Diss. ETH No. 12949

A Novel Method for N-terminal Ladder Sequencing and Subtractive Proteome Analysis of Oocytes from *Asterina pectinifera* after Resumption of Meiosis

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH
for the degree of Doctor of Natural Sciences

presented by

SJOUKE HOVING

M.Sc. Weizmann Institute of Science (Rehovot, Israel)
born January 9, 1966
(Vlagtwedde, The Netherlands)

Accepted on the recommendation of
Prof. Dr. Ernesto Carafoli, examiner
Prof. Dr. Josef Brunner, co-examiner
Dr. Peter James, co-examiner

Zürich - 1998

Summary

Recent developments in protein separation techniques and protein identification techniques have dramatically changed biology. New methods for high throughput protein identification methods are required. This thesis is divided in three sections and describes these advanced techniques in detail.

In the section "General Introduction" an overview about proteome research and the required techniques for proteome analysis is given. The proteome defines protein-based gene expression analysis and thus indicates the PROTEins expressed by a genOME or tissue under specified conditions. In order to analyse the proteome, methods that are able to separate and identify thousands of proteins at high resolution are required. Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is currently the only technique available for the separation of complex protein mixtures from cells or tissue. The speed and sensitivity of protein identification is accelerated by the development of new ionization techniques in mass spectrometry (MS). Using these new MS techniques it is possible to accurately measure the masses of peptides produced by specific protease digestions and match these against theoretically protein digests in a protein sequence database. This technique is known as *peptide mass fingerprinting*. Another technique known as *peptide fragment mass fingerprinting* produces (partial) sequence information that can be used for protein identification.

Although being powerful methods, in section "Part A: Partial N-terminal Sequencing" a novel method to generate a short amino acid sequence tag or sequence ladder as a tool in protein identification is described for high throughput. Protein digests are immobilized on reversed-phase material and partially degraded in parallel from the N-termini with a thioacetylating reagent. The partially degraded peptides are then eluted from the reversed-phase material and the peptide masses are accurately measured by MALDI-MS. For the identification, a new algorithm MassDynSearch has been developed. The method is applied to a set of ribosomal proteins isolated from *Escherichia coli*. As a control, independent analysis of the proteins by *peptide fragment mass fingerprinting* and using the program SEQUEST identified all the ribosomal proteins. Using the novel partial

degradation method, 50% of the proteins could be identified, compared to only 27% by *peptide mass fingerprinting* alone.

Finally, the section "Part B: 2-D Gel Analysis of Oocytes from Starfish" describes an example of a subtractive proteome analysis. The effect of the natural hormone 1-methyladenine (1-MA) on oocytes from *Asterina pectinifera* at the protein level is studied. It is known that upon addition of 1-MA to prophase arrested oocytes meiosis is resumed, this is accompanied with the breakdown of the nucleus. In this study, 2-D gels of control oocytes and oocytes treated with the hormone were compared. Analysis of some proteins that are degraded upon addition of 1-MA are identified as cytoskeleton related proteins. Further experiments showed that some of these proteins might be targets for the calcium dependent protease calpain. This could imply an important role for calcium in the cell nucleus, a highly debated issue in the literature.

Zusammenfassung

Neue Entwicklungen auf dem Gebiet der Proteinseparation und Proteinidentifikation haben zu einer signifikanten Änderung in der modernen Biologie geführt. Hierbei werden neue Screeningmethoden zur parallelen und schnellen Identifikation von Proteinen immer bedeutsamer. Diese Arbeit beschreibt ausführlich in drei Sektionen die Entwicklung neuer relevanter Techniken.

In der Sektion "Allgemeine Einführung" wird ein Überblick über die Proteomforschung und die hierzu benötigten Techniken zur Proteomanalyse gegeben. Unter Proteom ist eine Analyse der Genexpression auf Proteinebene zu verstehen und steht als Synonym für alle PROTEine exprimiert von einem GenOM, respektive eines Gewebes unter den spezifischen Bedingungen. Um ein Proteom zu analysieren, sind Techniken zur Separation und Identifikation tausender Proteine bei hoher Auflösung notwendig. Zur Zeit bietet die zweidimensionale Polyacrylamid Gelelectrophorese (2-D PAGE) die vielversprechendste Möglichkeit, komplexe Proteingemische von Zellen oder einzelner Gewebe aufzutrennen. Die Schnelligkeit und Sensitivität der Proteinidentifikation wurde durch die Entwicklung neuer Ionisationsmethoden der Massenspektrometrie (MS) beschleunigt. Unter Benutzung dieser neuen MS-Techniken ist es nun möglich, die Massen von Peptiden, die durch proteolytischen Abbau erhalten worden sind, zu messen und gegen eine Datenbank zu vergleichen. Man spricht hier von einem *peptide mass fingerprint*. Eine andere Technik, die als *peptide fragment mass fingerprinting* bekannt ist, erzeugt eine partielle Sequenzinformation, die ebenfalls zur Proteinidentifikation benutzt werden kann.

In der Sektion "Teil A: Partielle N-terminale Sequenzierung" wird eine neue Methodik zur Generierung kurzer Aminosäuresequenzmarkern oder kurzer Aminosäureleitern vorgestellt. Hierbei wird ein Peptidverdau auf hydrophobem Material immobilisiert und danach partiell und parallel vom N-terminus ausgehend über eine Thioacetylierungsreaktion degradiert. Die partiell abgebauten Peptide werden von der hydrophoben Oberfläche eluiert und die Peptidmassen mittels MALDI-MS bestimmt. Um eine Identifikation zu ermöglichen, wurde ein neuer Algorithmus MassDynSearch entwickelt. Angewandt wurde diese Methode erstmalig zur Bestimmung der ribosomalen Proteine aus *Escherichia coli*. Eine hiervon

unabhängig durchgeführte Analyse der Proteine mittels *peptide fragment fingerprinting* und SEQUEST führte zur Identifikation aller ribosomaler Proteine. Durch die neue Degradationsmethode konnten 50% der Proteine identifiziert werden, wohingegen durch *peptide fragment fingerprinting* lediglich 27% erhalten werden konnte.

Die letzte Sektion "Teil B: 2-D Gel Analyse der Oozyten vom Seestern" beschreibt ein Beispiel der subtraktiven Proteomanalyse. Hierbei wurde der Effekt des natürlichen Hormons 1-Methyladenine (1-MA) auf Oozyten von *Asterina pectinifera* auf Proteinebene untersucht. Es ist bekannt, daß bei Zugabe von 1-MA in der Prophase verweilende Oozyten die Meiose wieder aufnehmen, was mit einem Abbau des Nukleus einhergeht. In dieser Studie wurden 2-D Gele von Kontroll Oozyten mit solchen von hormonbehandelten Oozyten verglichen. Einige Proteine, die nach Zugabe von 1-MA abgebaut wurden, konnten als zytoskeletale Proteine identifiziert werden. Weitere Experimente zeigten, daß diese Proteine potentielle Angriffobjekte der kalziumabhängigen Protease Calpain sein könnten. Dies könnte für eine wichtige Rolle von Kalzium im Zellnukleus sprechen.

Riassunto

I recenti sviluppi nelle tecniche di separazione e di identificazione delle proteine hanno cambiato sensibilmente la ricerca in campo biologico. Si è perciò reso necessario lo sviluppo di nuove metodiche per identificare in modo rapido un numero elevato di proteine. Questa tesi è suddivisa in tre sezioni e descrive dettagliatamente queste tecniche avanzate.

Nella sezione "Introduzione Generale" viene presentata una panoramica della ricerca relativa al proteoma e delle tecniche indispensabili per la sua analisi. Un proteoma definisce l'insieme delle PROTEine espresse da un genOMA o da un tessuto in condizioni specifiche. Per poter analizzare un proteoma sono necessari metodi in grado di separare ed identificare ad alta risoluzione migliaia di proteine. L'elettroforesi bidimensionale su gel di poliacrilammide (2-D PAGE) è attualmente l'unica tecnica in grado di separare miscele complesse di proteine derivanti da cellule o da tessuti. La rapidità e la sensibilità nell'identificazione delle proteine è accelerata dallo sviluppo di nuovi metodi di ionizzazione nel campo della spettrometria di massa (MS). Utilizzando le nuove tecniche di spettrometria di massa è possibile misurare in modo accurato le masse dei peptidi prodotti per digestione proteica specifica e correlarle con i valori teorici relativi a digestioni di proteine contenuti in banche dati di sequenza. Questa tecnica è chiamata *peptide mass fingerprinting*. Un'altra tecnica nota come *peptide fragment mass fingerprinting* può fornire informazioni sulla sequenza (parziale) di una proteina che possono essere utili per la sua identificazione.

Nonostante i metodi appena citati siano di notevole validità, nella sezione "Parte A: Sequenziamento N-terminale Parziale" viene anche riportato, per la sua particolare rapidità, un nuovo metodo per generare una breve sequenza di amminoacidi (*sequence tag* o *sequence ladder*) come strumento di investigazione nell'identificazione di proteine. Le digestioni di proteine vengono immobilizzate su di un supporto solido a fase inversa e il gruppo ammino terminale di ogni peptide viene fatto reagire con un agente tioacetilante. I peptidi parzialmente reagiti vengono quindi eluiti dalla resina e le masse dei peptidi accuratamente misurate per mezzo della spettrometria di massa (MALDI-MS). Per l'identificazione delle proteine è stato sviluppato un nuovo algoritmo chiamato MassDynSearch. Questo metodo è stato applicato ad una serie di proteine ribosomiali isolate da *Escherichia coli*. L'analisi indipendente delle proteine attraverso il metodo

del *peptide fragment mass fingerprinting* ha consentito di identificare tutte le proteine ribosomiali in esame utilizzando il programma SEQUEST. Con l'utilizzo del nuovo metodo di sequenziamento parziale è stato possibile identificare il 50% delle proteine, mentre solo il 27% utilizzando il metodo del *peptide mass fingerprinting*.

La sezione "Parte B: 2-D Gel Analisi di Ovociti di Stella di Mare" descrive un esempio di analisi di un proteoma. E' stato investigato l'effetto dell'ormone naturale 1-metiladenina (1-MA) sugli ovociti di *Asterina pectinifera* a livello di espressione proteica. E' noto come in seguito all'aggiunta di 1-MA ad ovociti bloccati a livello di profase si ha una ripresa della meiosi. Questo effetto è accompagnato da una rottura del nucleo cellulare. Nel corso di questa ricerca sono stati messi a confronto 2-D gels di ovociti di controllo e di ovociti trattati con l'ormone. L'analisi di alcune delle proteine che vengono degradate in seguito all'aggiunta di 1-MA ne ha reso possibile l'identificazione come di proteine collegate al citoscheletro. Ulteriori esperimenti hanno evidenziato che alcune di queste proteine potrebbero rappresentare un bersaglio della calpaina, una proteasi calcio dipendente. Questi risultati indicherebbero una possibile funzione del calcio a livello di nucleo cellulare, argomento questo tuttora al centro di un vivace dibattito.

Samenvatting

Recente ontwikkelingen in technieken om eiwitten te scheiden en te identificeren hebben de moderne biologie drastisch veranderd. Nieuwe methoden om grote aantallen eiwitten te kunnen identificeren zijn noodzakelijk. Dit proefschrift bestaat uit drie delen en beschrijft deze nieuwe methoden om grote aantallen eiwitten te kunnen identificeren in detail.

In het eerste deel "Algemene Inleiding" wordt een overzicht van het proteoom onderzoek en de benodigde technieken voor de analyse van het proteoom gepresenteerd. Het proteoom is gedefinieerd als de PROTEinen (eiwitten) die tot expressie gebracht worden door een genOOM of weefsel onder bepaalde (fysiologische) omstandigheden. Het definieert de op eiwit-gebaseerde analyse van gen expressie. Om het proteoom te kunnen analyseren moeten methoden beschikbaar zijn om duizenden eiwitten te kunnen scheiden en te identificeren met een hoog oplosssend vermogen. Twee-dimensionale polyacrylamide gel electroforese (2-D PAGE) is momenteel de enige beschikbare methode om complexe mengsels van eiwitten uit cellen of weefsel van elkaar te scheiden. De snelheid en gevoeligheid waarmee eiwitten geïdentificeerd kunnen worden is in een stroomversnelling gekomen door de ontwikkeling van nieuwe ionisatie technieken in de massaspectrometrie (MS). Met deze nieuwe MS technologie is het mogelijk om zeer nauwkeurig de massa's van peptiden, die gegenereerd worden na behandeling met een specifiek protease, te meten. Via een speciaal zoekprogramma wordt vervolgens bepaald of er een eiwit in de database voorkomt dat deze massa's na digestie (theoretisch) zou opleveren. Deze techniek heet *peptide mass fingerprinting*. Een andere methode is *peptide fragment mass fingerprinting* en geeft informatie over de gedeeltelijke aminozuur volgorde van een eiwit, dat gebruikt kan worden voor de identificatie van het volledige eiwit.

Hoewel de hierboven beschreven methoden indrukwekkend zijn, wordt in het tweede deel "Deel A: Partiële N-terminale Sequencing" een nieuwe methode beschreven, waarmee grote aantallen eiwitten kunnen worden geïdentificeerd via een korte aminozuur volgorde, een zogenaamde *sequence tag* of *sequence ladder*. Digesties van eiwitten worden in parallel geïmmobilizeerd op reversed-phase materiaal en partieel gedegradeerd vanaf de N-terminus met een thioacetylerend reagens. De partieel

gedegradeerde peptiden worden van het reversed-phase materiaal geëlueerd en vervolgens worden de massa's van de peptiden zeer nauwkeurig met behulp van MALDI-MS bepaald. Voor de identificatie is een nieuw algorithme, genaamd MassDynSearch, ontwikkeld. De methode is toegepast op een set ribosomale eiwitten, die geïsoleerd zijn uit de darmbakterie *Escherichia coli*. Onafhankelijke analyse van de eiwitten met behulp van de *peptide fragment mass fingerprinting* en het programma SEQUEST ter controle, leidde tot identificatie van alle ribosomale eiwitten. Met de nieuwe gedeeltelijke degradatie methode konden 50% van de eiwitten worden geïdentificeerd, in vergelijking met slechts 27% door *peptide finger printing* alleen.

In het laatste deel "Deel B: 2-D Gel Analyse van Oöcyten van de Zeester" wordt een voorbeeld van een zogenaamde subtractieve proteoom analyse beschreven. Het effect van het natuurlijke hormoon 1-methyladenine (1-MA) op oöcyten van de zeester *Asterina pectinifera* is op eiwit niveau bestudeerd. Uit de literatuur is bekend dat na toevoeging van 1-MA aan oöcyten, die in de pro-fase geblokkeerd zijn, de meiose wordt hervat. Dit gaat gepaard met het afbreken van de celkern. In deze studie zijn 2-D gelen van normale oöcyten vergeleken met oöcyten, die behandeld zijn met het hormoon. De analyse van enkele eiwitten die zijn afgebroken na toevoeging van 1-MA konden geïdentificeerd worden als eiwitten, die met het cytoskeleton te maken hebben. Andere experimenten hebben aangetoond dat enige van deze eiwitten als substraat voor het calcium afhankelijke protease calpain kunnen dienen. Dit zou een grotere rol voor calcium in de celkern kunnen betekenen, wat een zeer omstreden kwestie in de huidige literatuur is.

Abbreviations

1-MA	1-methyladenine
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
α-CHC	α-cyano-4-hydroxy-cinnamic acid
a.m.u.	atomic mass unit
API	atmospheric pressure ionization
bp	base pair
cADPr	cyclic adenosine diphosphate ribose
CaM	calmodulin
CaM-kinase II	type II Ca^{2+}-calmodulin dependent protein kinase
CANP	Ca^{2+}-activated neutral protease (calpain)
cdk	cyclin dependent protein kinase
CHAPS	3-[(3-cholamideopropyl)dimethylammonio]-1-propane sulphonate
CID	collisionally induced dissociation
DE	delayed extraction
DTT	dithiothreitol
ESI	electrospray ionization
EST	expressed sequence tag
eV	electronvolt
HPLC	high performance liquid chromatography
IEF	isoelectric focusing
$InsP_3$	inositol 1,4,5-triphosphate
IPG	immobilized pH gradient
IT	ion trap
kDa	kilodalton
MALDI	matrix assisted laser desorption/ionization
MH^+-ion	single charged positive ion
MH^{2+}-ion	double charged positive ion
MPF	maturation or M phase promoting factor
M_r	relative molecular mass
MS	mass spectrometry, mass spectrometer
MS/MS	tandem mass spectrometry
m/z	mass to charge ratio
NC	nitrocellulose
NEM	N-ethylmorpholine

NMP	N-methylpiperidine
NMR	nuclear magnetic resonance
NP-40	Nonidet P-40
PCR	polymerase chain reaction
pI	isoelectric point
PIC	phenylisocyanate
PITC	phenylisothiocyanate
ppm	parts per million
PSD	post-source decay
PTFE	poly(tetrafluorethylene)
PTH	phenylthiohydantoin
PVDF	polyvinylidene difluoride
Q-TOF	quadrupole time-of-flight
SAGE	serial analysis of gene expression
SDS	sodium dodecylsulphate
TATE	thioacetylthioethane
TATG	thioacetylthioglycolic acid
TBP	tributylphosphine
TBTGA	S-(thiobenzoyl)thioglycolic acid
TEA	triethylamine
TEMED	N,N,N',N'-tetramethylethylenediamine
TFA	trifluoroacetic acid
TMA	trimethylamine
TOF	time-of-flight
Tris	Tris(hydroxymethyl)aminomethane
W W W	World-Wide Web

1. General Introduction

In this general introduction an overview is given about proteome research. The concept of the proteome is explained and the techniques for protein separation at high resolution and protein identification at a large scale are discussed. The results of proteome analysis show which genes are expressed, how the protein products are modified, and how they interact. This makes proteome research of fundamental importance for the biologist, the clinician and the pharmaceutical industry.

