 1994). More striking is the resemblance (at the sequence level) of TAF42 and TAF62 for the non-linker histone proteins H3 and H4 (Figure 5.7). Like histones H3 and H4, TAF42 and TAF62 strongly associate in solution, forming a tight heterodimer (TAF42/TAF62) with a ~10^{-9} M equilibrium dissociation constant. Such a tight association facilitated the co-expression and purification of the TAF42/62 complex. There is also strong evidence pointing to the existence of the heterotetrameric complex $(TAF42/TAF62)_2$, a species that would be analogous to the core histone $(H3/H4)_2$ complex. Ultracentrifugation experiments have shown an equilibrium dissociation constant of TAF dimer/tetramer complex to be ~10^{-6} M (Xie et al., 1996). Also, biochemical

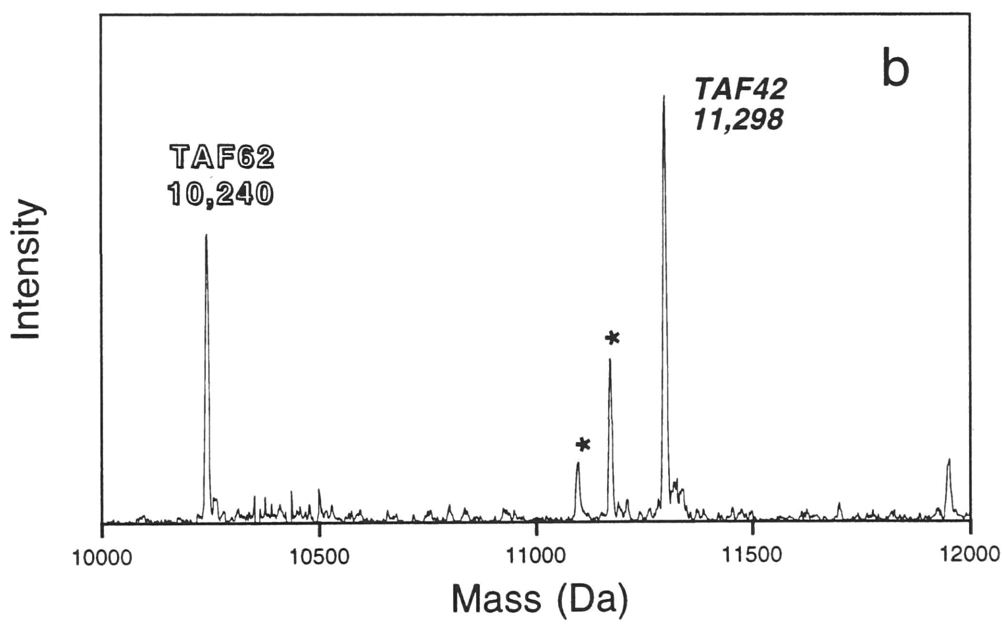
studies have revealed a third TAF (TAF30α in *Drosophila*) interaction with the TAF42/62 tetramer suggesting the existence of a histone octamer-like TAF complex (Hoffmann et al., 1996).

I undertook two studies of the TAF42/62 complex. (1) An investigation of the non-covalent interactions between the proteins by ESI-MS, and (2) domain elucidation of the TAF42/62 complex by MALDI-MS.

5.4.3.1 ESI-MS of Non-covalent Interactions between TAF42 and TAF62

As discussed in Chapter 1, under appropriate solution conditions, ESI-MS can probe the interactions between proteins in solution. Figure 5.8a shows the ESI-MS of the TAF42/62 complex under denaturing conditions (i.e., sprayed from an acidified methanolic solution). The proteins constructs used in this experiment were the TAF42 and TAF62 sequences that spanned the histone-similarity regions (TAF42/62 original constructs; see Materials and Figure 5.7). The spectrum shows two "envelops" of ion peaks —one from TAF42, the other from TFA62— that represent distributions of a highly multiply-charged protein species. Under acidic denaturing conditions (pH ~2.5), the protein is maximally unfolded in solution, exposing acidic and basic residues for protonation that results in a broad distribution of positive charge states for each protein. Figure 5.8b shows a deconvolution of the ESI mass spectrum of Figure 5.8a. Deconvolution mathematically 'combines' the multiple-charged peaks of each protein species into a single peak for each protein (Mann et al., 1989). The two