1.1 Introduction to the proteome

1.1.1 The proteome, a new word

Organisms as different as bacteria and human beings use the same building blocks to construct divergent macromolecules such as proteins and the flow of genetic information from DNA to mRNA to proteins is essentially the same in all organisms. Proteins play a crucial role in nearly all biological processes. Their significance and broad range of activity are shown in the following examples, such as enzymatic catalysis, transport and storage, co-ordinated motion, mechanical support, immune protection and control of growth and differentiation. The prominent position occupied by proteins in biological systems was correctly surmised by the Dutch chemist Gerardus Mulder, who in 1838, on the suggestion of Jons Jakob Berzelius, introduced the term "protein", derived from the Greek word "proteios", meaning "of the first rank". Mulder commented:

> *"There is present in plants and animals a substance which ..... is without doubt the most important of the known substances in living matter, and without it, life would be impossible on our planet. This material has been named Protein."*

Not a bad name when one considers all those functions proteins have in the living organism (Zubay et al., 1995).

A large part of this thesis deals with some novel techniques that are used to be able to identify large numbers of proteins. Recent advances in two-dimensional gel electrophoresis, protein microanalysis and

bioinformatics have made the large-scale systematic analysis of proteins from any organism or tissue possible. This is what "Proteome Research" is about. The concept of the proteome was recently proposed to define protein-based gene expression analysis. The proteome indicates the PROTEins expressed by a genOME or tissue under specified conditions (Wilkins et al., 1995). Although it was first used in 1994 at the Siena 2-D Electrophoresis meeting, the term proteome is already widely accepted and proteomics, the techniques involved in this large scale protein identification in a fast way at low picomole level is rapid rising and causing a revolution in biology.

1.1.2 Entering the post genome era

Each organism has its own complement of genes which provide the information how the organism is constructed. Since DNA sequencing became rapid and simple and can be fully automated, sequencing of the entire genome of some organisms has become a key issue in science. This progress of large genome sequencing projects has changed the scale of biology, since all these sequence data must be translated into the coding protein networks that define the physiology of the organism (James, 1997a). In a relatively short period of time, several complete nucleotide sequences have become available with many more to come in the near future (table 1.1). An up-to-date list of genomes that have been sequenced or are in the progress of being sequenced can be find on the World-Wide Web (WWW) site, http://www.tigr.org/tdb. It can be seen from table 1.1 that the genomes show a large variety in size from only 9 genes for $\phi X174$ to 5,885 genes for *Saccharomyces cerevisiae*. The yeast *S. cerevisiae* is the only eukaryote genome sequenced so far. But progress is being made for example on the human genome sequencing project (about 3 billion base pairs comprising about 70,000 to 100,000 genes). This ambitious project should be completed within the next five to seven years.

Nevertheless, genome sequencing projects are not an end in themselves. A great challenge that biologists face now is how to examine the expression of the thousands of gene products. The proteome, which is the product of all the genes is not a fixed feature in an organism compared to its genome. Instead, it changes with the development of the cell or tissue or even with the environmental conditions of the organism. Therefore, there are many more proteins in a proteome than genes in a genome. Especially in eukaryotes there can be various ways a gene is spliced in constructing

mRNA, and there are many ways that the same protein can be post-translationally modified. So the one-gene-one-protein hypothesis of Beadle and Tatum is not longer valid (Beadle and Tatum, 1941).

Table 1.1 Some organisms whose genomes are fully sequenced (modified from Williams and Hochstrasser, 1997).

Organism[a]	Genome size (Mb)	Number of Genes[b]	Reference
φX174	0.005	9	Sanger et al., 1977
Mycoplasma genitalium	0.58	470	Fraser et al., 1995
Methanococcus jannaschii	1.7	1,738	Bult et al., 1996
Haemophilus influenzae	1.8	1,743	Fleischmann et al., 1995
Bacillus subtilis	4.2	3,450	Kunst et al., 1997
Mycobacterium tuberculosis	4.4	3,924	Cole et al., 1998
Escherichia coli	4.6	4,285	Blattner et al., 1997
Saccharomyces cerevisiae	12.1	5,885	Goffeau et al., 1996

[a] Other organisms whose genome sequences will be available within two years include the fruitfly *Drosophila melanogaster,* the plant *Arabidopsis thaliana* and the nematode *Caenorhabditis elegans* and many more (see http://www.tigr.org/tdb).
[b] Predicted from the DNA sequence. Note that proteins smaller than approximately 10 kDa are often not predicted from genomic sequences.

The recent discovery of protein splicing has contributed to another complexity of the gene to protein concept. This post-translational reaction involves the precise excision from a precursor protein, coupled to peptide bond formation between the flanking N-terminal and C-terminal domains, to give a spliced protein product (Kane et al., 1990). Parallel to the nomenclature in RNA splicing the INternal removed proTEIN fragment is called the intein and the EXternal proTEIN fragments to form the mature protein is called the extein. An intein has been defined as a protein sequence

embedded in-frame within a protein precursor that is removed by protein splicing (Perler et al., 1994). Inteins, once considered an oddity, are now known to be widely distributed. Figure 1.1 shows the first observed example of protein splicing in the *S. cerevisiae* TFP1 gene, which encodes the 69 kDa catalytic subunit of the vacuolar H^+-ATPase. A continuous open reading frame predicted a protein of 119 kDa with the N- and C-exteins exhibiting 75% sequence identity to the equivalent V-ATPase subunit of *Neurospora crassa*. This homology was interrupted by a 454 codon stretch of sequence in the TFP1 gene, that lacked homology to any known ATPase subunit. The precise deletion of this codon stretch in the gene produced the V-ATPase subunit of the correct size. Northern blot analysis excluded the involvement of RNA splicing and mutagenesis showed that translation of the intein-coding sequence was required for production of the spliced protein. (Kane et al., 1990). The excised inteins are site-specific DNA endonucleases that catalyse genetic mobility of their DNA coding sequence by an intein homing mechanism. The mechanism by which protein splicing occurs is probably entirely encoded within the internal protein sequence or intein by means of auto catalysis and does not require other accessory molecules. Some mechanistic aspects were recently reviewed (Colston and Davis, 1994; Perler, 1998).

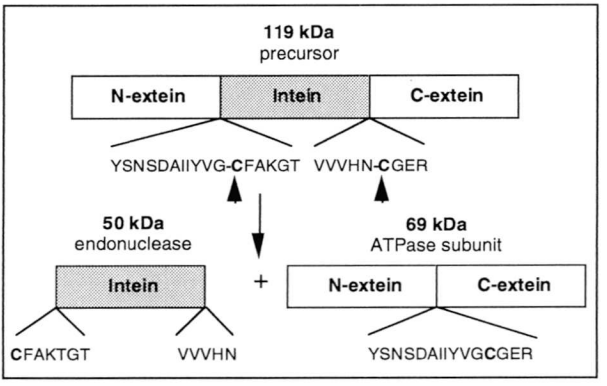


***Figure 1.1* Example of protein splicing.** The 119 kDa TFP1 precursor is spliced into the 69 kDa V-ATPase subunit (N-extein and C-extein) and 50 kDa endonuclease (intein). The arrowheads show the predicted cleavage points at the precursor. There is a high degree of conserved motifs at the intein-extein junctions found in several organisms (Cooper and Stevens, 1995).

There are two main approaches that can be used to examine gene expression on a large scale. One uses nucleic acid-based technology, the other protein-based technology. The most promising nucleic acid-based technology is differential display of mRNA, which uses polymerase chain reaction (PCR) with arbitrary primers to generate thousands of cDNA species, each corresponding to an expressed gene or part of a gene. Recently, two new approaches have been introduced. Advances in biochemistry and engineering have enabled the development of biological assays on micro chips.

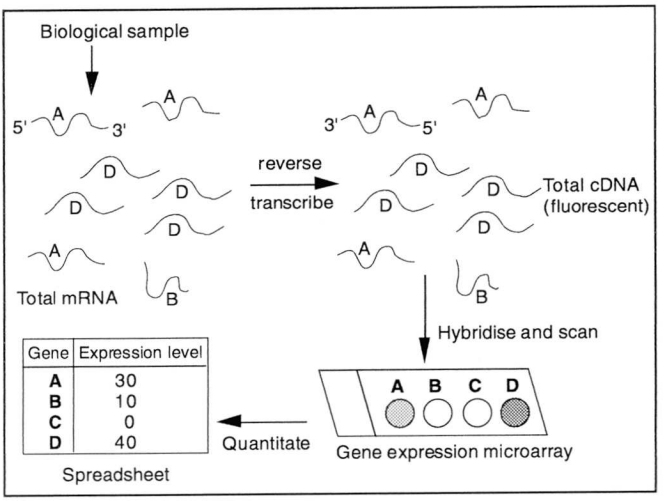


***Figure 1.2* Microarray assay for gene expression.** The steps used to monitor gene expression on a cDNA chip are outlined. Total mRNA is isolated from any biological sample, primed with oligo-dT and fluorescently labelled using a single round of reverse transcription in the presence of fluorescein-12-dCTP. The fluorescent probe is hybridized to a cDNA microarray containing specific hybridization targets, washed and scanned for fluorescence emission following laser excitation. Measurement of the intensity allows quantitation of gene expression. The data are displayed on a spreadsheet for further analysis.

One method describes the attachment of cDNA on such microarrays (DNA chips). These are printed on a glass surface using computer controlled, high-speed robotics. The cDNA's to be arrayed are first amplified in a 96-well format using PCR. Samples in the order of nL of the amplified cDNA's are then transferred from microtiter plates onto glass microscope slides using a robotic printhead (Schena et al., 1995). The source of cDNA's

for the microarrays may include fully sequenced clones, collections of partially sequenced cDNA's knows as expressed sequence tags (EST's, see also section 2.1.1 for a more detailed description of EST's) from any library of interest. Currently, it is feasible to produce microarrays containing 20,000 targets per day. The isolated mRNA from a cell or tissue is subjected to a single round of reverse transcription and fluorescence tagging. The resulting fluorescent cDNA is then hybridized with the microarray and scanned with a laser. Fluorescence indicates if hybridization has occurred with the intensity being an indication for the amount of cDNA binding. Figure 1.2 shows an overview of the microarray assay (Schena, 1996).

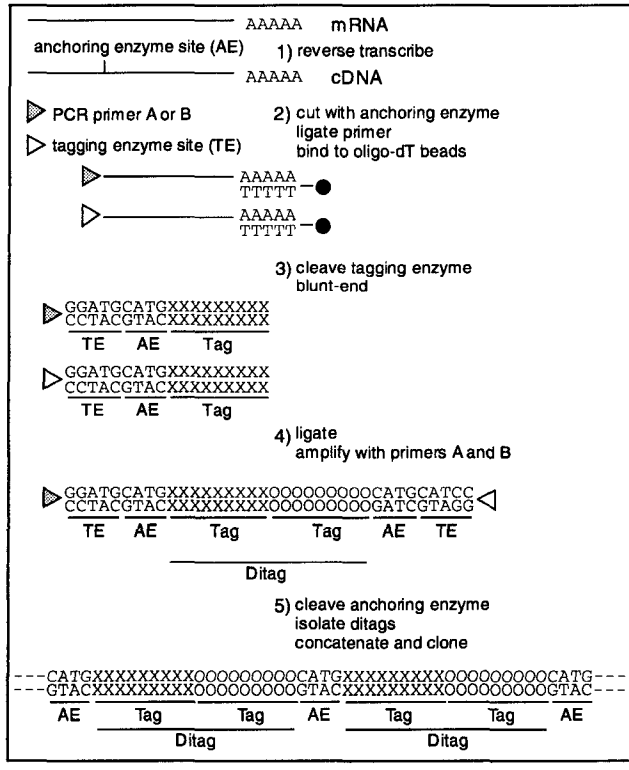


***Figure 1.3* Diagram of the SAGE protocol.** A detailed description of the procedure is given in the text. In this example NlaIII is the anchoring enzyme (AE) and FokI is the tagging enzyme (TE) (Velculescu et al., 1995; Adams, 1996).

Another method is called Serial Amplification of Gene Expression (SAGE). Figure 1.3 shows an overview of the SAGE procedure. The process starts with double-stranded cDNA constructed from a mRNA sample of choice. After cleavage with a frequent cutting restriction enzyme, termed the anchoring enzyme, cDNA is bound to streptavidin beads via a biotinylated oligo-dT primer. This ensures that tags will be generated from a defined position in the transcript, the most 3' occurence of the anchoring enzyme restriction site. The bead-bound cDNA is then split into two fractions, each of which is ligated to one of two PCR primers containing the five-base recognition site for a defined tagging enzyme. Each fraction is then digested with the tagging enzyme and blunt-ended. Velculescu et al. (1995) used the type IIS restriction enzyme BsmFI, which cleaves 12 bp 3' of its recognition site. This 12 bp sequence includes 3 bp of the anchoring enzyme recognition site and 9 bp that are specific to the cDNA. The 9 bp sequence is referred to as the SAGE tag. The two pools of PCR primers attached to SAGE tags are then ligated to each other to make ditags and amplified. By ligating prior to amplification, any bias in PCR can be eliminated by discounting any repeated ditags that may have arisen during amplification. Using this method, clones containing 10-50 tags are amplified by PCR and the sequence determined (Velculescu et al., 1995).

Perhaps the most controversial aspect of SAGE is the contention that a 9 bp sequence can be a unique identifier for a transcript. With 9 bp there are 262,144 (4^9) sequence combinations possible, more than the number of human genes. It was found that over 95% of the transcripts could be distinguished based upon 9 bp sequences adjacent to the most 3' anchoring enzyme sites. This means that in the majority of cases the 9 bp SAGE tag will be an effective link to more complete sequence and gene identification information. In order to determine if the technique accurately measures mRNA levels, SAGE results with pancreatic mRNA were compared to results from screening a cDNA library with a radiolabelled probe. Four probes representing abundant transcripts showed very similar representations by plaque hybridization to the library and by the number of SAGE tags (Velculescu et al., 1995).

Using SAGE, Velculescu et al. (1997) have analysed the set of genes expressed from the yeast genome. The impressive analysis of 60,633 transcripts revealed 4,665 genes, with expression levels ranging from 0.3 to over 200 transcripts per cell. Of these genes 1,981 had known functions, while 2,684 were previously not characterized. These studies provide insight

into global patterns of gene expression in yeast and demonstrate the feasibility of whole genome expression studies in eukaryotes.

While these two recently described approaches are extremely powerful and now becoming automated to allow massive screening, it is important to realize that mRNA based approaches do not measure the actual protein present. A protein cannot be synthesized without its mRNA present, but you can have protein present when its mRNA is no longer present and conversely, you can have lots of mRNA and no translation of the message into the protein. A recent study showed that there is no good relation between mRNA abundance and protein amount at a given time. Anderson and Seilhammer (1997) showed in human liver that mRNA's were enriched for secreted proteins, while mRNA's for cellular proteins were under represented. It was shown that the correlation coefficient between mRNA and protein abundancy was 0.48. Tew et al. (1996) studied this correlation for one gene product across 60 human cell lines and found a correlation coefficient of 0.43, with 10-fold variations in either protein or mRNA abundance. Further, mRNA's are much more labile than DNA, both in terms of chemical degradation (due to the possession of two adjacent hydroxyl groups on the ribose sugar ring) and in terms of enzymatic degradation (the ubiquitous RNase). Proteins are in general more stable and exhibit slower turnover in most tissues. For example, mRNA levels can decrease up to 200-fold in human brain during a 48 h post-mortem period, while the same samples show little or no decrease in native protein amount (Yolken and Johnston, 1997). The final and perhaps most potent argument in favour of protein-based analysis over nucleic acid-based analysis is function. Proteins implement almost all controlled biological functions and hence, are involved in all important activities, disease processes and drug effects. On the other hand, mRNA is just what its name implies, a disposable message, having no other function than to serve as a temporary piece of information, while being operated by proteins (Anderson and Anderson, 1998).