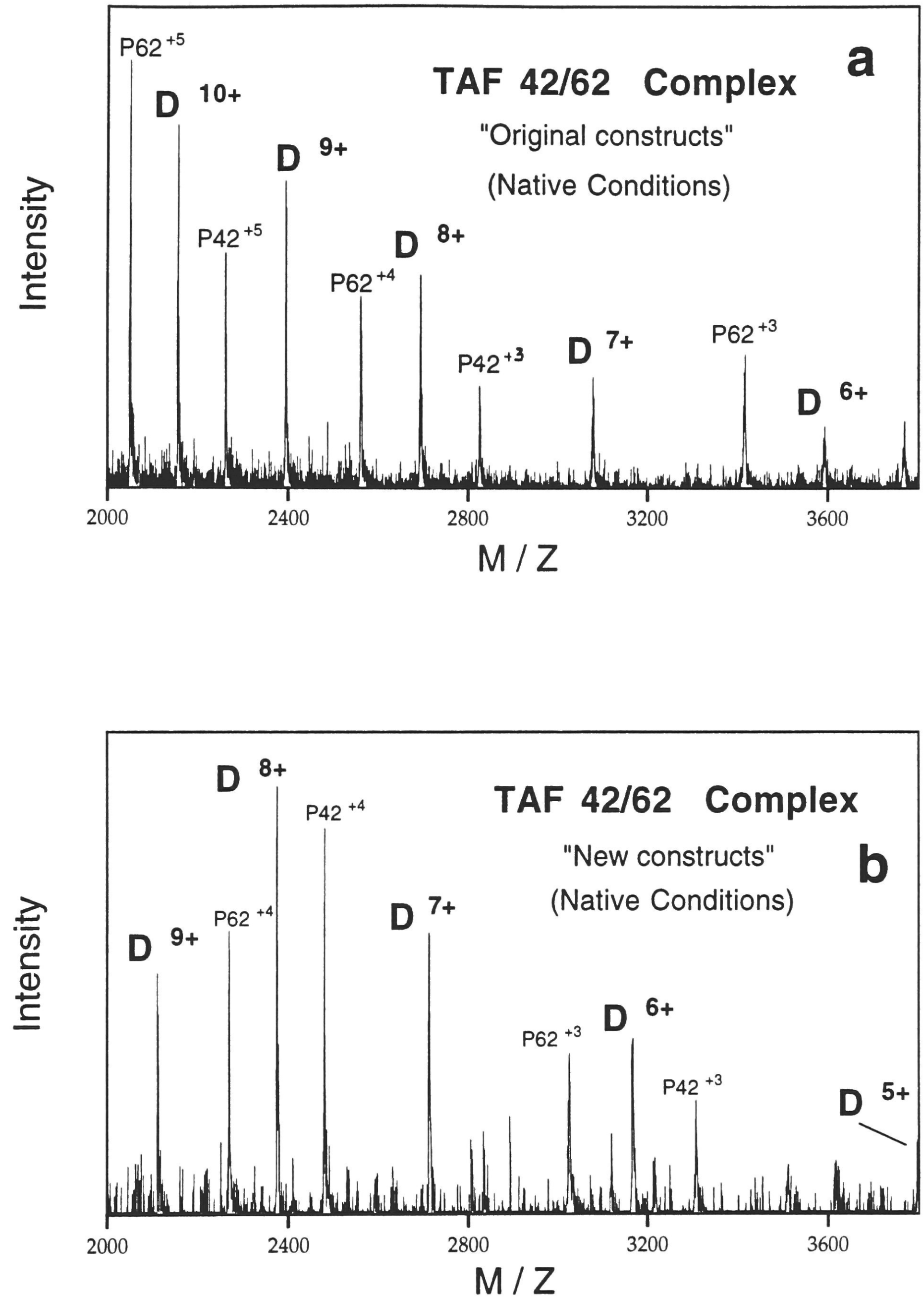


<u>Figure 5.8</u> *ESI-MS of dTAF 42/62 under denaturing conditions*

(a) Raw data labeled with multiple protonation charge states of dTAF42 (1-100)(shadowed #s) and dTAF62 (1-91)(italisized #s), and ***(b)*** deconvoluted data of *(a)* labeled with the measured masses of the TAFs. Expected masses: TAF62, 10,240 Da; TAF42, 11,297 Da. Peaks marked with asterisk are MS fragment ions from TAF42.

major peaks shown in the deconvoluted spectrum have masses that correlate to the TAF42 and TAF62 constructs.

Figure 5.9a shows ESI mass spectra of the TAF42/62 complex electrosprayed under non-denaturing conditions. Comparison with the spectra obtained under denaturing conditions (Figure 5.8a) reveals several important differences. (1) Protein ions with higher m/z values were measured under non-denaturing conditions (Figure 5.9a). No peaks were observed in this high m/z range under denaturing conditions (data not shown). The peaks in the higher m/z range in the non-denaturing conditions (Figure 5.9a) arise from protein with low charge states. Approaching neutral pH, the protein is expected to be in its native, folded state (i.e., TAF42/62 heterodimeric and tetrameric complexes). A more compact structure due to protein folding as well as non-covalent interactions between the subunits would likely reduce the protonation of basic groups due to the 'burial' of protein residues at the interaction interface. It would be energetically highly unfavorable to 'bury' a residue with a net charge, especially into a hydrophobic region such as a globular domain. (2) A set of peaks that correlate to the TAF monomers is observed under denaturing (Figure 5.8a) and non-denaturing (Figure 5.9a) conditions. (3) A set of peaks whose mass correlates to the heterodimeric complex of TAF42/62 is observed under non-denaturing conditions (Figure 5.9a). There is no evidence for homodimers or any form of tetramers. No dimeric peaks were observed under the acidic denaturing conditions. (4) By adjusting the capillary temperature and skimmer voltage, the intensity of the heterodimer peaks (for the non-denaturing solution)



<u>**Figure 5.9**</u> *ESI-MS of dTAF 42/62 under native conditions.*

(a) Original constructs dTAF42 (1-100)/dTAF62 (1-91) and ***(b)*** New constructs dTAF42 (11-95)/dTAF62 (1-82) used for X-ray crystallography of the protein complex. Monomer protein peaks are labeled with their charges as P^{n+}. Non-covalent heterodimer peaks (TAF42/62 complex) are labeled with their charge as D^{n+}. There is no indication of homodimers for either samples.

were made to completely dominate (at reduced temperatures and voltages) or to completely disappear (at elevated temperatures and voltages) (data not shown). The dissociation of the complex at elevated capillary temperatures and skimmer voltages confirmed that the heterodimer TAF42/62 complex measured under non-denaturing solution conditions was non-covalent and that the specificity of interaction (i.e. heterodimerization) is maintained into the gas phase.

Further attempts were made to observe the tetramer complex $(TAF42/TAF62)_2$ by ESI-MS. The most 'gentle' ESI conditions (i.e. low capillary temperature and skimmer voltage) are required to observe dimer-dimer non-covalent interactions which are expected to be much weaker than monomer-monomer interactions. Under extremely gentle conditions, there were only weak indications of a tetrameric species. As mentioned before, the dimer/tetramer equilibrium dissociation constant was determined by ultracentrifugation to be $\sim 10^{-6}$ M —three orders of magnitude weaker than the monomer/dimer dissociation constant. Even the 'gentlest' ESI-MS conditions were probably still too harsh to maintain the interactions of a moderately-weak tetrameric complex into the gas phase.

5.4.3.2 Domain Elucidation of the TAF42/62 Complex for High-Resolution Studies

Xiaoling Xie set out to crystallize the TAF42/62 complex in order to address the possibility that a pair of TAFs has a histone-like topology. [A low

resolution crystal structure of the core histone octamer was available for comparison (Arents et al., 1991).] The first job was selecting initial constructs for crystallization. The task was simplified through the histone TAF42/62-H3/H4 similarity alignments. However, because strong sequence homology was lacking, the exact lengths of the helices and loops in the TAFs were not clear. To accommodate the 'boundary' uncertainties, the original TAF constructs included 8-20 additional residues extending beyond the histone-like portions at the N- and C-termini (except for the N-terminal of TAF62). Throughout the crystallization project, MALDI- and ESI-MS was used to check the identity and purity levels of the recombinant TAF proteins.

After four months of protein expression, isolation and co-purification the TAF42/62 constructs were shown to be monodisperse and tetrameric by dynamic light scattering methods (Xie/Burley). High monodispersion usually means little or no protein aggregation, an important condition for a successful crystallization (Ferré-D'Amaré and Burley, 1997). The TAF proteins readily crystallized, but the crystals were small and diffracted weakly to 4-5 Å. Flexible portions of the protein were suspected to be reducing the packing efficiency into the protein crystal lattice, compromising the quality of the X-ray diffraction. Additional evidence came from MALDI-MS which showed that recombinant TAF42/62 is susceptible to substantial degradation in solution. The MALDI data revealed a complex mixture of degradation products whose identities correlated to proteolytic cleavage at the protein termini (data not shown).

Limited proteolysis was applied to the protein complex to better elucidate its folding domains. Separate digests were performed using endo Asp-N, Glu-C, trypsin, chymotrypsin and subtilisin. MALDI-MS was used to map the digests. Figure 5.10 summarizes the proteolysis results. The digest data made clear that three of four termini were proteolytically susceptible, whereas the N-terminal of TAF62 was highly resistant to digestion.

The proteolysis data provided a rational basis for redesigning constructs of the TAF proteins. Several new constructs were designed, and ultimately the most 'compact' ones —dTAF42 (11-95) and dTAF62 (1-82) (see Figure 5.10)— yielded positive results. From five to nine residues were 'trimmed' from three of the four termini of the original constructs; the TAF62 N-terminal was left unchanged. In less than two months after redesigning the constructs, recombinant protein was available for crystallization trials. The trials promptly produced large crystals that diffracted to better than 1.4 Å resolution (Xie et al., 1996). The TAF42/62 structure solved to 2 Å resolution is shown in Figure 5.11a. For comparison, the H3/H4 heterodimer structure [extracted from the low resolution histone-core octamer structure (Arents et al., 1991)] is also depicted (Figure 5.11b). The folds are very similar. The TAF42/62 structure represents the first high-resolution TAF structure as well as the first high-resolution view of the histone fold, showing that the tetrameric $(dTAF42/dTAF62)_2$ complex (not shown) strongly resembles the $(H3/H4)_2$ heterotetrameric core of the histone octamer (Xie et al., 1996).

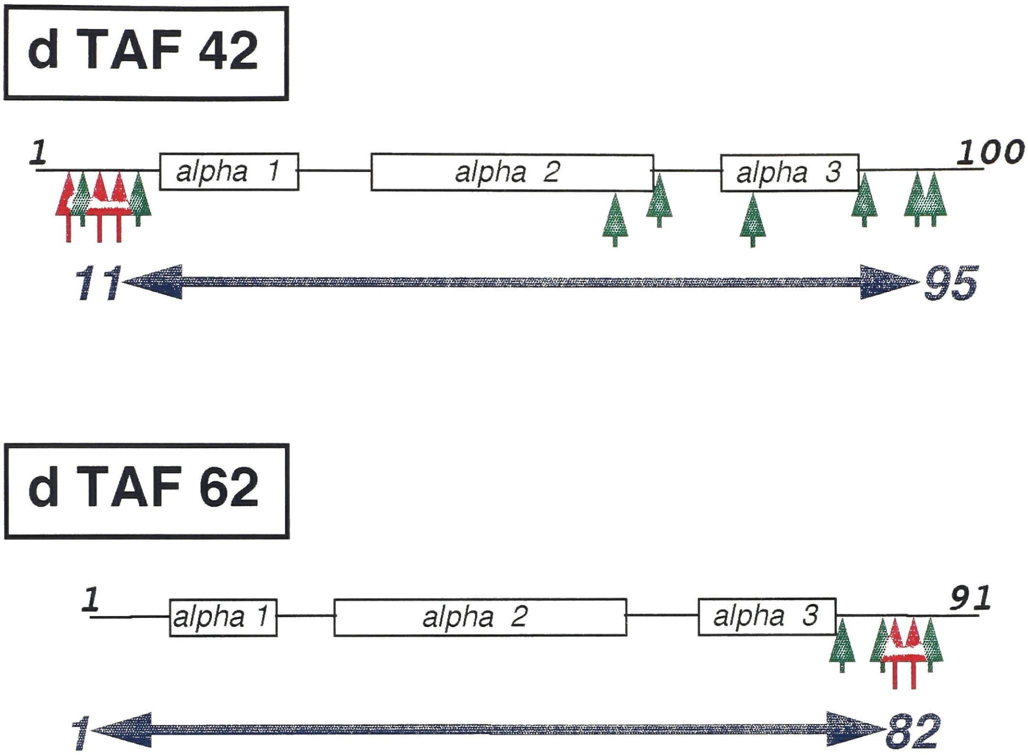
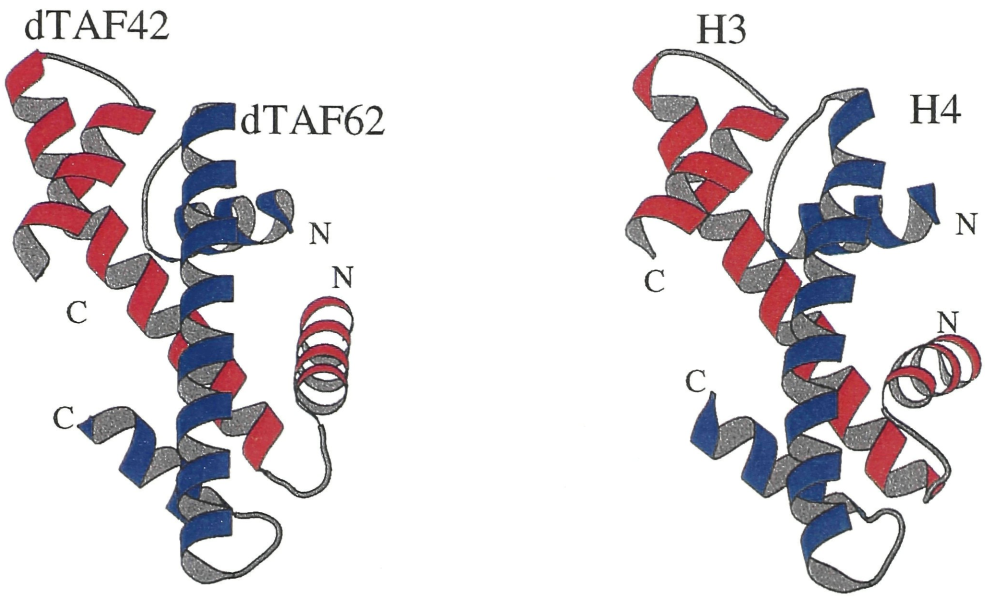


Figure 5.10 *Domain elucidation of the dTAF42/62 complex.*

The linear diagrams outlined in black depict the protein constructs {dTAF42 (1-100) and dTAF62 (1-91)} initially designed for crystallization based on histone H3 and H4 sequence similarities (Fig. 5.7). MALDI-MS mapped the proteolytically susceptible sites. Red arrows denote sites of rapid (min-hr) and green arrows denote sites of slow (hr-1 day) proteolysis. Solid blue lines designate the new constructs used for high resolution X-ray crystallography. The α-helical regions, determined from the crystal structure, are labeled.