Thus, at this point it might be clear that one needs to screen the products of genes from cells, i.e. the proteins. This is the protein-based approach. Up to now protein science has been a slow and sometimes frustrating art. Unlike the swift developments in DNA sequencing and mRNA screening, where thousands of genes can be rapidly analysed, this has not been possible for proteins so far.

Only until recently methods have become available to allow massive screening of proteins. These methods include amino acid analysis (Shaw, 1993; Wilkins et al., 1996a) and peptide mass fingerprinting (James et al., 1993; Yates et al., 1993; Mann et al., 1993; Henzel et al., 1993; Pappin et al., 1993). In the next sections these techniques will be discussed in more detail. Protein technology is much more complex than nucleic acid-based technology. The basic alphabet is bigger (4 nucleotides for DNA compared to 20 unmodified and many more modified amino acids for proteins). Additionally, there are also many ways in which proteins are modified after synthesis. Almost all eukaryotic proteins are post-translationally modified in some way, for example by addition of sugars, phosphate, sulphate, methyl, acetyl, lipid groups etc. The technologies required in order to be able to separate and to identify these large numbers of proteins are not straightforward. It is the combination of mainly two technologies that make proteome analysis possible. These include two-dimensional polyacrylamide gel electrophoresis for the separation of complex protein mixtures and mass spectrometry for the identification of proteins.

1.2 Two-dimensional polyacrylamide gel electrophoresis

1.2.1 Introduction

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is the heart of proteome technology, because it is the only method currently available which is capable of simultaneously separating thousands of proteins. The technique involves the separation of proteins by isoelectric focusing (IEF) in the first dimension. Proteins are separated in a pH gradient until they reach a stationary position where their net charge is zero. The pH at which a protein has zero net charge is called its isoelectric point (pI). In the second dimension the proteins are separated according to their molecular weight by SDS-PAGE. The detergent sodium dodecylsulphate (SDS) binds to the proteins at a ratio of about one SDS molecule per two amino acid residues in such a way, that they all have the same net negative charge density and migrate in a polyacrylamide gel according only to their relative molecular mass. In the early 1970's the development of 2-D PAGE started when Kenrick and Margolis (1970) combined native isoelectric focusing with pore gradient SDS gel electrophoresis (SDS-PAGE) to obtain a separation of serum proteins. The 2-D technique which is mostly used today originated from the work of O'Farrell (1975), Klose (1975) and Scheele (1975), who used denaturing IEF in the first dimension. The resolving power of the technique was already demonstrated by O'Farell showing a separation of over 1,000 proteins from *E. coli* cell extracts (O'Farell, 1975).

It has become the method of choice for the separation of complex mixtures of proteins from tissues and even whole cells, because of its enormous high resolution, due to the independent parameters for the two dimensions.

1.2.2 Technical limitations and solutions for 2-D PAGE

The main drawback for 2-D PAGE for a long time was reproducibility, not only between different laboratories, but even within the same laboratory. This was mainly due to the instability of the pH gradient formed by carrier ampholytes during the first dimension. Furthermore, preparative amounts of proteins could not be separated very well on these type of gels. The introduction of immobilized pH gradient (IPG) gels largely overcame

the problem of gradient instability and poor sample loading capacity, making low abundant proteins difficult to visualize (Bjellqvist et al., 1982; Görg et al., 1988; Righetti, 1990). The IPG's are formed by co-polymerization of the pH gradient and the polyacrylamide support. This results in extremely stable pH gradients over a large range, except for the very alkaline (> 12) pH values. Only very recently IPG's have been described for the separation of proteins in a pH range of 4 to 12 in a highly reproducible manner (Görg et al., 1998). Since the IPG's are commercially available in a broad range of pH, both linear, non-linear, stepwise or sigmoidal, this has increased to a large extent the reproducibility in 2-D PAGE nowadays. Using IPG's in the first dimension is the recommended method of choice in proteome analysis.

One of the most important steps before starting 2-D PAGE is the sample preparation. The treatment of the samples involves solubilization, denaturation and reduction to completely breakdown the interactions between proteins and to remove non-protein components from the sample. In the ideal case the pretreatment is a single step. This is the case for soluble samples. Many proteins are insoluble, like membrane bound proteins, or are highly resistant to denaturation due to the tissue. Adding (thio)urea and surfactants like 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulphonate (CHAPS) improve the solubilization of the proteins and increase the transfer of proteins to the second dimension gel (Rabilloud et al., 1997). Another important feature is to completely reduce inter- and intra disulphide bonds. This is achieved by adding reducing agents such as β-mercaptoethanol or dithiothreitol (DTT). Reagents like DTT are charged and migrate out of the pH gradient during the first dimension, reducing the solubility of certain proteins. Replacing the thiol containing agents by a non-charged reducing agent, such as tributhylphosphine (TBP) can enhance protein solubility (Herbert et al., 1997). Finally, the presence of nucleic acids, especially DNA has negative effects on the separation of proteins. DNA increases the viscosity of the solution, inhibiting protein entry into the IPG gel. Additionally, DNA can bind to proteins and cause streaking (Rabilloud et al., 1996). Usually, nucleic acids are removed by treatment of the sample with endonucleases (Rabilloud et al., 1986).

There are several ways of applying the sample to the surface of an IPG gel. Because IEF is an equilibration technique, the actual point of applying the sample does not matter. When loading the sample at a discrete point on the IPG gel it is usually better to load at one of the pH extremes, where most of the proteins will be charged and thus minimize sample loss through

precipitation. For this reason, loading of the sample was achieved with a sample cup at one end of the IPG strip. However, even with the most soluble samples it is difficult to avoid sample loss at the loading point, especially with high sample loads (Bjellqvist et al., 1993a). To increase the amount of protein entering the gel when loading with cups it is necessary to radically change the dimension of the IPG gel, especially the loading area and the entry time at low voltage. Using this approach it is possible to load up to 15 mg protein, although the IPG gels and sample cups ("tennis racket" shaped cups) are non standard, which does not improve inter-laboratory reproducibility. Another disadvantage is that loading the sample this way causes smearing of the protein spots on the gel (Bjellqvist et al., 1993a).

The fact that the commercially available IPG strips are supplied in a dehydrated form and that sample can be applied at any position on an IPG allows the sample to be loaded during the gel rehydration over the whole IPG. This avoids sample loss by precipitation in sample cups (Rabilloud et al., 1994; Sanchez et al., 1997). This method of sample loading offers many advantages over cup loading and is rapidly becoming the loading method of choice. The elimination of sample loss provides a means of quantitative sample loading for analytical purposes and very high (> 10 mg) micro-preparative sample loading on a standard IPG gel.

1.2.3 Applications of 2-D PAGE

Many processes like cell signalling cannot be monitored by looking at the mRNA expression levels, since the mRNA levels do not represent the amount of the proteins present in the cell as described before. Another important point is that many proteins are post-translationally modified. By studying at the mRNA level one can not distinguish the modified proteins from the native ones. However, by using 2-D PAGE one can differentiate between the modified proteins. Post-translationally modified proteins appear usually as a series of spots on 2-D gels.

In general there is an increasing interest in the application of 2-D PAGE, not only in biochemistry but also in clinical and biomedical research. Many biochemical pathways, such as cell signalling, involve phosphorylation or dephosphorylation of proteins by kinases and phosphatases. Many drugs or certain toxins can modulate or significantly change these phosphorylation and dephosphorylation events. Since such events involve post-translational modification of the protein, mRNA

analysis is not very useful, however, 2-D PAGE allows these modifications to be monitored and analysed. For example, several cytokines interact with receptors of the hematopoietin receptor superfamily, the signal transduction pathways including several phosphorylated proteins (Linnekin et al., 1992). Many hormones trigger a cascade of phosphorylation or dephosphorylation in cells. Human chorionic gonadotropin induces protein phosphorylation in chorionic tissue (Yamawaki and Toyoda, 1994). These are only a few examples where the study of protein phosphorylation on human samples was done using 2-D PAGE technology.

Multiple genetic modifications are required for cancer development. But not only oncogene and oncogene product alterations, but cell cycle specific protein modifications play a major role in tumorogenesis and cancer progression (Hartwell and Kastan, 1994). The reliable measurement of protein expression in patient cells or tissue biopsies relies almost entirely on quality of the sample preparation. It was found that nonenzymatic methods for the preparation of tumour cells have advantages over methods using enzymatic extraction of cells. Nonenzymatic methods are rapid and show reduced loss of high molecular weight proteins (Franzen et al., 1991, 1993, 1995). Further, it was found that features, such as local homogeneity, amounts of connective tissue and serum proteins were critical factors for the successful preparation of the sample and high quality of the separation and analysis of the proteins. Clear guidelines are now available for sample preparation of patient cells and biopsies (Franzen et al., 1995). In many cases it is possible to make a direct comparison of proteins expressed in normal and diseased tissue from a single patient. The comparison of 2-D separations of such tissues can immediately highlight proteins that are present in greater or lesser quantities, new proteins only expressed in the cancerous cells and changes in protein post-translational modifications. A summary of studies on cancerous tissues using a proteomic approach was recently reviewed (Hochstrasser, 1997).

In many other fields of medicine the 2-D PAGE technology is applied as well. In the molecular epidemiology of viruses and bacteria (Cash, 1991), to determine changes in protein synthesis patterns in murine organs during post-implantation development (Praxmayer et al., 1992) and in studying the immune response (Kovarova et al., 1992).

Comparative protein maps, derived from 2-D PAGE from for example normal cells and pathological cells have been created to be used as diagnostic tools (Appel et al., 1991).

1.2.4 *Methods of protein detection*

There are several ways for visualizing proteins in 2-D gels. The method of choice is dependent on the amount of protein loaded, the purpose of the gel (protein quantification or blotting) or the sensitivity required.

The most common means of protein detection and their applications are summarized in table 1.2. In general the dye Coomassie Blue is used to detect the proteins. It results in a fairly clear background suitable for gel scanning, that is required for further analysis. Furthermore, proteins can be subjected to protease digestion or N-terminal sequencing. The detection limit lies in the picomole range. When higher sensitivities are required silver staining is usually done. The detection limit is a few orders of magnitudes higher than for Coomassie Blue, but further analysis is problematic due to irreversible modifications of the silver stained protein. Although most methods of detection give some indication of the quantities of protein present, in general they cannot be used for quantitation. This is because protein stain binding is not linear over a wide range of protein concentration, binding is dependent on amino acid composition and on post-translational modifications of the protein (Goldberg et al., 1988; Li et al., 1989).

Table 1.2 Common stains for 2-D gels or blots and their applications (sensitivity is per protein spot or band in the gel).

Detection method	Main applications	Sensitivity	Reference
[^{35}S] Met or ^{14}C radiolabelling	cell lines, cultured organisms	20 ppm of radiolabel	Latham et al., 1993
[^{35}S] thiourea silver	ultra high sensitivity gel staining	0.4 ng protein	Wallace and Saluz, 1992
Silver	very high sensitivity gel staining	4 ng protein	Doucet and Trifaro, 1988

Table 1.2 (continued) Common stains for 2-D gels or blots and their applications (sensitivity is per protein spot or band in the gel).

Detection method	Main applications	Sensitivity	Reference
Coomassie Blue	staining of gels and PVDF membranes	40 ng protein	Schägger and von Jagow (1987)
Zinc imidazole	reverse staining of gels or PVDF membranes (useful for MALDI-MS of peptides)	higher than Coomassie Blue	Ortiz et al., 1993
Ponceau S	staining higher protein loads on PVDF	100 ng protein	Sanchez et al., 1992
India ink	staining of NC or PVDF membranes	1-10 ng protein	Hughes et al., 1988
Stains-all	staining glyco-proteins or Ca^{2+} binding proteins	100 ng protein	Campbell et al., 1983

1.2.5 *Two-dimensional gel analysis and proteome databases*

After completion of the electrophoresis and visualizing of the proteins, images of the gels are made using a laser scanner or a charge-coupled device (CCD) camera and digitized for computer analysis. Gel images can be manipulated by improving the spot intensities against background. Several gel analysis and database programs were (and still are) developed such as TYCHO (Anderson et al., 1981), GELLAB (Lemkin and Lipkin, 1981), MELANIE (Appel et al., 1991) and QUEST (Garrels, 1989) Alternatively, the gel images can be analysed manually. The latter might sound obscure, but when comparing for example 2-D gels of cells grown under different conditions the manual method turned out to be very reliable (Quadroni et al., 1996).

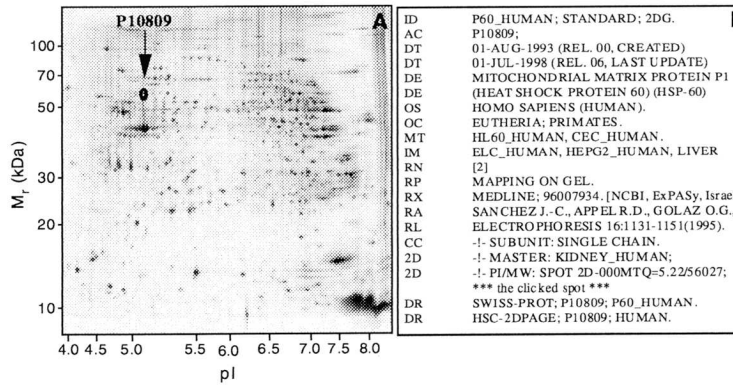


***Figure 1.4* The 2-D gel master image of human kidney and an excerpt from its annotation from the SWISS-2DPAGE database.** (A) The gel image is available from the URL: http:/www.expasy.ch/cgi-bin/map2/def?KIDNEY_HUMAN. Protein spots that have been identified can be selected by clicking the mouse. (B) After selecting an identified protein spot the annotation linked to the spot is displayed. This annotation shows how the protein was identified and provides further links to the SwissProt database.

Anderson and Anderson (1982) postulated the idea of the "human protein index". This was to be a database of all the spots resolved by a 2-D gel (from a certain tissue) together with structural, functional and clinical information. Especially since the introduction of the IPG's in the first dimension, which made inter-laboratory gel comparisons possible there is a rapid rise in the development of such 2-D gel databases. One of the first and to date most complete projects to attempt to map systematically all of the spots on a 2-D gel to protein sequences started about twenty years ago for *E. coli* (Pederson et al., 1978). Other databases followed and recently a standard format has been proposed to create linked databases that are accessible through the World-Wide Web (Appel et al., 1996; Hoogland et al., 1998). In figure 1.4 an example of such a 2-D PAGE database entry is shown. Clicking on a protein spot that has been identified displays the annotation of the corresponding protein.

1.3 Protein identification in proteome analysis

1.3.1 Overview of protein identification strategies

In proteome projects the identification and to a lesser extent the characterization of all proteins expressed by an organism or tissue is the central point of attention (Wilkins et al., 1995; James, 1997b). The challenge is now how to identify hundreds or even thousands proteins from a 2-D gel in an economical way. Traditionally, proteins from 2-D gels have been identified by techniques such as immunoblotting, N-terminal sequencing, internal peptide sequencing, co-migration of unknown proteins with known proteins or overexpression of homologous genes of interest in the organism under study (Matsudaira, 1987; Rosenfeld et al., 1992; Van Bogelen et al., 1992; Celis et al., 1993; Honore et al., 1993). Although these techniques can be very powerful, they are too expensive, too slow and labour intensive for screening purposes of large amounts of proteins

Alternatively, the identification of proteins on a large scale can be done by looking at different parameters (species of origin, pI, apparent molecular mass, N- and C-terminal sequence) of the proteins and use these, either alone or in combination, to match protein databases in different ways. Most of these parameters relate directly from the protein sequence. A fundamental property of any protein is its species of origin. Nowadays, with some completely genomes known, and many more to come in the future, one can match proteins from a single species only to the proteins in the database from that species, this way eliminating the tens of thousands of other proteins in the database. From 2-D gels the pI of the protein can be estimated. The pI is highly dependent on amino acid sequence, its N- and C-terminal and any post-translational modifications. The unknown proteins are matched against known proteins on the 2-D gel. By using the known proteins a grid is formed on the gel. This allows the unknown proteins to be assigned. Clearly, pI alone is not sufficient for protein identification, but it can exclude many other proteins in the database (Bjellqvist et al., 1993b, 1994; Wilkins et al., 1996a).

Closely related to an accurate pI is the determination of the amino acid composition of a protein. This can be carried out on PVDF membranes or glass-fibre. Proteins separated by 2-D PAGE were selected from the gel and extracted with 0.1 M NaOH and 2% thiodiglycol. The extracted proteins were

then acid hydrolysed and applied to an automated amino acid analyser (Manabe et al., 1982). The method regained popularity when more sensitive amino acid analysers were developed. The method is useful for the more abundant proteins, especially in combination with other identification techniques, such as N-terminal sequencing. Eckershorn et al. (1988) have applied the method to the identification of mouse brain proteins by 2-D PAGE. After the amino acid composition of the protein has been determined, computer programs are used to match it against the calculated compositions of proteins in databases (Eckershorn et al., 1988).

A potential very sensitive method which would allow thousands of proteins to be monitored simultaneously was put forward by Latter et al. (1983). By using 20 different cultures of a cell line, each grown with a different ^{14}C- or ^{35}S-labelled amino acid, the amino acid composition of each of the proteins could be determined by computer analysis of the autoradiograms of the 2-D gels. A protein can be identified in a database search using the ratio of four pairs of amino acids, determined using ^{14}C (for any amino acid) and ^{35}S (methionine or cysteine). Since the decay times are drastically different (half-lifes of 5,730 years for ^{14}C and 88 days for ^{35}S) the relative amounts of each amino acid can be determined for each spot on the gel. The main drawback of this method is the even introduction of the isotope labelled amino acids due to the metabolic scramble (James, 1997b).

In a similar way as for the determination of pI, the apparent mass of the protein can be determined from the 2-D gel. However, one has to take care, since proteins, especially if they are small or are post-translationally modified, show different migrations than should be expected from their theoretical mass. This can differ from 30% for small proteins to 50% for glycosylated proteins (Wilkins et al., 1996a). More accurate mass determinations can be obtained by using mass spectrometry. This will be described later in more detail (see section 1.4).

1.3.2 *N-terminal sequencing*

For a long period of time, the method of choice for determining the sequence of a protein was N-terminal sequencing, developed by Per Edman (Edman, 1949). The Edman degradation sequentially removes one residue at the time from the N-terminal of the peptide. Figure 1.5 shows the chemistry of the Edman degradation.

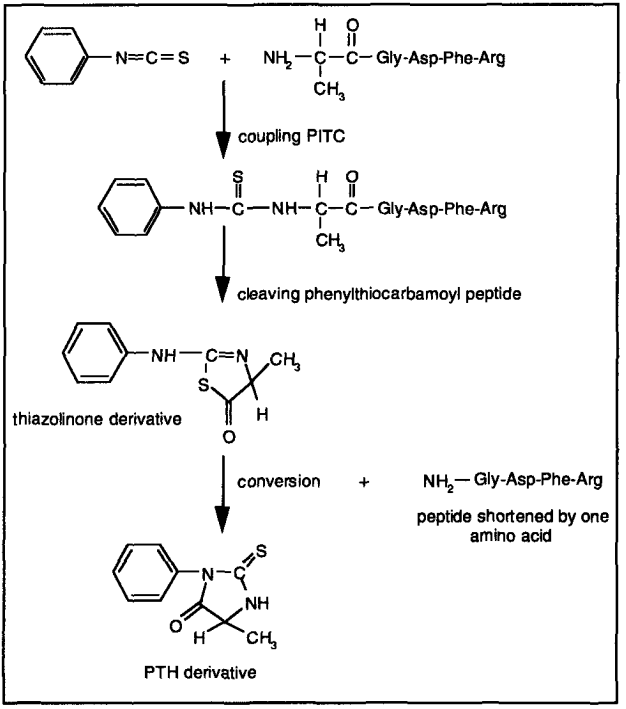


***Figure 1.5* The Edman degradation reaction.** Scheme of the main reactions: coupling of phenylisothiocyanate in alkaline conditions to the N-terminal part of the peptide or protein followed by cleavage of the first amino acid residue by strong acid and conversion of the unstable thiazolinone to the stable phenylthiohydantoin for subsequent HPLC analysis. The cycle of reactions on the intact peptide, shortened by one amino acid residue is then repeated.

The principle is based on coupling of the N-terminal of the protein with phenylisothiocyanate (PITC) followed by cleavage of the labelled residue without disrupting the remaining peptide chain. PITC reacts with the uncharged N-terminal group of the peptide to form a phenylthiocarbamoyl derivative. Under strong acidic conditions a thiazolinone derivative of the terminal amino acid is released, which leaves an intact peptide shortened by one residue. The unstable cyclic amino acid derivative is converted into the stable phenylthiohydantoin (PTH) that can be identified by chromatographic methods. Although the chemistry has remained nearly unchanged over the decades, many improvements have been introduced to increase the sensitivity of the method. One of the first breakthroughs was the automation of the procedure resulting in

commercially available instruments (Edman and Begg, 1967). Other improvements include the different methods of peptide immobilization, the way the chemicals are delivered and especially miniaturization of the instrumental parts, which lower losses of material and allow much higher coupling and cleaving efficiencies. A modern instrument can produce sequences at the 200 fmole level. This is a tremendous improvement when one considers that in the original Edman procedure about 2 mmole of protein was required.

A major development was the direct sequencing by Edman degradation from proteins electroblotted onto glass fibres or polymer membranes from 1-D and 2-D gels respectively (Aebersold et al., 1986; Bauw et al., 1987). Subsequently, Aebersold et al. (1987) showed that internal sequence information could be obtained from proteins blotted onto nitrocellulose (NC) membranes. The method overcomes the problem of obtaining amino acid sequence data from N-terminally blocked proteins and provides multiple, independent stretches of sequence that can be used to generate oligonucleotide probes for molecular cloning or used to search sequence data bases for related proteins. The main development was the use of a polymer, polyvinylpyrrolidone (PVP-40) to block the membranes, preventing the adsorption of the protease. The use of nitrocellulose as a support which is hydrophobic enough to retain proteins during electroblotting but hydrophilic enough to allow peptide release for subsequent HPLC before sequencing was another major development. Much effort has been made to improve supports for electroblotting and sequencing (Aebersold, 1990). Although as a technique very useful, for the analysis of a large number of proteins from 2-D gels it is not a suitable method being too slow and insensitive, as many proteins from 2-D gels are present in amounts below the current limits of the Edman sequencers.

In section 2.1.3 an alternative chemistry to N-terminal Edman degradation is described for generating short sequence tags to be used to identify large number of proteins.

1.3.3 C-terminal sequencing

Complementary to N-terminal sequencing is C-terminal sequencing. The chemistry described by Schlack and Kumpf (1926) did not change significantly, as the method is still performed this way (Stark, 1968). In figure 1.6 an outline of the C-terminal chemistry according to Stark is shown. The

method involves the activation of the C-terminal carboxyl group with acetic anhydride by protonation. This reacts with thiocyanate in the presence of the strong acid TFA to yield a thiohydantoin derivative, which is cleaved under basic conditions with trimethylamine (TMA).

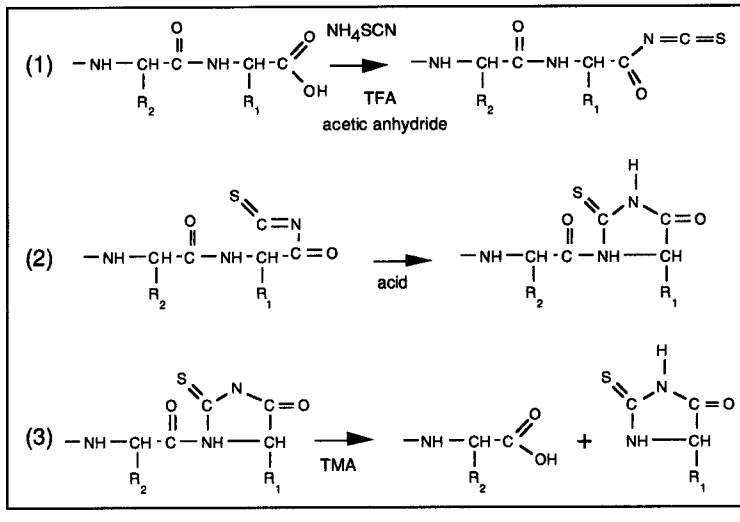


***Figure 1.6* An outline of C-terminal degradation according to Stark (1968).** (1) Ammonium thiocyanate vapour reacts in strong acid (TFA) with the activated carboxyl group, that is protonated by the acetic anhydride to generate a thiocyanate derivative. (2) The C-terminal thiocyanate rearranges spontaneously to give the C-terminal thiohydantoin under the acidic conditions. (3) The thiohydantoin is released by treatment with gaseous trimethylamine (TMA).

The thiocyanate does not react with proline or aspartic acid. New coupling reagents, trimethylsilylisothiocyanate and diphenylphosphoro-isothiocyanatidate were introduced to overcome this problem (Bailey and Shively, 1990; Bailey et al., 1992). The coupling reagent reacts with the carboxyl group to form an acylthiohydantoin which cyclizes to form the thiohydantoins which loses the amide proton. Since proline lacks this amide proton the C-terminal proline is regenerated upon cleavage. It was found that the proline thiohydantoin can be synthesized, the key being the protonation of the thiohydantoin by acid which is then cleavable by water vapour (Inglis et al., 1992). This was adapted by Bailey et al. (1995) to produce the first automated sequencer, that can analyse all 20 amino acids. Another way of C-terminal degradation was described by Boyd et al. (1992). Again, a thiocyanate based coupling reaction was used but an S-alkylation of the

hydantoin before cleavage was introduced. This improves the leaving group potential of the thiohydantoin and allows the introduction of fluorescent markers to increase the sensitivity.

Unfortunately, in both methods the repetitive yields are low and only sequences up to four residues can be obtained, making the method less useful for sequencing proteins from 2-D gels. In general, chemical C-terminal sequencing did not find widespread use so far. Currently, it only finds use in quality control, showing that a protein has the correct C-terminal or as a complement to N-terminal tagging since low repetitive yield gives sequence ladders (see section 2).

1.4 Mass spectrometry

1.4.1 Introduction

Since the introduction of two novel ionization methods, protein analysis is possible by mass spectrometry (MS) and came into widespread use in proteome analysis. These methods of ionization are matrix assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp, 1988) and electrospray ionization (ESI) (Fenn et al., 1989). Table 1.3 shows a comparison of the different MS methods and SDS-PAGE for mass determination (adapted from Rademaker, 1996).

Table 1.3 Comparison of techniques for protein molecular mass determination (numbers shown are approximate and represent typical values).

	MW limit	*accuracy*	*sample amount*
ESI-MS	100,000	0.1%	0.0001 - 10 pmole
MALDI-TOF	500,000	0.05%	0.05 - 1 pmole
SDS-PAGE	5,000,000	5%	0.0001 - 500 pmole

1.4.2 Matrix assisted laser desorption/ionization (MALDI)

The MALDI-MS technique is mostly used with a time-of-flight (TOF) analyser. The principles of the MALDI-TOF are illustrated in figure 1.7. In this method the protein or peptide sample is embedded in a matrix such as α-cyano-4-hydroxy-cinnamic acid or 3,5,-dimethoxy-4-hydroxycinnamic acid. A laser pulse, usually from a nitrogen laser at 334 nm is used to desorb the sample from the matrix to cause ionization.

The ions are accelerated by a high potential (20 - 30 kV) in a field free drift region towards the detector. The time required to reach the detector is proportional to the voltage and the mass of the ion. It is calculated by the following equation:

$$t = s\left(\frac{m}{(2KE)n}\right)^{1/2}$$

where t is the drift time, s the drift distance, m the mass, KE the kinetic energy of the ion and n the number of charges on the ion.

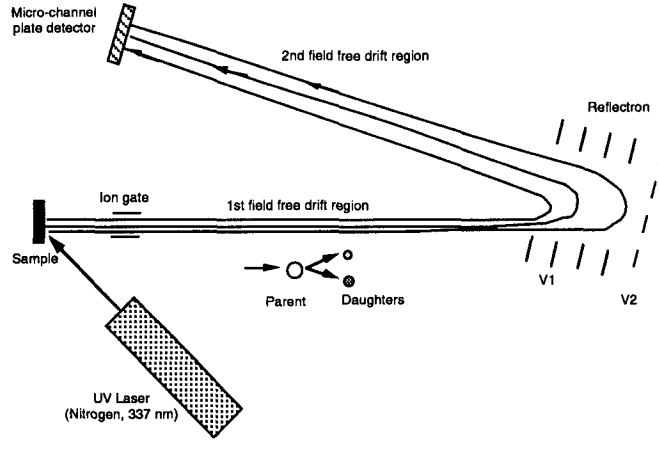


***Figure 1.7* A schematic diagram of a reflectron MALDI-TOF mass spectrometer.** A laser pulse is used to desorb and ionize the sample, which is embedded in an UV absorbing matrix. The ions are accelerated by a large difference in potential and pass through the evacuated flight tube to the reflectron. This ion mirror reflects the ions in a mass dependent way to the detector. The mass is determined by measuring the time needed for the ion to reach the detector from the target. Parent ions can be selected using an ion gate and fragments which have decomposed during the flight to the ion reflector, are measured by lowering the voltage in the reflector from its normal value, thus sending fragments of various sizes to the detector. This process is known as post-source decay (PSD) and is used to obtain sequence information (from James, 1997b).

The recent introduction of pulsed ion extraction, known as delayed extraction (DE) greatly increases the resolution. This is achieved by separating the desorption and acceleration events by applying the accelerating field with a slight delay (in the order of 200 - 500 ns) relative to the laser pulse. The ions have then the same initial velocity. Combined with a reflecting TOF analyser this results in very accurate mass determinations in routine peptide analysis (Vestal et al., 1995; Jensen et al., 1996).

1.4.3 Electrospray ionization (ESI)

ESI-MS makes use of a triple quadrupole (TSQ) or ion trap (IT) analyser, although recently an ESI coupled to a time-of-flight detector has been described (Verentchikov et al., 1994). ESI-MS is based on an elegant idea

that had already been proposed in the 1960's by Dole et al. (1968) by introducing a very fine spray of solvent, containing the molecule of interest, into the mass spectrometer. Peptides or proteins are ionized at atmospheric pressure by generating a very fine spray of the analyte under a potential difference between the sample exit and the MS entrance. The ions are desolvated either by passage through a heated capillary or with a counter current of gas before entering the high vacuum area of the MS. These modes are called electrospray ionization (ESI, used on instruments from Finnigan) and atmospheric pressure ionization (API, used on instruments from Sciex), respectively. The method makes coupling of HPLC to the MS possible. Figure 1.8 shows a scheme of a ESI-TSQ-MS.

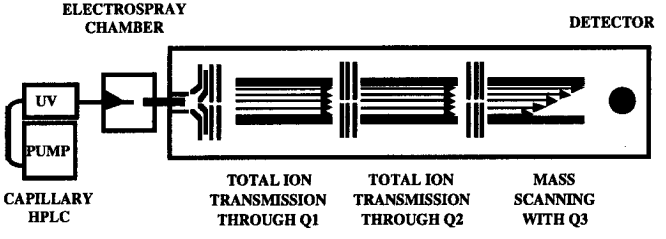


***Figure 1.8* A scheme of a ESI-TSQ mass spectrometer coupled to an HPLC system operating in the normal mode.** The triple quadrupole can be imagined as three consecutive mass filters. Ions are generated in the electrospray chamber at atmospheric pressure before entering the high vacuum of the MS. In the normal scanning mode the first two filters are set to pass the ions through, whereas the spectrum is accumulated by scanning a fixed width window over the mass range using the third filter.

The most important recent improvement in ESI-MS is the development of the nano-ESI source (Gale and Smith, 1993; Kriger et al., 1995; Wilm and Mann, 1996; Wilm et al., 1996). It makes use of a 1 - 2 μm spraying orifice, that is achieved by pulling the spraying capillary to a fine tip. No solvent pumps and inlet valves are required. Due to its low flow rate of about 20 nL/min. it generates very small droplets (< 200 nm, about 500 x smaller than the drops generated by conventional ESI) that facilitate more efficient desolvation and ionization of analyte molecules, resulting in higher sensitivity. Small droplets have a high surface-to-volume ratio, which makes a large proportion of analyte molecules available for desorption. Using this source a stable spray is obtained for 30 minutes from only 0.5 μL analyte solution.

1.4.4 Peptide fragmentation using mass spectrometry

Mass separation in the TSQ is achieved by establishing an electric field in which ions of a certain m/z value have a stable trajectory through the field. The electric fields are created by simultaneously applying a d.c. voltage and an oscillating R_f voltage on four parallel metal rods, the quadrupoles. Adjacent rods have opposite d.c. polarity. Ions move with complex trajectories containing characteristic frequencies as they drift down the axis of the array of rods. By increasing the magnitude of the d.c. and R_f voltages while maintaining the ratio, stable trajectories are created for ions of different m/z to pass through the quadrupole array and exit to the detector. Using the TSQ (or IT) analyser one can select the ion of interest (the parent ion), even from a mixture of peptides in the first quadrupole by filtering out other peptides. The parent ions are accelerated in the second quadrupole which contains a collision gas such as argon (at 1 - 4 mTorr). In the second quadrupole the parent ions undergo fragmentation (collisionally induced dissociation, CID) (James, 1997b; Yates, 1998). The fragmented ions (the daughter ions) are analysed in the third quadrupole. This is known as MS/MS or tandem mass spectrometry. Figure 1.9 shows a scheme of the TSQ mass spectrometer in the MS/MS mode. For an extensive explanation of the operating principles of the TSQ one is referred to Miller and Denton (1986).