<u>Figure 5.11</u> *The dTAF42/62 and histone H3/H4 structures.*

X-ray crystallographic structures of the histone-like regions of dTAF42/dTAF62 complex **(LEFT)**, solved to 2.0Å resolution, showing the binary heterodimeric complex between dTAF42(11-95) *(red)* and dTAF62 (1-91) *(blue),* the 'minimal folding domains' of the proteins as determined by limited proteolysis and mass spectrometry; and **(RIGHT)** the corresponding view of the complex between histones H3 *(red)* and H4 *(blue)* for comparison. The ribbons outline the polypeptide backbones of each protein (Xie et al., 1996).

5.4.4 Domain Elucidation by Mass Spectrometry-Additional Applications

Domain elucidation by mass spectrometry has proven to be an indispensable tool for high resolution studies (described above). Subsequent X-ray crystallographic projects have also benefited from domain elucidation including the HIV I Nef/SH3 protein complex (undertaken in the Chait and Kuryian Labs) (Lee et al., 1996) and two structural studies (in progress) of protein-RNA complexes: the Nova1 protein-RNA complex (Lewis/Burley) and the eukaryotic initiation factor 4E (eIF-4E; the small cap-binding protein) (Marcotrigiano/Burley). In the HIV I Nef/SH3 study, MALDI- and ESI-MS were used to map the limited digests. Both mass spectrometric methods yielded similar information concerning the compact folding domain of the Nef-SH3 complex (U. A. Mirza, personal communication).

5.5 Conclusions —Mass Spectrometry as a Tool for Protein X-ray Crystallography

Figure 5.12 outlines an overall strategy of applying mass spectrometry to high-resolution protein X-ray crystallography. A comparable diagram could also be drawn for multidimensional NMR spectroscopy. MS can facilitate the route to a high resolution structure (Chait, 1994) at almost every stage of crystallography. Figure 5.12 summarizes these stages. (1) Protein expression,

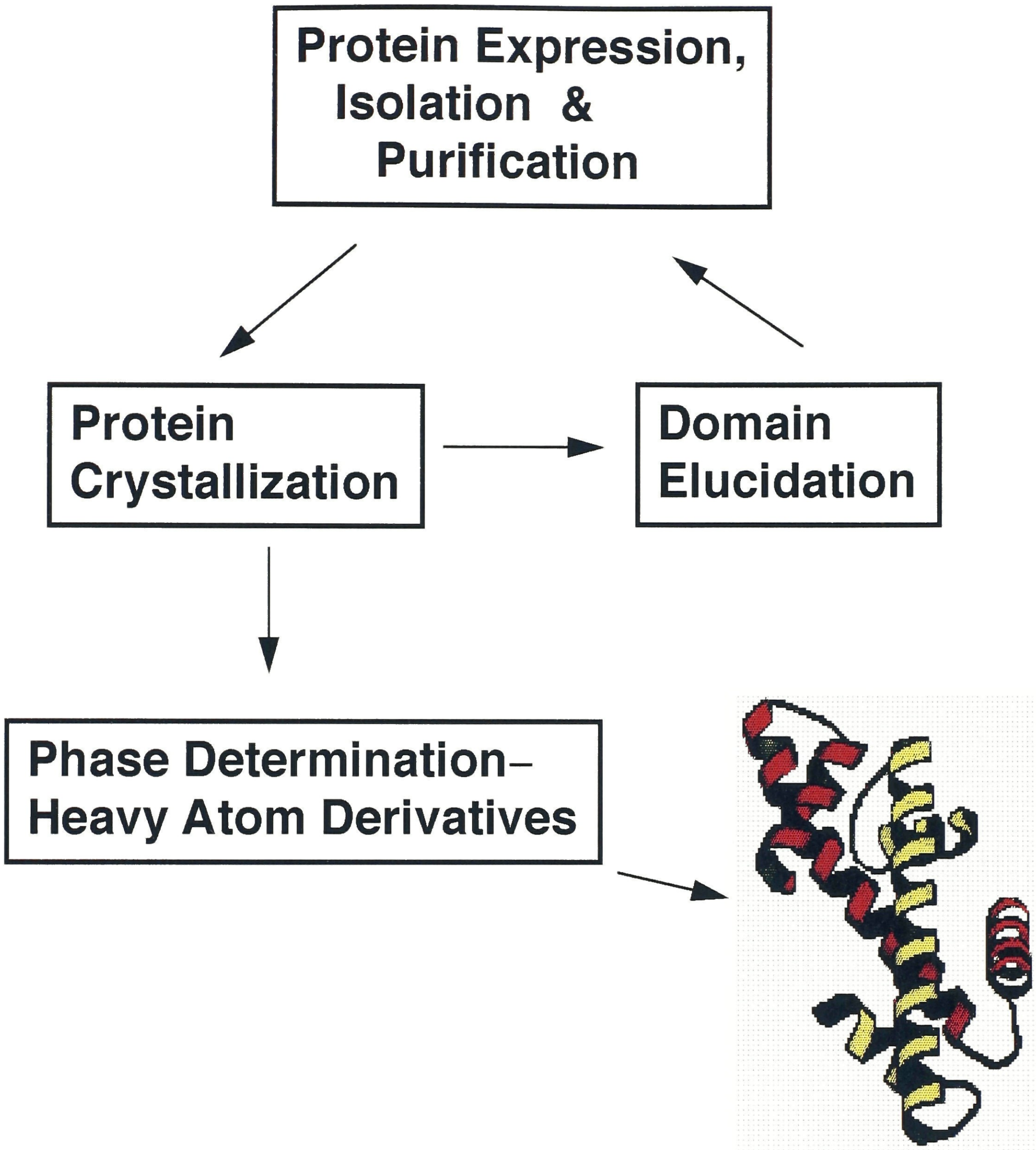


Figure 5.12 *Protein X-ray crystallography & mass spectrometry*

The four stages of protein X-ray crystallography where mass spectrometry has played an important facilitating role leading to a high resolution structure.

isolation and purification (discussed in this chapter). (2) Protein crystallization trials. Protein crystals can be directly analyzed by MALDI-MS (an example is given at the end of Chapter 2). (3) Domain elucidation (presented in this chapter). (4) Phase determination. Heavy atom derivatives (e.g., selenomethionine or mercury derivatized proteins) can be assayed by MALDI- or ESI-MS. These are extremely useful measurements since there are no other simple analytical methods available to assay heavy atom incorporation. (For example, this was particularly useful in the TAF 42/62 project). In summary, mass spectrometry has developed into an invaluable tool that should be included in the arsenal of strategies used for attaining high resolution X-ray crystallographic and NMR structures.

Bibliography

Abate, C., Baker, S.J., Lees-Miller, S.P., Anderson, C.W., Marshak, D.R., and Curran, T. (1993). Dimerization and DNA binding alter phosphorylation of Fos and Jun. *Proc. Natl. Acad. Sci. USA* **90,** 6766-6770.