Hunt et al. (1986) pioneered protein sequencing by tandem mass spectrometry. Although the mechanism of gas phase peptide fragmentation is not well understood, a few generalizations can be made. If the collision energies used are fairly low (10 - 40 eV) fragmentation occurs only along the peptide backbone. Within the backbone fragmentation can occur in different modes. Relative high energies (> 30 eV) favours fragmentation of single bonds, yielding b-type ions. The generally accepted nomenclature for fragmented ions is described by Roepstorff and Fohlman (1984) and Biemann (1988, 1990). Appendix 1 shows this nomenclature for fragmentation ions in mass spectrometry. If the energy is lower, y-type ions, formed by proton transfer to the amide nitrogen and elimination of a ketene, will preferrentially occur since this involves simultaneous bond formation and cleavage. However, distribution of the type of ions is also dependent on the amino acid composition. High energy collisions (1000 eV) as used in four sector magnetic instruments are sufficient to cause side chain fragmentation which can allow the isobaric amino acids leucine and isoleucine to be distinguished. All of the ion types described can eliminate

small molecules such as ammonia (-17 a.m.u.), water (-18 a.m.u.) and carbon monoxide (-28 a.m.u.) resulting in a pair of signals. (Papayannopoulos, 1995).

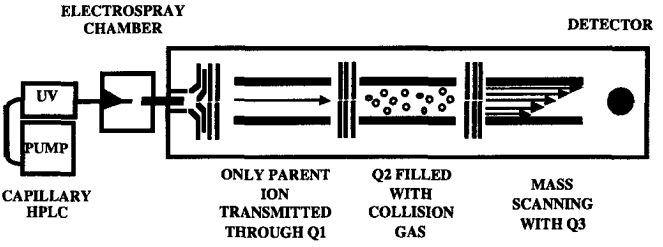


***Figure 1.9* A scheme of a ESI-TSQ mass spectrometer coupled to an HPLC system operating in the daughter scanning or MS/MS mode.** In this mode the first filter is used to select a mass window to allow to pass only the ion of interest (parent ion). The ion is accelerated to the second quadrupole containing a collision gas such as argon, where it undergoes multiple collisions causing fragmentation. The daughter ions are analysed by scanning the third filter.

The great advantage of performing MS/MS is that *de novo* sequence information can be obtained from these fragmentation patterns from mixtures of peptides, whereas by using Edman degradation a single pure peptide is required. From the differences in masses between a series of y-ions or b-ions an amino acid sequence can be deducted. Secondly, MS techniques are much more sensitive than conventional Edman techniques, increasing from the picomole range (Hunt et al., 1986) up to the femtomole range (Hunt et al., 1992). The development of the nano-ESI sources has dropped the sensitivity further into the low femtomole range (Wilm et al., 1996) whereas attomole sensitivity has been recorded using an ion cyclotron instrument (Valaskovic et al., 1996). Another advantage of obtaining sequence information using MS/MS over the Edman degradation is the possibility to sequence post-translationally modified proteins or otherwise N-terminally blocked proteins (estimated to account for about 70% of all eukaryotic proteins), phosphorylation sites and other modifications. Disadvantages of tandem MS are the difficulties in interpreting MS/MS spectra (reviewed by Papayannopoulos, 1995). The isobaric amino acids leucine and isoleucine cannot be distinguished, whereas amino acids such as glutamine and lysine must be differentiated by acetylation.

The ion trap can be considered as a 3-D analogue of the quadrupole mass filter and allows multiple analyses on selected ions (Jonscher and

Yates, 1997). MS/MS can be carried out in a way similar to the TSQ. The parent ion is selected by using d.c. and R_f fields to make all ions except the parent to be ejected from the trap. The trapped ion undergoes collisions with the helium gas (at about 1 mTorr) that is always present in the trap. The mass spectrum of the daughter ions is recorded by sequentially ejecting the product ions. Since the separation of ions is separated by time and not by multiple MS stages, multiple dissociation experiments can be carried out, MS^n (James, 1997b; Yates, 1998).

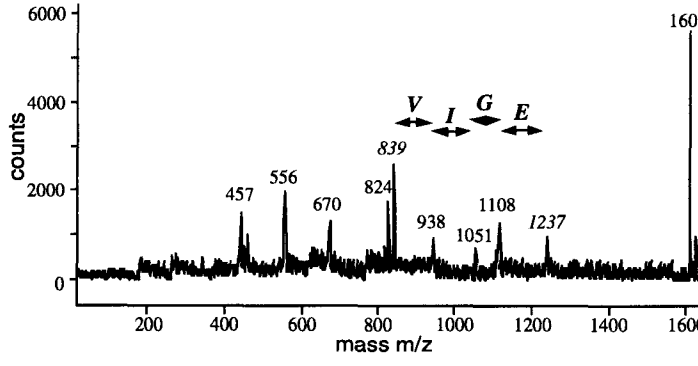


Figure 1.10 **An example of a PSD spectrum.** PSD was done on the peak with m/z = 1607 from a tryptic digest of a protein separated by 2-D PAGE from the bacterium *Bradhyrozobium japonicum*. A short partial sequence (EGIV) could be assigned.

For MALDI-MS an equivalent to MS/MS exists which is called post-source decay (PSD). TOF mass spectrometers can be fitted with an ion mirror, called reflectron to increase the effective path length of the ion from the target to the detector (see figure 1.7). A longer flight path results in higher mass resolution. Using a laser energy increased by a factor 1.5 - 2 over the energy normally required for ionization results in considerable fragmentation of the ions. It was demonstrated in MALDI that a large fraction of the desorbed ions undergo fragmentation reactions during the flight in the time-of-flight tube due to multiple collisions of the peptides with the matrix (Kaufmann et al., 1993). This process is called post-source decay (PSD). Fragments created in the field-free region of the flight tube have the same velocity (but kinetic energies which vary with their masses) as the precursor ion, since the fragmentation occurs outside the accelerating field of the ion source. Since the accelerating voltage for the ions is in the order of 20 - 30 kV, very high energy collisions occur and the peptides have a

relatively long time interval in which to decay. This collision energy is in the order of magnitude as for the magnetic sector instruments. Fragments of higher m/z value will penetrate deeper into the reflectron and exit later than the lighter product ions. A separation based on m/z value is thereby effected in the reflectron. Figure 1.10 shows an example of a PSD spectrum of the ion with m/z = 1607 from a tryptic digest of a protein separated by 2-D PAGE from the bacterium *Bradyrhozobium japonicum*. Although no complete sequence was obtained, a short partial sequence (EGIV) could be assigned. PSD spectra containing enough information to determine the entire sequence of a peptide are very rare, due to low fragmentation, although partial sequences may be obtained which can be useful for database searching (see also section 2). In general, the novel advanced mass spectrometry methods have largely changed protein chemistry (James, 1997b; Yates, 1998).

1.5 Database searching methods in proteome projects

1.5.1 Peptide mass fingerprinting

A completely different and complementary approach for the identification of proteins is called peptide mass fingerprinting. The technique of peptide mass fingerprinting was recently independently described by several groups (James et al., 1993; Yates et al., 1993; Mann et al., 1993; Henzel et al., 1993; Pappin et al., 1993). Proteins separated by 2-D gel electrophoresis are digested either in-gel or on membranes by enzymatic or chemical cleavage methods. The key issue is that the cleavage is specific for certain amino acids. A commonly used enzyme is trypsin, which cleaves only at the C-terminal side of arginine or lysine residues. Chemical digestion methods such as cyanogen bromide (methionine specific) or formic acid (aspartic acid - proline specific) have not found widespread use in peptide mass fingerprinting (Nikodem and Fresco, 1979; Vanfleteren et al., 1992). Complete trypsin digestion results usually in peptides with masses between 500 and 3000 a.m.u. since trypsin cleaves at the abundant amino acids arginine or lysine. Cyanogen bromide is methionine specific, a much less frequent amino acid in proteins, resulting in larger peptides. Further, it is known that protein cleavage with cyanogen bromide leads to many side reaction products. The sequence aspartic acid - proline is even more rare, so cleavage with formic acid is not very useful to generate a set of small peptides.

The exact masses of the resulting peptides are very accurately measured by mass spectrometry. This technique can achieve mass accuracies better than 0.005 mass units for small peptides. The obtained set of peptide masses are characteristic for a protein and serves as a "fingerprint", which is then matched against theoretical peptide libraries generated from protein sequences in databases such as the SwissProt or trEMBL database. As proteins have different amino acid sequences, the peptides from a specific digestion method should produce characteristic fingerprints. Table 1.4 shows the effect of the accuracy of peptide mass determination on the results of such a database search.

Table 1.4 The effect of mass accuracy on peptide mass fingerprint searching. The table shows the results from two searches carried out with the same data, but obtained at different mass accuracies. The higher the peptide mass accuracy, the better the score and the difference between the correct highest scoring hit and the next non-related protein. Currently, mass accuracies in the range of 0.001 are possible, greatly improving selectivity. AC stands for ACcession number of the database entry, DE stands for DEscription (name) of the protein (adapted from James, 1997b).

Score	AC	DE
Mass accuracy ± 0.2 a.m.u.		
Searching with masses: 673.5, 836.5, 915.5, 932.5, 1064.2		
93.1	P02755	Beta-lactoglobulin, water buffalo
78.1	P02754	Beta-lactoglobulin precursor, bovine
78.7	P02757	Beta-lactoglobulin, sheep
76.0	P02756	Beta-lactoglobulin precursor, goat
51.2	P10834	Pet 54 protein, *S. cerevisiae*
Mass accuracy ± 1.5 a.m.u.		
Searching with masses: 674.1, 838.0, 913.5, 931.3, 1064.5		
58.2	P02755	Beta-lactoglobulin, water buffalo
57.8	P18163	Long-chain-fatty-acid-coaligase, rat
49.4	P02754	Beta-lactoglobulin precursor, bovine
49.4	P02757	Beta-lactoglobulin, sheep
49.3	P05413	Fatty acid-binding protein, human

Both ESI-MS and MALDI-MS can be used for peptide fingerprinting. MALDI is preferable because of its higher sensitivity and tolerance to contaminating substances from gels or buffers. If necessary, samples can be directly washed on the target by pipetting ice-cold water on the target spot, which is then removed by suction using a fine pipette after 5 s. (Vorm et al., 1994). Also the interpretation of MALDI mass spectra is relatively easy compared to ESI mass spectra, because MALDI leads usually to single charged ions, leading to single peaks, whereas in ESI multiple charged ions are formed, making the mass spectrum more complicated. Furthermore, most MALDI instruments are equipped with targets that can hold many samples (up to 10,000) simultaneously and automatic sample acquisition and analysis can be done, which ensures high throughput screening of samples. For example, Shevchenko et al. (1996) have reported the analysis of 150 proteins from *S. cerevisiae* from 2-D gels from which 80% could be

identified. In Appendix 2 some database search programs and their availability on the Internet are shown.

Table 1.5 Increasing search accuracy of DNA databases by using orthogonal data sets. By using orthogonal data sets from the protein a high confidence level can be restored. AC stands for ACcession number of the database entry, Pos. the position of the sequence in the search output and Delta is the difference in score between the correct and next highest scoring non-related sequence (adapted from James, 1997b).

			Single digest		Dual digest	
Protein	AC	Digest	Pos.	Delta	Pos.	Delta
Lambda receptor	P02943	LysC/Tryp	1	+ 3.9	1	+ 58.3
Citrate carrier	P31602	Tryp/DxH[a]	2	- 2.1	1	+ 29.8
10 kDa Chaperonin	P15020	V8/Tryp	1	+ 5.6	1	+ 119.8
Na/K-ATPase Alpha1	P06685	LysC/CNBr	1	+ 15.8	1	+ 81.6
Lipid binding protein	P07926	Tryp/+ 42[b]	2	- 7.7	1	+ 40.9
Apolipoprotein AI	P02647	LysC/Tryp	1	+ 10.1	1	+ 102.5
		Average		+ 4.3		+ 72.2

[a] DxH is deuterium exchange

[b] + 42 is acetylation of first digest

Although peptide mass finger printing is becoming widely used in proteome analysis there are problems that can arise. For example, masses from peptides that do not match those in the database may be due to protein modifications (Roepstorff, 1997). Hydrophobic peptides may not be extracted from the gel or give poor ionization during mass spectrometry. Mismatching can be overcome by applying a second, orthogonal set of data from a digest with another enzyme or by deuterium exchange of the first digest (James et al., 1994). The difference in the confidence levels obtained using single and orthogonal database searches is shown in table 1.5. When peptide mass fingerprinting is carried out using DNA databases instead of protein databases the confidence level drops when using a single set of data. In order to extract the protein data from a DNA database all six reading frame translations must be searched. Secondly, the number of entries in DNA databases is much larger, especially when containing EST's.

Figure 1.11 compares the difference in mass accuracy and resolution. High resolution does not necessarily mean high accuracy. Accuracy can only be reached by good calibration methods, either internal or external. For tryptic digests it is very useful to take autolysis products from trypsin as internal standards. The term resolution gives more insight in the distribution of the peptide isotopic masses. At high resolution the isotopic distribution of peptide masses is perceivable, especially for small peptides (< 5000 a.m.u.) this is the case (Vestal et al., 1995).

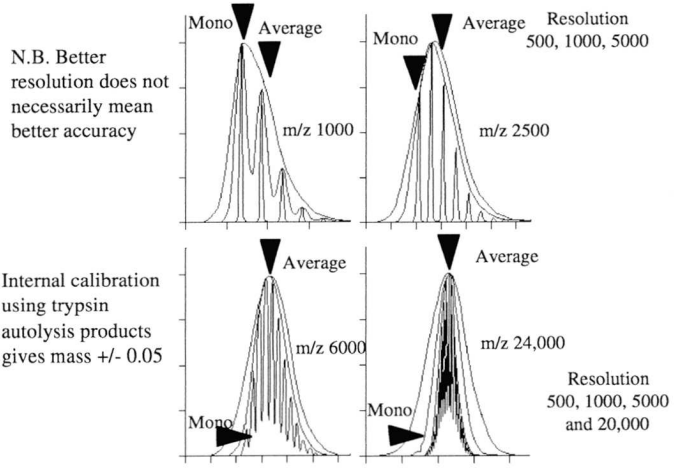


***Figure 1.11* Mass accuracy and resolution.** The mass accuracy and isotope distribution at masses m/z 1,000 - 2,500 - 6,000 and 25,000 a.m.u. is shown. At lower masses (typical for peptides from tryptic digests) the isotopic distribution is clearly resolved, whereas at higher masses the resolution is lower and the isotopic distribution is not seen anymore. It is important to note the difference between monoisotopic mass and the average mass.

One of the main problems of mass mapping is determining the confidence level of the search result. Digests that produce only a few peptides can produce inconclusive results, as can proteins that are not in the database. One established method that greatly increase the confidence levels in database searching is hydrogen-deuterium exchange. The number of hydrogens in a peptide is sequence dependent, so peptides with similar masses may be distinguished after exchange, except proline. The amino acids A, F, G, I, L, M and V all have one exchangeable hydrogen, C, D, E, H, S, T, W and Y have two, K, N and Q have three and R has five exchangeable

hydrogens. After deuterium exchange of a digest, the new masses are matched against protein data in the deuterated protein database (James et al., 1994).

1.5.2 *Peptide fragment mass fingerprinting*

A single MS/MS experiment of a protein digest may contain up to 50 different peptide MS/MS spectra. Manual interpretation of each spectrum to determine the sequence for a database search is very time consuming (about 30 min. per spectrum). One way to overcome this problem is the "shotgun" identification approach (Yates, 1998).

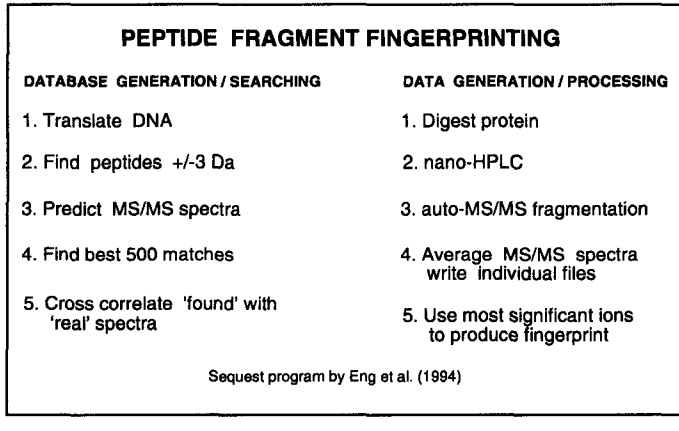


***Figure 1.12* SEQUEST MS/MS database search outline.** The figure outlines the data processing and searching procedure using the program SEQUEST (Eng et al., 1994) to allow fully automated protein identification using uninterpreted MS/MS data (adapted from James, 1997b).