Amati, B., Brooks, M.W., Levy, N., Littlewood, T.D., Evan, G.I., and Land, H. (1993). Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell* **72,** 233-245.

Anthony-Cahill, S.J., Benfield, P.A., Fairman, R., Wasserman, Z.R., Brenner, S.L., Stafford, W.F., 3d, Altenbach, C., Hubbell, W.L., and DeGrado, W.F. (1992). Molecular characterization of helix-loop-helix peptides. *Science* **255,** 979-983.

Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. USA* **88,** 10148-10152.

Ayer, D.E., Kretzner, L., and Eisenman, R.N. (1993). Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* **72,** 211-222.

Ayer, D.E., Lawrence, Q.A., and Eisenman, R.N. (1995). Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* **80,** 767-776.

Barber, M., Bordoli, R.S., Elliot, G.J., Sedgwick, R.D., and Tyler, A.N. (1981). Fast atom bombardment of solids (F.A.B.): A new ion source for mass spectrometry. *J. Chem. Soc. Chem. Commun.* 325-327.

Baron, C. and Thompson, T.E. (1975). Solubilization of bacterial membrane proteins using alkyl glucosides and dioctanoyl phosphatidylcholine. *Biochimica. et. Biophysica. Acta* **382,** 276-285.

Baxevanis, A.D. and Vinson, C.R. (1993). Interactions of coiled coils in transcription factors: where is the specificity?. *Curr.Opin.Genet.Dev.* **3,** 278-285.

Beavis, R.C. and Bridson, J.N. (1993). Epitaxial protein inclusion in sinapic acid crystals. *J. Phys. D: Appl. Phys.* **26,** 442-447.

Beavis, R.C. and Chait, B.T. (1989). Factors affecting the ultraviolet laser desorption of proteins. *Rapid Commun. Mass Spectrom.* **3,** 233-237.

Beavis, R.C. and Chait, B.T. (1990a). Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proc. Natl. Acad. Sci. USA* **87,** 6873-6877.

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

Beavis, R.C. and Chait, B.T. (1991). Investigations of matrix isolated, (UV) laser induced polymer sublimation using a time-of-flight mass spectrometer. In Methods and mechanisms for producing ions from large moelcuels. K.G. Standing and W. Ens, eds. (New York: Plenum Press), pp. 227-234.

Beavis, R.C. and Chait, B.T. (1996). Matrix-assisted laser desorption-ionization mass spectrometry of proteins. *Methods Enzymol.* **270,** 519-551.

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

Beckey, H.D. and Schulten, H.-R. (1975). Field desorption mass spectrometry. *Angew. Chem. Int. Ed. Engl.* **14,** 403-415.

Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA-binding proteins. *Cell* **61,** 49-59.

Benore-Parsons, M., Seidah, N.G., and Wennogle, L.P. (1989). Substrate phosphorylation can inhibit proteolysis by trypsin-like enzymes. *Arch. Biochem. Biophys.* **272,** 274-280.

Berberich, S.J. and Cole, M.D. (1992). Casein kinase II inhibits the DNA-binding activity of Max homodimers but not Myc/Max heterodimers. *Genes Dev.* **6,** 166-176.

Bernards, R. (1995). Transcriptional regulation. flipping the myc switch. *Curr. Biol.* **5,** 859-861.

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

Blackwood, E.M., Luscher, B., and Eisenman, R.N. (1992). Myc and Max associate in vivo. *Genes Dev.* **6,** 71-80.

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., and Ebright, R.H. (1994). Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78,** 889-896.

Bousset, K., Henriksson, M., Lüscher-Firzlaff, J.M., Litchfield, D.W., and Lüscher, B. (1993). Identification of casein kinase II phosphorylation sites in Max: effects on DNA-binding kinetics of Max homo- and Myc/Max heterodimers. *Oncogene* **8,** 3211-3220.

Brockerhoff, S.E., Edmonds, C.G., and Davis, T.N. (1992). Structural analysis of wild-type and mutant yeast calmodulins by limited proteolysis and electrospray ionization mass spectrometry. *Protein Sci.* **1,** 504-516.

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

Burley, S.K. and Roeder, R.G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Ann. Rev. of Biochem.* **65,** 769-799.

Burlingame, A.L., Boyd, R.K., and Gaskell, S.J. (1996). Fundamental Reviews: Mass Spectrometry. *Anal. Chem.* **68,** 599R-651R.

Chait, B.T. (1989). ^{252}Cf plasma desorption mass spectrometry. Contributions from the Rocefeller University. *Intl. J. Mass Spectrom. Ion Processes* **92,** 297-329.

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

Chait, B.T. (1994). Mass spectrometry--a useful tool for the protein X-ray crystallographer and NMR spectroscopist. *Structure* **2,** 465-467.

Chou, J.Z., Kreek, M.J., and Chait, B.T. (1994). Matrix-assisted laser desorption mass spectrometry of biotransformation products of Dynorphin A in vitro. *J. Am. Soc. Mass Spectrom.* **5,** 10-16.

Chowdhury, S.K., Katta, V., and Chait, B.T. (1990). An electrospray-ionization mass spectrometer with new features. *Rapid Commun. Mass Spectrom.* **4,** 81-87.

Cohen, S.L., Ferré-D'Amaré, A.R., Burley, S.K., and Chait, B.T. (1995). Probing the solution structure of the DNA-binding protein Max by a combination of proteolyis and mass spectrometry. *Protein Sci.* **4,** 1088-1099.

Cohen, S.L. (1996). Domain elucidation by mass spectrometry. *Structure* **4,** 1013-1016.

Cohen, S.L. and Chait, B.T. (1996). Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal. Chem.* **68,** 31-37.

Cohen, S.L. and Chait, B.T. (1997). Mass spectrometry of whole proteins from SDS-PAGE gles. *Anal. Biochem.* (In press).

Cohen, S.L., Halaas, J.L., Friedman, J.M., Chait, B.T., Bennett, L., Chang, D., Hecht, R., and Collins, F. (1996). Human leptin characterization. *Nature* **382,** 589.

Cole, M.D. (1986). The *myc* oncogene: its role in transformation and differentiation. *Ann. Rev. Genet.* **20,** 361-384.

Considine, R.V., Sinha, M.K., Heiman, M.L., Kriauciunas, A., Stephens, T.W., Nyce, M.R., Ohannesian, J.P., Marco, C.C., Mckee, L.J., Bauer, T.L., and Caro, J.F. (1996). Serum immunoreactive leptin concentrations in normal-weight and obese humans. *New Engl. J. Med.* **334,** 292-295.

Devos, R., Richards, J.G., Campfield, L.A., Tartaglia, L.A., Guisez, Y., van der Heyden, J., Travernier, J., Plaetinck, G., and Burn, P. (1996). Ob protein binds specifically to the choroid plexus of mice and rats. *Proc. Natl. Acad. Sci. USA.* **93,** 5668-5673.

Dzandu, J.K., Johnson, J.F., and Wise, G.E. (1988). Sodium dodecyl sulfate-gel elcetrophoresis: Staining of polypeptides using heavy metal salts. *Anal. Biochem.* **174,** 157-167.

Eisenberg, D. and Hill, C.P. (1989). Protein crystallography: more surprises ahead. *Trends Biochem. Sci.* **14,** 260-264.

Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S.C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* **8,** 970-980.

Feick, R.G. and Shiozawa, J.A. (1990). A high-yield method for the isolation of hydrophobic proteins and peptides from ployacrylamide gels for protein sequencing. *Anal. Biochem.* **187,** 205-211.

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

Ferré-D'Amaré, A.R., Prendergast, G.C., Ziff, E.B., and Burley, S.K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363,** 38-45.

Ferré-D'Amaré, A.R., Pognonec, P., Roeder, R.G., and Burley, S.K. (1994). Structure and function of the b/hlh/z domain of USF. *EMBO J.* **13,** 180-189.

Ferré-D'Amaré, A.R. (1994). DNA recognition by helix-loop-helix proteins. (Thesis; Rockefeller University, NYC.)

Ferré-D'Amaré, A.R. and Burley, S.K. (1997). Dynamic Light Scattering as a Tool for Evaluating Crystallizability of Macromolecules. *Methods Enzymol.* **276,** 157-166.

Fontana, A. (1989). Limited proteolysis of globular proteins occurs at exposed and flexible loops. In Highlights in Modern Biochemistry. A. Kotyk, J. Skoda, V. Paces, and V. Kostka, eds. (Zeist, The Netherlands: VSP), pp. 1711-1726.

Fruton, J.S. (1975). The specificity of proteinases toward protein substrates. In Proteases and biological control. E. Reich, D.B. Rifkin, and E. Shaw, eds. (New York: Cold Spring Harbor Laboratory), pp. 33-50.

Goodrich, J.A. and Tjian, R. (1994). TBP-TAF complexes: selectivity factors for eukaryotic transcription. *Curr. Opin. Cell Biol.* **6,** 403-409.

Gusev, A.I., Wilkinson, W.R., Proctor, A., and Hercules, D.M. (1995). Improvement of signal reproducibility and matrix/comatrix effects in MALDI analysis. *Anal. Chem.* **67,** 1034-1041.

Hager, D.A. and Burgess, R.R. (1980). Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: Results with sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.* **109,** 76-86.

Halaas, J.L., Gajiwala, K.S., Maffei, M., Cohen, S.L., Chait, B.T., Rabinowitz, D., Lallone, R.L., Burley, S.K., and Friedman, J.M. (1995). Weight-reducing effects of the plasma protein encoded by the *obese* gene. *Science* **269,** 543-546.

Hames, B. D., and Rickwood, D. (Eds.) (1989). Proteolytic enzymes: a practical approach (Oxford: IRL Press).

Harrington, W.F., Von Hippel, P.H., and Mihalyi, E. (1959). Proteolytic enzymes as probes of the secondary structure of fibrous proteins. *Biochem. Biophys. Acta.* **32,** 303-304.

Henderson, L.E., Oroszlan, S., and Konigsberg, W. (1979). A micromethod for complete removal of dodecyl sulfate from proteins by ion-pair extraction. *Anal. Biochem.* **93,** 153-157.

Heukeshoven, J. and Dernick, R. (1982). Reversed-phase high-performance liquid chromatography of virus proteins and other large hyrdrophobic proteins in formic acid containing solvents. *J. Chromatogr.* **252,** 241-254.

Hillenkamp, F., Karas, M., Beavis, R.C., and Chait, B.T. (1991). Matrix-assisted laser desorption ionization mass spectrometry of biopolymers. *Anal. Chem.* **63,** 1193a-1203a.

Hoffmann, A., Chiang, C.M., Oelgeschlager, T., Xie, X., Burley, S.K., Nakatani, Y., and Roeder, R.G. (1996). A histone octamer-like structure within TFIID. *Nature* **380,** 356-359.

Hubbard, S.J., Eisenmenger, F., and Thornton, J.M. (1994). Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci.* **3,** 757-768.

Joachimiak, A. and Sigler, P.B. (1991). Crystallization of protein-DNA complexes. *Methods Enzymol.* **208,** 82-99.

Juhasz, P. and Biemann, K. (1994). Mass spectrometric molecular-weight determination of highly acidic compounds of biological significance via their complexes with basic polypeptides. *Proc. Natl. Acad. Sci. USA* **91,** 4333-4337.

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

Kato, G.J., Lee, W.M., Chen, L.L., and Dang, C.V. (1992). Max: functional domains and interaction with c-Myc. *Genes Dev.* **6,** 81-92.

Katta, V. and Chait, B.T. (1991). Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Amer. Chem. Soc.* **113,** 8534-8535.

Katta, V., Chowdhury, S.K., and Chait, B.T. (1991). Use of a single-quadrupole mass spectrometer for collision-induced dissociation studies of multiply charged produced by electrospray ionization. *Anal. Chem.* **63,** 174-178.

Keil, B. (1992). Specificity of Proteolysis (Berlin: Springer-Verlag).

Klarskov, K. and Roepstorff, P. (1993). Plasma desorption mass spectrometry of proteins transferred from gels after sodium dodecyl sulphate-polyacrylamide gel electrophoresis. *Biol. Mass Spectrom.* **22,** 433-440.

Kokubo, T., Gong, D.W., Wootton, J.C., Horikoshi, M., Roeder, R.G., and Nakatani, Y. (1994). Molecular cloning of Drosophila TFIID subunits. *Nature* **367,** 484-487.

Koskinen, P.J., Vastrik, I., Makela, T.P., Eisenman, R.N., and Alitalo, K. (1994). Max activity is affected by phosphorylation at two NH2-terminal sites. *Cell Growth & Differentiation* **5,** 313-320.

Kretzner, L., Blackwood, E.M., and Eisenman, R.N. (1992). Myc and max proteins possess distinct transcriptional activities. *Nature* **359,** 426-429.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 682-685.

Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240,** 1759-1764.

le Maire, M., Deschamps, S., Moller, J.V., Le Caer, J.P., and Rossier, J. (1993). Electrospray ionization mass spectrometry on hydrophobic peptides eluted from sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Application to the topology of the sarcoplasmic reticulum Ca^{2+} ATPase. *Anal. Biochem.* **214,** 50-57.