Proteins are subjected to protein digestion to yield a complicated mixture of peptides. These peptides are then analysed by HPLC, coupled to a tandem mass spectrometer. The obtained MS/MS spectra are, without interpretation, used to search the algorithm SEQUEST, to correlate these uninterpreted MS/MS data of peptides to sequences in a protein database in a fully automated fashion (Eng et al., 1994; Yates et al., 1995a). The experimentally obtained spectrum from each parent ion is compared with the theoretically predicted spectrum for that sequence and the best matches are subjected to cross correlation analysis. Figure 1.12 outlines the steps in database searching using MS/MS data according to the SEQUEST approach.

In a single HPLC run over 100 MS/MS spectra can be obtained and processed.

The algorithm was extended to deal with MS/MS data of modified peptides (Yates et al., 1995a) and to search DNA databases (Yates et al., 1995b). The technique is very powerful when inaccurate or small DNA sequence stretches are being searched and normal protein fingerprinting data fails.

A different approach is searching the database with the use of peptide sequence tags as shown in figure 1.13 (Mann and Wilm, 1994). The MS/MS spectrum must be manually interpreted to find a group of ions which form a series, from which a small sequence, the tag, can be derived and used with the intact peptide mass and the tag sequence start and end masses to search the database using the program PepSearch. Even when incomplete fragmentation occurs, or noisy MS/MS spectra are obtained, enough peaks are resolved to assign the short sequence tag. The power of both the SEQUEST and the PepSearch algorithm is that they can be used for identification of proteins in complex mixtures (McCormack et al., 1997). The type of ions (b- or y-ions), and thus the ion series to assign a sequence, obtained in MS/MS spectrum depends on the collision energy used. Special versions of the SEQUEST program, that take into account which fragmentation energy is used, have been developed for database searching with MALDI-PSD spectra (Griffin et al., 1995) and tandem mass spectra (Yates et al., 1996).

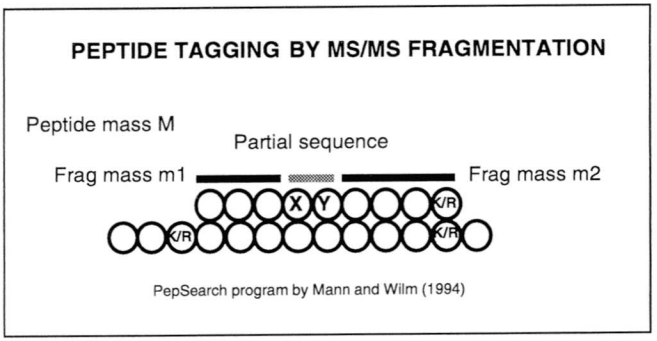


***Figure 1.13* Peptide tag searching.** The figure shows the data parameters used by the program PeptideSearch (Mann and Wilm, 1994). Each MS/MS spectrum is manually inspected and four search parameters are extracted: parent mass, mass at the start of sequence, a small sequence found by the inspection (the tag) and the mass at the end of the sequence.

Another way to simplify the interpretation of MS/MS spectra is the use of isotopic labelling techniques. This was demonstrated by using ^{18}O containing water when performing tryptic digestion. The ^{18}O atom is then incorporated into the C-terminal carboxy groups of the peptides. Subsequent fragmentation by MS/MS reveals a pair of y-ions separated by a characteristic 2 mass unit shift, which facilitates interpretation of the MS/MS spectra (Shevchenko et al., 1997). While this technique simplifies spectrum interpretation, the use of a scanning analyser, such as the quadrupole, limits the sensitivity that can be obtained. For this reason the quadrupole time of flight (Q-TOF) mass spectrometer is developed. It combines the quadrupole with the reflector time-of-flight analyser (Verentchikov et al., 1994). This new concept combines the simplicity of a quadrupole with the inherent sensitivity (up to 100 fold), resolution and mass accuracy of an orthogonal acceleration TOF analyser. High mass resolution and accuracy were reported, as expected from a TOF instrument. This type of mass spectrometer will play an important role in rapid *de novo* sequencing of proteins (Shevchenko et al., 1997).

Increasing the reliability of the database search can be done by generating a short sequence tag on the N- and/or C-terminal of the peptides from the protein digest. Since protein sequence is extremely specific a short sequence results in reliable protein identification. In Part A this will be described in more detail. It is clear that mass spectrometry and protein databases play a crucial role in protein identification on a large scale.

1.6 Perspectives

1.6.1 Outlook for proteome analysis

The last ten years or so have shown a dramatic change in analytical protein chemistry. Two-dimensional gel electrophoresis is the technique of choice for the separation of complex protein mixtures from cells or tissue and mass spectrometry is by far the fastest and most sensitive technique for the identification of proteins.

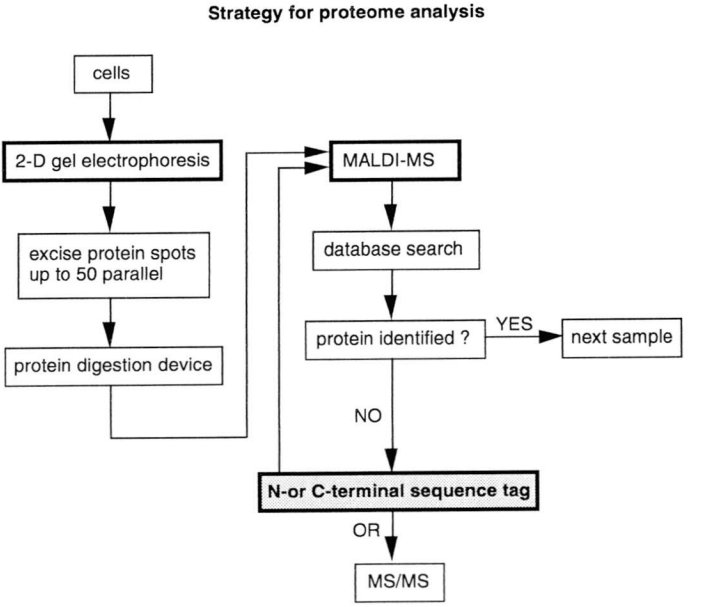


Figure 1.14 **Strategy for proteome analysis.** This scheme outlines the steps that are involved, the two main techniques in proteome analysis are two-dimensional gel electrophoresis and MALDI mass spectrometry in this flow scheme. Part A of this thesis is devoted on the development of a novel N-terminal sequence tag generation method (shaded box). Other techniques in this flow scheme are either still in developmental stage (protein digestion device) or are subject to automation (MS/MS and database search).

Figure 1.14 summarizes the main steps in our strategy for proteome analysis. Some techniques are still in early stages, such as the protein digestion device. Using this device the proteins are concentrated and

digested with a protease, such as trypsin. The resulting peptides are directly bound on reversed-phase material and can be both measured directly by MALDI-MS and subjected to partial N-terminal degradation before analysis by MALDI-MS.

When handling proteins and peptides at the low picomole level special care must be taken. Proteins and peptides at these concentrations are easily lost, for example on the walls of the tubes. Adding detergents or siliconization of the tubes can overcome this, but these interfere drastically with the MS techniques available. Therefore, sample handling must be minimized at all time (Staudenmann et al., 1998). Concentration of proteins serves another goal. Most proteases have a K_m between 5 - 50 µM. Low abundant proteins isolated from 2-D gels sometimes have a lower concentration than this K_m. The following formulas illustrate the enzyme kinetics as described by Michaelis-Menten:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P$$

$$V = V_{max} \frac{[S]}{[S] + K_m}$$

According to Michaelis-Menten, when S >> E, so ES dissociation is independent of [S]. But if S << K_m, then $V_0 = k_0 / K_m$ [ET][S], where ET = E + ES (Michaelis and Menten, 1913; Stryer, 1987). In words, this means that when the protein concentration is considerably lower than the K_m for the protease, no digestion will occur, so protein concentration prior to digestion is required. A prototype of such a concentrater/digester is developed in our laboratory. Since a patent is pending no figure can be shown of the prototype. The main advantage of this single step approach is that sample handling is minimized.

Miniaturization and increased throughput will become more important in the near future. Miniaturization can be reached by developing microchips for high speed separation and sensitive detection of proteins (Manz, 1997). Figeys et al. (1997, 1998a, 1998b) recently described several integrated microfabricated fluidics-mass spectrometry devices for automated protein analysis. These microfluidics devices are modules of specific function and design which are micromachined by photolithography/etching

of glass. Samples, typically tryptic digests of gel separated proteins are applied to specific reservoirs and directed by computer-controlled electro-osmotic pumping through a network of etched channels (30 μm deep) to the MS detector where selected peptide ions were subjected to tandem mass spectrometry. The whole process is fully automated. Limits of detection were in the low femtomole range (Figeys et al., 1998b). The developments in technology for high-throughput screening are rapid rising with advances in robotics and miniaturization. For example, the conventional 96-well plate can be replaced using robotics with the 9,600-well plate for screening purposes (Persidis, 1998).

1.6.2 The ultimate goal of proteome analysis

Reaching the goals for protein identification in proteome projects is not an end in itself. The ultimate goal is to understand how proteins function. To reach this it is necessary not only to know its amino acid sequence (primary structure), but also the shape of the protein and how and where it interacts with other proteins or substrates. This is approached by first resolving the three-dimensional (3-D) structure of the protein and then changing the protein sequence to determine how the structure and function is changed. Site-directed mutagenesis experiments are generally designed to provide information about essential amino acid residues in enzyme-based reactions, molecular recognition, protein stability and drug design. However, the experimental determination of the 3-D structure of a protein either by X-ray crystallography or by NMR techniques is not an easy task, because of difficulties in obtaining sufficient amounts of pure protein or diffracting crystals and the size of the protein. This is illustrated by the fact that the SwissProt/trEMBL database contains about 200,000 sequences (Bairoch and Apweiler, 1997), while the Brookhaven Protein Data Bank (PDB) contains about 6,000 3-D structures from only 1,700 different proteins (Abola et al., 1996).

While the number of solved 3-D structures of proteins is growing very slowly compared to the rate of sequencing novel cDNA's there is a growing interest in methods for predicting 3-D structures of proteins. One of the most reliable methods is comparative protein modelling. This is based on the finding that proteins with similar sequences have similar structures (Chothia and Lesk, 1986). But protein modelling requires expensive computer hardware and software and expert knowledge how to manipulate

structural information. In order to simplify this there is software available nowadays to many scientists for comparative protein modelling. This software can be found on the following web sites, http://www.expasy.ch/swissmod/SWISS-MODEL.html and http://www.expasy.ch/spdbv/mainpage.html (Peitsch and Guex, 1997).

At this point it may be clear that it is important to link the different databases that are available for protein identification. The introduction of the World-Wide Web (WWW) in recent years greatly enhanced the power of cross-references in databases (Appel et al., 1993, 1994). Using special graphical browsers such as Netscape Navigator it became very easy to navigate across different databases that are available on the Internet through the display of active hypertext cross-references. There is no need to download databases and keep copies of databases up-to-date on local computers.

Proteomics results into data that make biology better defined. It is the core of functional genomics and will have a dramatic impact on all fields of biological sciences in the future.

2. Part A: Partial N-terminal Sequencing

In this section the use of short amino acid sequence tags as a tool for protein identification is described. It deals with the description of a novel thioacetylating method for N-terminal degradation on peptides or protein digests in a parallel way to generate a short sequence ladder to be used for rapid protein identification by searching the database. For this purpose a new algorithm MassDynSearch, which searches the SwissProt and trEMBL by peptide mass after N- or C-terminal degradation, was developed.

2.1 Introduction

2.1.1 Peptide sequence tags

As mentioned before, protein sequence is highly specific. For example, a sequence of three amino acids results in 8,000 possible combinations (20^3), four amino acids in 160,000 (20^4) and five amino acids in a huge 3,200,000 (20^5) combinations.

A powerful approach to the exploration of the coding potential of a genome is the use of expressed sequence tags (EST's) (Adams et al., 1991). EST's are short sequences of 250 to 400 base pairs obtained by random, single-pass sequencing of cDNA libraries. Many eukaryotic genomes have been or currently are the target of large scale EST sequencing projects. Currently, there are about two million EST's in the nucleotide sequence databases of which most are derived from human cDNA libraries. By definition, EST's are partial sequences and do not code for complete protein sequences (except in some rare cases). Since the tags have a relative high error rate (about 2%) because of the single-pass sequencing they seem not very useful, but in conjunction with high throughput partial sequencing using tandem mass spectrometry, EST databases are used where normal protein fingerprinting data fails (see section 1.5). Searching programs have been developed that scan all six potential reading frames of EST's to find matches with partial protein sequences (James et al., 1994; Gevaert et al., 1996). However, although the number of EST's in databases is very large this does not mean that every protein is represented. Proteins produced from low abundance mRNA's may not be present in EST databases. Another possibility is that

mRNA sequences are present that contain only non-coding information from the 3' region.

Small sequences can be generated by tandem MS fragmentation of peptides followed by database searching (Mann and Wilm, 1994). More recently, short sequences from either the N- or C-termini from intact proteins have been proposed as tools for protein identification (Gooley et al., 1997; Wilkins et al., 1996b, 1998). The specificity of protein terminal sequence tags is rather high, 60% for N-terminal tags and up to 90% for C-terminal tags of four residues when searching in protein databases restricted to *E. coli* proteins only. This confidential level drops when searching in non-redundant databases. Most protein N-terminal sequence tags have been generated by conventional Edman degradation on proteins blotted onto membranes. This is a very reliable method although being slow and expensive. For high throughput it is less useful (Wilkins et al., 1996c; Gooley et al., 1997). Again, as C-terminal sequencing is not as common as N-terminal sequencing, generating protein C-terminal sequence tags remains difficult. Chemical C-terminal sequencing requires more material in the order of hundreds of picomoles (Burkart et al., 1996). Carboxypeptidase digestion is successful for generating sequences in peptides and small proteins (Patterson et al., 1995; Thiede et al., 1997).

2.1.2 Ladder sequencing

An alternative method to peptide sequencing by mass spectrometry instead of using MS/MS or PSD is the chemical or enzymatic degradation of peptides to generate a set of peptides. The difference in the masses between adjacent peaks define the amino acid removed and thus the sequence (figure 2.1). In Appendix 3 a list of the exact masses, both monoisotopic and average masses, of the amino acids is presented. To generate a ladder of degradation products, sequencing with a low repetitive yield is required. The principle of low repetitive yield ladder sequencing chemistry is illustrated by the following example: if a model peptide ABCDEFG is degraded with a repetitive yield of 30% then after four cycles the theoretical distribution of peptides will be 24.0% intact ABCDEFG, 41.2% BCDEFG, 26.4% CDEFG, 7.6% DEFG and 0.8% EFG respectively.

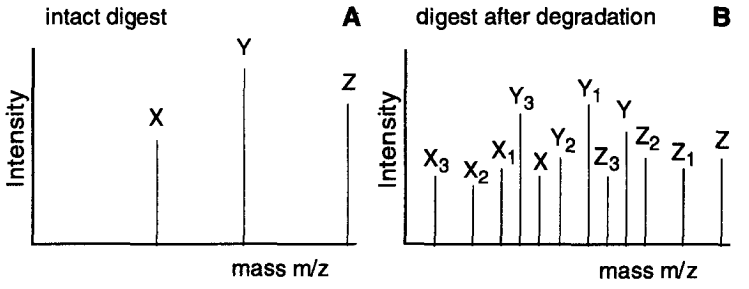


***Figure 2.1* Principle of the ladder sequencing.** (A) shows schematically a typical mass spectrum of a tryptic protein digest and (B) shows the same digest after three degradation cycles. The mass differences between for example X, X_1, X_2 and X_3 define a short sequence tag which can be used to search the database.

Since the availability of MALDI-MS the idea of ladder sequencing gained interest. Chait et al. (1993), who defined the concept of ladder sequencing first, used a modified variant of the Edman degradation by introducing a small amount of a sequence terminator (phenylisocyanate) in each cycle to create a ladder. This requires a careful balance between the reaction kinetics of sequencing and blocking. The method has some disadvantages. The degradation is carried out on a membrane and it is difficult to recover all peptides. For low level analysis this can cause problems. Due to the terminating reagent (PIC) the N-terminal of the set of peptides are blocked by a non charged group. This causes a drop in sensitivity in most mass spectrometric methods. Alternatively, a ladder can be generated by the addition of new peptide after each cycle (Bartlet-Jones et al., 1994). They used a volatile fluoroisothiocyanate avoiding the problem of adding a non charged group on the N-terminal of the peptides. Instead, after generating the ladder the free N-termini are modified with a quaternary alkylamine to improve ionization in mass spectrometry. Main drawback of this method is the multiple sample handling which leads to potential sample losses when using low levels of peptides. In this thesis a novel method for ladder sequencing is described that should overcome the disadvantages of these two methods. This is done by immobilizing the peptides on reversed-phase material and using a thioacetylation method for N-terminal degradation.

2.1.3 A thioacetylation method for partial N-terminal sequencing

Although the most popular step-wise degradation method is Edman degradation there are some limitations to this procedure when extending it to ladder generating procedures. An alternative chemistry is a thioacetylation method. The possibility of using N-thioacyl peptides for sequence analysis as an alternative to isothiocyanate degradation dates back to Kenner and Khorana (1952), who noted the simplicity how to regenerate the free amino acids in their attempt to stepwise degrade peptides from the N-terminal. Selective removal of the terminal amino acid involves two stages: 1) attack by a reagent which either modifies the terminal amino-acyl group so as to reduce its stability or attaches a residue so constructed that its reaction with the neighbouring peptide linkage is facilitated by formation of a stable cyclic structure and 2) removal of the modified amino acid (Kenner and Khorana, 1952). It was then shown by Barrett (1967) that a N-thiobenzoyl group could be removed from the peptide in a similar way as the phenylthiocarbamoyl peptide in Edman degradation by acid hydrolysis with trifluoroacetic acid.

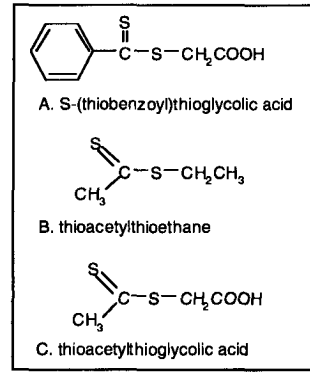


***Figure 2.2* Thioacetylating compounds.** *These compounds are used in this study.* (A) S-(thiobenzoyl)thioglycolic acid and (B) thioacetylthioethane are commercially available, (C) thioacetylthioglycolic acid was synthesized according to Mross and Doolittle (1977).

Figure 2.2 shows the thioacetylating compounds that were used in this study. The method of degradation by thioacetylation was chosen since it allows to perform the chemistry entirely in the aqueous or gas phase. For our approach this was a necessary prerequisite because the peptides from a

protein digest are immobilized on reversed-phase material prior to the degradation reaction. Immobilizing the peptides minimizes sample handling and thus potential losses of material. Furthermore, excess reagents and reaction by-products which are water soluble can easily washed away with water while the peptides remain bound to the reversed-phase material.

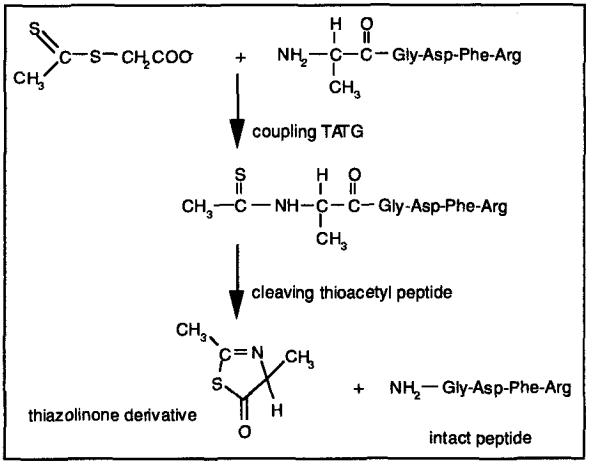


***Figure* 2.3 Thioacetylation chemistry.** Scheme of the main reactions: coupling of TATG in alkaline conditions to the N-terminal part of the peptide or protein followed by cleavage of the first amino acid residue by strong acid and release of the thiazolinone. Note the similarity between the Edman degradation (figure 1.5).

Figure 2.3 shows the reaction of the thioacetylation procedure (Barrett, 1967). Thioacetylthioglycolic acid (TATG) can be considered as an active ester of thionacetic acid and as such can react with various nucleophiles. Under alkaline conditions the TATG is coupled to the unprotonated N-termini of the peptides followed by a cleavage step with trifluoroacetic acid (TFA) gas. This releases a thiazolinone, which can be washed away. The chemistry is very similar to the Edman degradation.

After obtaining the sequence tags of the peptides a database search is performed. For this purpose the algorithm MassDynSearch was developed (Korostensky et al., 1998). The algorithm uses a combination of peptide masses and a few associated sequence tags to match proteins in the SwissProt or trEMBL database. Including the sequence tag highly improves the accuracy of the mass fingerprinting as described before.

2.2 Materials and Methods

2.2.1 Materials

Acetonitrile, mercaptoacetic acid, N-ethylmorpholine, pyridine, thioacetylthioethane and trifluoroacetic acid (for protein sequence analysis) were purchased from Fluka AG (Buchs, Switzerland). α-cyano-4-hydroxycinnamic acid, N-methyl-piperidine and S-(thiobenzoyl)thioglycolic acid were purchased from Aldrich GmbH (Buchs, Switzerland). Trizma base and angiotensin were purchased from Sigma Chemical Co. (Buchs, Switzerland). Ammonium hydrogen carbonate, β-mercaptoethanol, sodium sulphide and sodium sulphate were purchased from Merck AG (Darmstadt, Germany). Acetic acid, carbon tetrachloride, diethyl ether, hydrochloric acid (min. 37%) and HPLC grade acetonitrile were purchased from Riedel-de Haën AG (Seelze, Germany). Sequencing grade modified trypsin was purchased from Promega (Zürich, Switzerland). DNase was purchased from Boehringer (Mannheim, Germany). Nucleosil 100 - 2540 C_{18} reversed-phase resin was purchased from Macherey-Nagel AG (Oensingen, Switzerland) and Poros 10 R2 Reversed-Phase Packing was purchased from Perseptive Biosystems (Framingham, MA, USA). The synthetic peptide C21W was a kind gift of Holger Schmid (Institute of Biochemistry III). All other chemicals were of the highest purity commercially available.

2.2.2 *Synthesis of thioacetylthioglycolic acid (TATG)*

The synthesis was performed as described before (Mross and Doolittle, 1977). Acetonitrile reacted with mercaptoacetic acid to yield carboxymethylthioimidate which was than converted to the product by treatment with H_2S in pyridine at 0 °C (figure 2.4).

Mercaptoacetic acid (46 g, 0.50 mole) and acetonitrile (22.5 g, 0.55 mole) were mixed in a reaction flask and cooled to ice temperature. The mixture was overlaid with 2 - 3 cm petroleum ether and HCl gas was bubbled for 1 h through the lower phase until a white solid appeared (carboxymethyl-thioimidate). The solvent was then evaporated under vacuum with a rotary evaporator and 250 mL of dry pyridine was added. H_2S was bubbled through for 4 h until the reaction was completed. H_2S was generated by dropping concentrated hydrochloric acid to a concentrated solution of Na_2S. During

this time the white slurry liquified and turned yellow, following which a white precipitate (ammonium chloride) appeared.

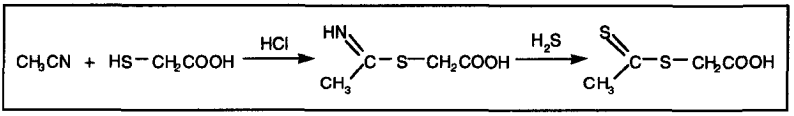


***Figure 2.4* Scheme of the synthesis of thioacetylthioglycolic acid (TATG).** Acetonitrile reacted with mercaptoacetic acid to yield carboxymethylthioimidate which was than converted to the product by treatment with H_2S in pyridine at 0 °C.

When the reaction was completed, 100 mL ice water was added and stirred to dissolve the ammonium chloride. The liquid was then poured over a mixture of 300 mL concentrated hydrochloric acid, 100 mL water and 300 g ice. The product was extracted with 300 mL diethyl ether. The aqueous phase was extracted twice more with 150 mL diethyl ether each. The combined diethyl ether layers were then washed with 3 M hydrochloric acid, following which the diethyl ether was dried with $NaSO_4$. The diethyl ether was removed by rotary evaporation, yielding about 20 g of an oil. TATG was crystallized from warm carbon tetrachloride, the yield was 5 g of yellow solid. ^{1}H-NMR and ^{13}C-NMR as well as ESI-MS were done to check the purity of the product.

^{1}H-NMR ($CDCl_3$, 300 MHz): δ = 2.89 (s, CH_3), 4.13 (s, CH_2) ppm.
^{13}C-NMR ($CDCl_3$, 75 MHz, proton decoupled): δ = 38.57 (CH_3), 38.86 (CH_2), 173.9 (COOH), 230.8 (C=S) ppm.
ESI-MS: 45 (6.88), 59 (80.9), 76 (4.86), 91 (7.51), 117 (4.83), 150 (100) M^+.

2.2.3 *Isolation of ribosomes from E. coli*

Escherichia coli MC4100 [F^- *araD*139 *Δ*(*argF-lac*) U169 *rspL*150 *relA*1 *deoC*1 *ptsF*25 *rpsR* *flbB*5301] was obtained from the Cold Spring Harbor Laboratory collection (Silhavy et al., 1984). Bacteria were cultivated in a sulphur-free, synthetic glucose-salts medium, with the addition of 500 μM inorganic sulphate as described before (Kertesz et al., 1993). The cultures were grown aerobically on a rotary shaker (180 rpm) at 37 °C, and growth was monitored spectrophotometrically at 650 nm. Cells were harvested in the mid-exponential phase (A_{650} = 0.5) by centrifugation at 9,000 rpm for 10 min. and washed with 50 mM Tris-HCl, pH 7.0. The cells (10 g wet cells)

were resuspended in 70 mL extraction buffer (20 mM Tris-HCl, pH 7.4, 40 mM NH_4Cl, 10 mM $MgCl_2$ and 7 mM β-mercaptoethanol). One tablet of a cocktail of protease inhibitors (Boehringer, Mannheim, Germany) was added and the cells were ruptured by two passes through a chilled French pressure cell at 20,000 psi. To the ruptured cells 10 µL DNase (10 µg/µL) was added and incubated at 37 °C for 40 min. Cell debris was removed by two centrifugation steps at 16,000 rpm for 30 min. at 4 °C. Ribosomes were isolated as previously described (Traub et al., 1979). The supernatant was centrifuged at 45,000 rpm for 2 h at 4 °C. The pellet contains the crude ribosomes and were resuspended in 20 mL high salt buffer (20 mM Tris-HCl, pH 7.4, 400 mM NH_4Cl, 10 mM $MgCl_2$ and 7 mM β-mercaptoethanol). Then, 5 mL of the crude ribosomal solution were layered onto a 7 mL 17.5% sucrose cushion (in high salt buffer) and centrifuged at 45,000 rpm for 3 h at 4 °C. The pellet (containing 70S ribosomes) was resuspended in 10 mL low Mg^{2+} buffer (20 mM Tris-HCl, pH 7.4, 40 mM NH_4Cl, 0.3 mM $MgCl_2$ and 7 mM β-mercaptoethanol) and layered onto four continuous 10-30% sucrose gradients (in low Mg^{2+} buffer) and centrifuged in a swing-out rotor at 18,000 rpm for 14 h at 4 °C. In the presence of low Mg^{2+} the 70S ribosomes dissociate into the 50S and 30S subunits. One mL fractions were taken and checked for protein content at 260 nm. No real separation between the two subunits was obtained and so the fractions containing protein were pooled (60 mL total) and dialysed overnight against 5 L of low Mg^{2+} buffer. The dialysed solution was centrifuged at 45,000 rpm for 4 h at 4 °C and the ribosomes were resuspended in 7.5 mL of low Mg^{2+} buffer and stored at -20 °C. Protein was determined according to a modified method of Lowry (Markwell et al., 1978) as 7.5 mg/mL.

2.2.4 Separation of ribosomal proteins by reversed-phase HPLC

To separate the ribosomal proteins by reversed-phase HPLC the RNA had to be extracted first. This was done by adding $MgCl_2$ and acetic acid to the ribosomes to a final concentration of 67 mM and 67% respectively and followed by incubation on ice for 1 hour. The mixture was centrifuged at 10,000 rpm for 10 min. at 4 °C. The procedure was repeated once more with the pellet (Hardy et al., 1969). The supernatants were combined and concentrated in the speed-vac. The proteins were redissolved in 3% acetic acid, centrifuged at 10,000 rpm for 20 minutes at 4 °C to remove any undissolved particles and injected onto a preparative reversed-phase HPLC

system (L-6220 Intelligent Pump, L-4250 UV-VIS Detector, Merck AG, Darmstadt, Germany) in the following conditions: a gradient of 10 - 25% B in 30 min., 25 - 35% B in 40 min., 35 - 36% B in 30 min., 36 - 40% B in 40 min., 40 - 55% B in 60 min. and 55 - 90% B in 40 min. at 2 mL/min. (A = 0.1% TFA, B = 80% acetonitrile/0.08% TFA), a C_{18} preparative column (250 x 21 mm, Nucleosil 100 - 12 µm, Macherey-Nagel AG, Oensingen, Switzerland). The absorbance was measured at 220 nm. Fractions of 2 mL were collected (Kamp et al., 1984; Kamp and Wittmann-Liebold, 1984).

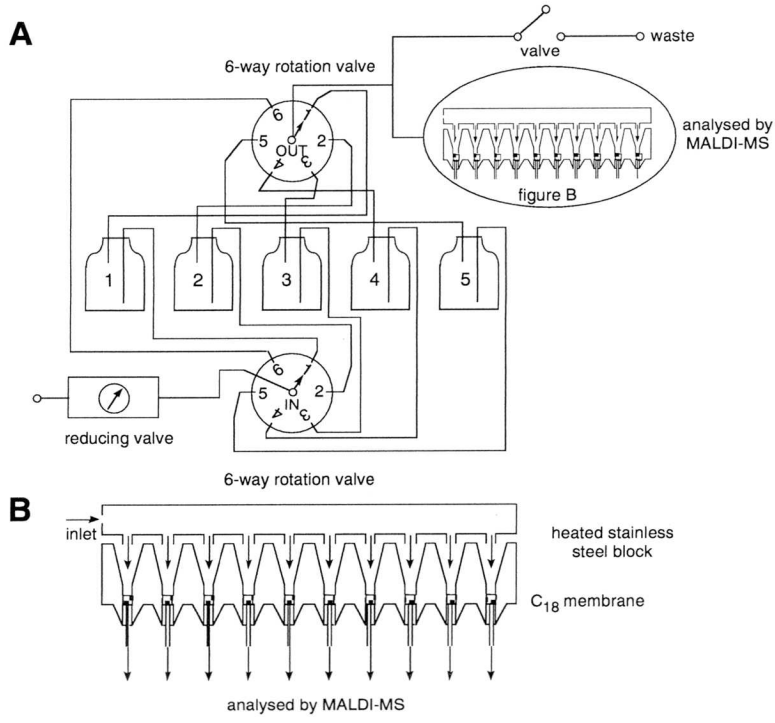
2.2.5 *Tryptic digestion of ribosomal proteins*

Fractions containing protein from the preparative reversed-phase HPLC were dried in the speed-vac and redissolved in 100 µL water. For tryptic digestion in solution 20 µL of protein was taken (assume approximately 120 µg protein), porcine trypsin (sequencing grade) was added at 2% with respect to the ribosomal protein in 100 mM NH_4HCO_3 buffer, pH 8.0 final concentration. Digestion was performed for 48 h at 37 °C. The solutions were dried in the speed-vac to remove the volatile buffer, washed with 100 µL water, dried again and finally the peptides were redissolved in 50 µL water and stored at -20 °C. The peptides obtained from the ribosomal proteins were used for partial N-terminal sequencing.

2.2.6 *Ladder generating chemistry on peptides*

Two devices were constructed, one to deliver N-methylpiperidine (NMP) or N-ethylmorpholine (NEM) and trifluoroacetic acid (TFA) in the gas phase and another to perform parallel on ten samples the coupling and cleaving reactions respectively (figure 2.5).