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

Lee, C.H., Saksela, K., Mirza, U.A., Chait, B.T., and Kuriyan, J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell* **85,** 931-942.

Liang, X., Bai, J., Liu, Y.-H., and Lubman, D.M. (1996). Characterization of SDS-PAGE-separated proteins by matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **68,** 1012-1018.

Liao, P.-C. and Allison, J. (1994). Enhanced detection of peptides in matrix-assisted laser desorption/ionization mass spectrometry through the use of charged-localized derivatives. *J. Mass. Spectrom.* **30,** 511-512.

Linderstrøm-Lang, K., Hotchkiss, R.D., and Johansen, G. (1938). Peptide bonds in globular proteins. *Nature* **142,** 996.

Lindsey, J.S., Chaudhary, T., and Chait, B.T. (1992). ^{252}Cf Plasma desorption mass spectrotmetry in the synthesis of porphyrin model systems. *Anal. Chem.* **64,** 2804-2814.

Littlewood TD and Evan, G. (1994). Transcription factors 2:Helix-loop-helix. *Protein Profile* **1,** 639-709.

Longman, J. (1996). Drug Lab Passes Test for Atlanta Drug Test. *The New York Times,* July 6, 1996.

Loo, R.R.O., Stevenson, T.I., Mitchell, C., Loo, J.A., and Andrews, P.C. (1996a). Mass spectrometry of proteins directly from polyacrylamide gels. *Anal. Chem.* **68,** 1910-1917.

Loo, R.R.O., Mitchell, C., Andrews, P.C., Stevenson, T.I., and Loo, J.A. (1996b). MALDI Mass Measurement Accuracy of Proteins Directly from Polyacrylamide Gels and Membranes. *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics; Portland, OR.* 350.

Ma, P.C., Rould, M.A., Weintraub, H., and Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77,** 451-459.

Maffei, M., Halaas, J., Ravussin, E., Pratley, R.E., Lee, G.H., Zhang, Y., Fei, H., Kim, S., Lallone, R., Ranganathan, S., Kern, P.A., and Friedman, J.M. (1995). Leptin levels in human and rodent: measurement of plasma leptin and *ob* RNA in obese and weight-reduced subjects. *Nature Med.* **1,** 1155-1161.

Maffei, M., Stoffel, M., Barone, M., Moon, B., Dammerman, M., Ravussin, E., Bogardus, C., Ludwig, D.S., Flier, J.S., Talley, M., Auerbach, S., and Friedman, J.M. (1996). Absence of mutations in the Human OB gene in obese/diabetic subjects. *Diabetes* **45,** 679-682.

Mann, M., Meng, C.K., and Fenn, J.B. (1989). Interpreting mass spectra of multiply charged ions. *Anal. Chem.* **61,** 1702-1708.

Marcu, K.B., Bossone, S.A., and Patel, A.J. (1992). Myc function and regulation. *Annu. Rev. Biochem.* **61,** 809-860.

Martin de Llano, J.J., Jones, W., Schneider, K., Chait, B.T., Manning, J.M., Rodgers, G., Benjamin, L.J., and Weksler, B. (1993). Biochemical and functional properties of recombinant human sickle hemoglobin expressed in yeast. *J. Biol. Chem.* **28,** 27004-27011.

Martin, J.F., Schwarz, J.J., and Olson, E.N. (1993). Myocyte enhancer factor (Mef) 2c: a tissue-restricted member of the Mef-2 family of transcription factors. *Proc. Natl. Acad. Sci. USA* **90,** 5282-5286.

McLafferty, F.W. and Turecek, F. (1993). Interpretation of Mass Spectra (Mill Valley, CA: University Science Books).

Meisner, H. and Czech, M.P. (1991). Phosphorylation of transcriptional factors and cell-cycle- dependent proteins by casein kinase II. *Curr. Opin. Cell Biol.* **3,** 474-483.

Mihalyi, E. (1978). Application of Proteolytic Enzymes to Protein Structure Studies, 2nd Edition (Boca Raton, Fl: CRC Press, Inc.).

Mirza, U.A., Chait, B.T., and Landers, H.M. (1995). Monitoring reactions of nitric oxide with peptides and proteins by electrospray ionization-mass spectrometry. *J. Biol. Chem.* **270,** 17185-17188.

Mock, K.K., Sutton, C.W., and Cottrell, J.S. (1992). Sample immobilization protocols for matrix-assisted laser-desorption mass spectrometry. *Rapid Commun. Mass Spectrom.* **6,** 233-238.

Munson, M.S.B. and Field, F.H. (1966). Chemical ionization mass spectrometry. I. General Introduction. *J. Amer. Chem. Soc.* **88,** 2621-2630.

Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobin enhancer binding, *daughterless, myoD,* and *myc* proteins. *Cell* **56,** 777-783.

Nakayama, H., Uchida, K., Shinkai, F., Shinoda, T., Okuyama, T., Seta, K., and Isobe, T. (1996). Capillary column high-performance liquid chromatographic electrospray ionization triple-stage quadrupole mass spectrometric analysis of proteins separated by two-dimensional polyacrylamide gel electrophoresis - application to cerebellar protein mapping. *J. Chromatogr.* **730,** 279-287.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA element ternary complex. *Nature* **377,** 119-128.

Novotny, J. and Bruccoleri, R.E. (1987). Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Letters* **211,** 185-189.

Ortiz, M.L., Calero, M., Patron, C.F., Castellanos, L., and Mendez, E. (1992). Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified elctroblotted proteins. *FEBS Lett.* **296,** 300-304.

O'Shea, E.K., Rutkowski, R., and Kim, P.S. (1989). Evidence that the leucine zipper is a coiled coil. *Science* **243,** 538-542.

Pellegrini, L., Tan, S., and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. *Nature* **376,** 490-498.

Pelleymounter, M.A., Cullen, M.J., Baker, M.B., Hecht, R., Winters, D., Boone, T., and Collins, F. (1995). Effects of the *obese* gene product on body weight regulation in *ob/ob* mice. *Science* **269,** 540-543.

Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. *Cell* **65,** 395-407.

Prendergast, G.C., Hopewell, R., Gorham, B.J., and Ziff, E.B. (1992). Biphasic effect of Max on Myc cotransformation activity and dependence on amino- and carboxy-terminal Max functions. *Genes Dev.* **6,** 2429-2439.

Przybylski, M. and Glocker, M.O. (1996). Electrospray mass spectrometry of biomacromolecular complexes with noncovalent interactions - new analytical perspectives for supramolecular chemistry and molecular recognition processes. *Angew. Chem. Int. Ed. Engl.* **35,** 806-826.

Qin, J. (1996). Trapping gene products: New tools for studying proteins. (Thesis; Rockefeller University, NYC.)

Qin, J., Wang, Y., and Chait, B.T. (1996a). Cooper stain provides an effective interface bewteen gel electrophoresis ans mass spectrometry. *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics; Portland, OR.* 333.

Qin, J., Zhao, Y., Fenyo, D., and Chait, B.T. (1996b). High confidence identification of gel-separated proteins. *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics; Portland, OR.* 334.

Rabbitts, T.H. (1994). Chromosomal translocations in human cancer. *Nature* **372,** 143-149.

Record, M.T., Jr., Ha, J.H., and Fisher, M.A. (1991). Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol.* **208,** 291-343.

Reddy, C.D., Dasgupta, P., Saikumar, P., Dudek, H., Rauscher, F.J., 3d, and Reddy, E.P. (1992). Mutational analysis of Max: role of basic, helix-loop-helix/leucine zipper domains in DNA binding, dimerization and regulation of Myc-mediated transcriptional activation. *Oncogene* **7,** 2085-2092.

Sawadogo, M. and Roeder, R.G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA-box region. *Cell* **43,** 165-175.

Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27,** 157-162.

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

Schreiber-Agus, N., Chin, L., Chen, K., Torres, R., Rao, G., Guida, P., Skoultchi, A.I., and DePinho, R.A. (1995). An amino-terminal domain of Mxi1 mediates anti-Myc oncogenic activity and interacts with a homolog of the yeast transcriptional repressor Sin3. *Cell* **80,** 777-786.

Schuhmacher, M., Glocker, M.O., Wunderlin, M., and Przybylski, M. (1996). Direct isolation of proteins from sodium dodecyl sulfate-polyacrylamide gel electrophoresis and analysis by electrospray-ionization mass spectrometry. *Electrophoresis* 17, 848-854.

Severinova, E., Severinova, K., Fenyö, D., Marr, M., Brody, E.N., Roberts, J.W., Chait, B.T., Darst, S.A. (1996). Domain organization of the *Escherichia coli* RNA polymerase σ^{70} subunit. *J. Mol. Biol.* **263,** 637-647.

Sha, M., Ferré-D'Amaré, A.R., Burley, S.K., and Goss, D.J. (1995). Anti-cooperative biphasic equilibrium binding of transcription factor upstream stimulatory factor to its cognate DNA monitored by protein fluorescence changes. *J. Biol. Chem.* **270,** 19325-19329.

Sillero, A. and Ribeiro, J.M. (1989). Isoelectric points of proteins: theoretical determination. *Anal. Biochem.* **179,** 319-325.

Smith, R.D. and Light-Wahl, K.J. (1993). The observation of non-covalent interactions in solution by electrospray ionization mass spectrometry: Promise, pitfalls and prognosis. *Biol. Mass Spectrom.* **22,** 493-501.

Sorensen, S.B., Sorensen, T.L., and Breddam, K. (1991). Fragmentation of proteins by S. aureus strain V8 protease. ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid. *FEBS Letters* **294,** 195-197.

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

Starovasnik, M.A., Blackwell, T.K., Laue, T.M., Weintraub, H., and Klevit, R.E. (1992). Folding topology of the disulfide-bonded dimeric DNA-binding domain of the myogenic determination factor MyoD. *Biochemistry* **31,** 9891-9903.

Strupat, K., Karas, M., Hillenkamp, F., Eckerskorn, C., and Lottspeich, F. (1994). Matrix-assisted laser desorption ionization mass spectrometry of proteins electroblotted after polyacrylamide gel electrophoresis. *Anal. Chem.* **66,** 464-470.

Sun, X.H. and Baltimore, D. (1991). An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell* **64,** 459-470.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988). Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2,** 151-153.

Tang, K., Allman, S.L., Jones, R.B., and Chen, C.H. (1993). Quantitative analysis of biopolymers by matrix-assisted laser desorption. *Anal. Chem.* **65,** 2164-2166.

Torgerson, D.F., Skowronski, R.P., and Macfarlane, R.D. (1974). New approach to the mass spectroscopy of non-volatile compounds. *Biochem. Biophys. Res. Commun.* **60,** 616-621.

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

Vestling, M.M. and Fenselau, C. (1994). Poly(vinylidene difluoride) membranes as the interface between laser desorption mass spectrometry, gel electrophoresis, and in situ proteolysis. *Anal. Chem.* **66,** 471-477.

Vorm, O., Roepstorff, P., and Chait, B.T. (1993). Mass spectrometry of protein samples containing detergent. *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*; San Francisco, CA. 621.

Vorm, O., Roepstorff, P., and Mann, M. (1994). Improved resolution and very high sensitivity in MALDI-TOF of matrix surfaces made by fast evaporation. *Anal. Chem.* **66,** 3281-3287.

Wang, X., Sato, R., Brown, M.S., Hua, X., and Goldstein, J.L. (1994). SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis. *Cell* **77,** 53-62.

Weinberger, S.R., Boernsen, K.O., Finchy, J.W., Robertson, V., and Musselman, B.D. (1993). An evaluation of crystallization methods for matrix-assisted laser desorption/ionization of proteins. *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics; San Francisco*, CA. 775.

Whittal, R.M. and Liang, L. (1995). High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer. *Anal. Chem.* **67,** 1950-1954.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379,** 466-469.

Xiang, F. and Beavis, R.C. (1993). Growing protein-doped sinapinic acid drystals for laser desorption: An alternative preparation method for difficult samples. *Org. Mass Spectrom.* **28,** 1424-1429.

Xiang, F. and Beavis, R.C. (1994). A method to increase contaminant tolerance in protein matrix-assisted laser desorption/ionization by the fabrication of thin protein-doped polycrystalline films. *Rapid Commun. Mass Spectrom.* **8,** 199-204.

Xie, X., Kokubo, T., Cohen, S.L., Mirza, U.A., Hoffmann, A., Chait, B.T., Roeder, R.G., Nakatani, Y., and Burley, S.K. (1996). Crystal structure of the $dTAF_{II}42/dTAF_{II}62$ complex resembles the histone $(H3/H4)_2$ heterotetramer. *Nature* **380,** 316-322.

Yates III, J.R., McCormack, A.L., and Eng, J. (1996). Mining genomes with MS. *Anal. Chem.* **68,** 534A-540A.

Yokoyama, C., Wang, X., Briggs, M.R., Admon, A., Wu, J., Hua, X., Goldstein, J.L., and Brown, M.S. (1993). SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene. *Cell* **75,** 187-197.

Zervos, A.S., Gyuris, J., and Brent, R. (1993). Mxi1, a protein that specifically interacts with Max to bind Myc- Max recognition sites. *Cell* **72,** 223-232.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., and Friedman, J.M. (1994). Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372,** 425-432.

End