Standard peptides or tryptic digests (usually 50 - 100 pmoles in 0.1% TFA) were loaded slowly to allow proper binding of the peptides (2 - 5 µL/min.) onto PTFE membranes with embedded C_{18} reversed-phase material (Empore 3M C_{18} Extraction Disk from Varian, Zug, Switzerland) on a multiple sample loader and washed with 50 µL 0.1% TFA. Nitrogen gas at 0.5 bar pressure was used to deliver either gas or liquid reagents. The ladder generating steps were performed at 40 °C in the following conditions:


A
waste
valve
6-way rotation valve
analysed by MALDI-MS
figure B
OUT
IN
1
2
3
4
5
reducing valve
6-way rotation valve
B
inlet
heated stainless steel block
C_{18} membrane
analysed by MALDI-MS
C

(a) washing with N_2 gas for 10 min. at 1.0 bar, (b) NMP base (gas) delivery for 3 min. at 0.5 bar, (c) loading of 2 µL 50 mM TATG in 1% NEM (cycle 1 and 2) or 2 µL 100 mM TATG in 2% NEM (cycle 3 and 4) and incubating 5 min., (d) washing with N_2 gas for 5 min. at 1 bar, (e) TFA gas delivery for 3 min. at 0.5 bar, (f) washing with N_2 gas for 10 min. at 0.5 bar and (g) washing with 50 µL water to remove any side products. Then, the cycle was repeated (steps a to g). When the ladder generating chemistry was completed the ragged peptides were slowly eluted with 50 µL 60% acetonitrile/0.1% TFA and measured by MALDI-MS.

2.2.7 MALDI-TOF mass spectrometry

The stepwise eluted degraded peptides were spotted onto the 100-position MALDI sample target. For each sample approximately 0.5 µL was required and the same amount of a saturated matrix solution was added (10 mg α-cyano-4-hydroxy-cinnamic acid in 50% acetonitrile and 1.25% TFA in water) and allowed to dry at ambient temperature.

To obtain better spectra the spots on the MALDI target were washed by pipetting ice-cold water on the target spot, which was then removed by suction using a fine pipette after 5 s.

Mass spectra were recorded using a Voyager Elite MALDI-TOF mass spectrometer (Perseptive Biosystems, Framingham, MA, USA). Samples were analysed in delayed extraction reflector mode using an accelerating voltage of 20 kV, a pulse delay time of 75 ns, a grid voltage of 55% and a guide wire voltage of 0.05%. Spectra were accumulated for 32 or 64 laser shots. For each sample two datasets were obtained, the mass spectrum of the tryptic digest of the protein and the mass spectrum of the stepwise degraded peptides of the digest.

***Figure* 2.5 Prototype of gas delivery apparatus and multiple sample loader.** (A) Scheme for the gas delivery apparatus. Up to five different chemicals can be delivered in the gas phase by turning the two 6-way rotation valves (IN and OUT), each position of the valves corresponding to a different chemical in the equal numbered bottle. Position 6 is used to flush the system with N_2 gas. The chemicals are then introduced into the multiple sample loader, which is shown enlarged in (B). The multiple sample loader is made of stainless steel containing two heating elements inside to control the reaction temperature at 40 °C. Inside the multiple sample loader is placed a polymer membrane with the reversed-phase material C_{18} embedded. For each sample this membrane is renewed to avoid cross contamination. Ten reactions can be performed in parallel. (C) Photograph of the set-up of the equipment for partial N-terminal sequencing which shows the gas delivery apparatus and the multiple sample loader. The complete set-up is placed in a fume-hood.

2.2.8 ESI-MS/MS sequencing

Tryptic digests (picomole amounts) from the ribosomal proteins were desalted on the C_{18} membranes and redissolved in 50% methanol and 0.5% acetic acid before introducing into the mass spectrometer at a flowrate of 0.2 µL/min. with a syringe pump. MS/MS sequencing was performed on a Finnigan MAT LC-Q ion trap mass spectrometer (San Jose, CA, USA). The peaks of interest were selected with a mass window of ± 5 a.m.u. and fragmentation of each peptide was established using a relative collision energy of 35 - 60 for MH^+-ions and 20 - 30 for MH^{2+}-ions. Each protein digest yielded 5 - 20 MS/MS spectra, taking on average 10 min. to record.

The obtained MS/MS fragmentation spectra were used to search the protein and nucleic acid databases using the program SEQUEST in a fully automated mode (Eng et al., 1994).

2.2.9 N-terminal Edman sequence analysis

Intact ribosomal proteins (0.5 - 1 µL) from the preparative HPLC separation were spotted onto PVDF membrane. N-terminal sequence analysis was directly performed on a Hewlett Packard G1000A protein analyser, equipped with four cartridges. Released PTH amino acids were analysed on a Hewlett Packard HPLC series 1100 (Palo Alto, CA, USA). Five to six sequence cycles were performed according to the standard protocols provided by the manufacturer.

2.2.10 The MassDynSearch program

The database searching program MassDynSearch was written using the DARWIN language. DARWIN stands for Data Analysis and Retrieval With Indexed Nucleotide/peptide sequences database searching language. MassDynSearch searches the SwissProt and trEMBL by peptide mass after N- or C-terminal degradation. The algorithm, that uses multiple sequence tags for database searching was developed in collaboration with the group of Prof. Gaston Gonnet (Computation Biology Research Group, ETH Zürich). A detailed description of the algorithm was recently published (Korostensky et al., 1998). Information for obtaining the code for MassDynSearch is available at the web site, http://cbrg.inf.ethz.ch. Appendix 4 shows an example of the

MassDynSearch page on the WWW, as shown with Netscape Navigator. Note that this site is still under construction.

2.3 Results and Discussion

2.3.1 Discussion of progress of ladder generating chemistry

In this section the results of the progress of developing the ladder generating chemistry are presented. Where the conditions are not the same as described in section 2.2.6 this is stated. In general, trials were performed on a synthetic peptide C21W, which is part of the calmodulin (CaM) binding domain of the Ca^{2+}-ATPase and has the amino acid sequence WFRGL NRIQT QIRVV NAFRS S and a monoisotopic mass of 2547.39 a.m.u. In figure 2.6 the mass spectrum is shown of this synthetic peptide.

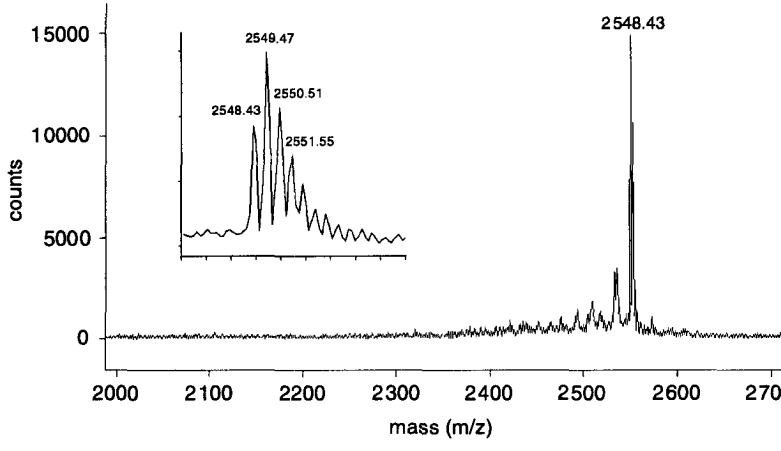


***Figure 2.6* Mass spectrum of the synthetic peptide C21W.** The measured mass is 2548.43 a.m.u. (this is the single charged MH^+-ion). From the measured value and its theoretical mass it can simply be calculated that the deviation of the mass is only 0.0015% (15 ppm). The inset shows an enlargement of the mass range around 2550 a.m.u. to show at high resolution the isotopic distribution of C21W. This is a typical MALDI mass spectrum, acquired in the reflectron mode.

The initial experiments to generate short sequence tags were done with the peptide C21W and the commercially available thioacetylating reagent S-(thiobenzoyl)thioglycolic acid (TBTGA) in solution. These manually performed reactions carried out in an eppendorf tube resulted in the generation of a ladder sequence of three to four residues (data not shown). Since these preliminary experiments were promising, the same

kind of reactions were repeated on the peptide, immobilized on reversed-phase material. Both Nucleosil C_{18} and Poros 10 R2 have been tested for this purpose. Small capillary columns (I.D. = 150 μm, L = 3 - 10 mm) were prepared using a laser puller to create a cone small enough to keep the resin, but wide enough to give a good flow. The columns were packed with the resin followed by loading the peptide on the resin. Such small capillaries can bind up to 500 pmoles of peptide. It must be noted that in many cases these small capillaries show a high back pressure. Problems arised when TFA in the liquid phase was used to cleave the N-terminal amino acid as a thiazolinone. Either the resin is not resistant against pure TFA at higher temperatures or peptides simply are eluted off the resin because of the hydrophobic character of the TFA. It turned out to be necessary to use TFA in the gas phase for the cleavage reaction. A small device was constructed in house to make TFA gas delivery possible. This was achieved by pipetting a small volume of liquid TFA onto a glass filter. Nitrogen gas was directed through this filter and evaporated TFA gas was then directed through the capillary with reversed-phase material to cleave the residue off. Then, the reaction cycle can be repeated to generate a short sequence.

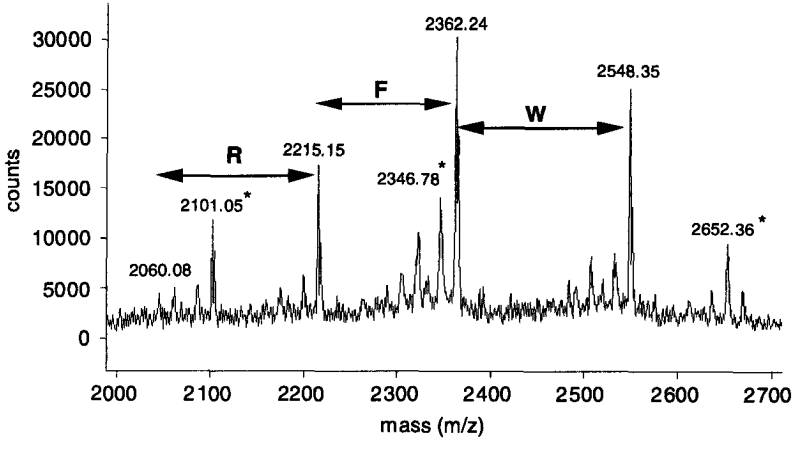
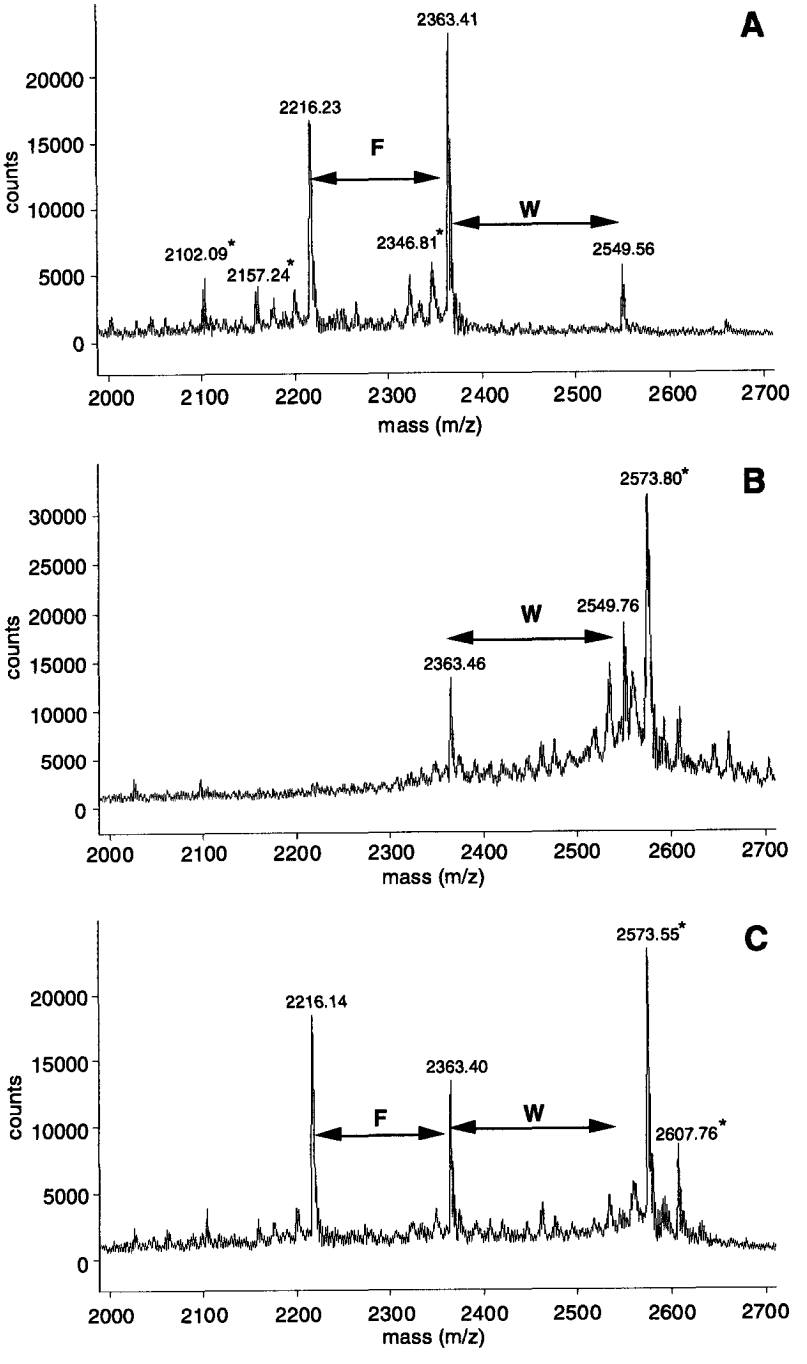


***Figure 2.7* Ladder generation on C21W immobilized on reversed-phase material.** Three reaction cycles were performed using the reagent TBTGA while the peptide was immobilized on Poros 10 R2 material. The peaks marked with an * are either products of side reactions or incomplete cleavage reactions. For example, the peak at m/z = 2652.36 is the peptide mass plus uncleaved TBTGA. Clearly, the N-terminal amino acid sequence WFR can be read from this mass spectrum (compare with the mass spectrum of C21W in figure 2.6).

Figure 2.7 shows an example of a mass spectrum of an experiment when three reaction cycles were performed using the reagent TBTGA on C21W, immobilized on Poros 10 R2 reversed-phase material. Although the mass spectrum in figure 2.7 shows a nice partial sequence there were some problems when extending the technique to protein digests from for example myoglobin or calmodulin. The peptide C21W is hydrophobic and therefore binds strongly to the reversed-phase material.

However, some peptides from protein digests are more hydrophilic and tend to bind less well or efficiently to the capillary column, resulting in considerable loss of peptides after performing the ladder generation chemistry using the hydrophobic reagent TBTGA. Since TFA cleavage has to be done in the gas phase this lead to the idea of turning the chemistry completely into the gas phase. Therefore, the commercially available thioacetylating reagent thioacetylthioethane (TATE) was used, since this is a volatile reagent. A prototype of a gas delivery apparatus was developed and constructed (see figure 2.5) to be able to deliver the bases NEM or NMP, the thioacetylating reagent TATE and TFA in the gas phase. Simultaneously, the small capillary columns were changed for columns with a larger diameter (I.D. = 1 mm, L = 10 mm) to overcome the problem of high back pressure. Although TATE does react in the gas phase with the N-terminal amino acid it was never possible to generate a sequence tag on a peptide. After the first cycle no further reaction is observed, making this reagent less useful in this study. It was found that after using TATE in basic conditions to couple and subsequent cleavage with acid a build-up of hydrophobic salts (base - TFA) occurs. This salt is difficult to remove from the column without significant peptide losses and causing a buffering effect, resulting in lowering the coupling efficiencies.

***Figure 2.8* Comparison of the three thioacetylating reagents on C21W.** (A) The mass spectrum after two reaction cycles with TBTGA. Two cycles are observed (residues WF). (B) The mass spectrum after two reaction cycles with TATE. Only one reaction cycle is observed (residue W). (C) The mass spectrum after two reaction cycles with the synthesized TATG. Two cycles are observed (residues WF) and the spectrum looks more uncontaminated compared to the spectrum in (A). The peaks marked with an * are either products of side reactions, incomplete cleavage reactions or Na^+-ion adduct peaks.


A
2363.41
2216.23
F
W
2102.09*
2157.24*
2346.81*
2549.56
counts
mass (m/z)
B
2573.80*
W
2549.76
2363.46
C
2573.55*
2216.14
2363.40
F
W
2607.76*

Another thioacetylating reagent, thioacetylthioglycolic acid (TATG) was synthesized according to the literature (Mross and Doolittle, 1977). Although this reagent is crystalline, which makes it not possible to be used in the gas phase, it is much more hydrophilic compared to its benzoyl analogue and it shows very good reactivity towards the N-termini of both C21W and other protein digests. Figure 2.8 shows a comparison of the reactivity on C21W of the three thioacetylating reagents used in this study. In this example, where only two cycles were performed, TATG shows the best mass spectrum. The coupling reaction was carried out in the aqueous phase, the cleavage reaction in the gas phase. All the reagents and by-products are water soluble and therefore easily removed from the reversed-phase column by washing with water. This avoids peptide losses during the reaction cycles. As can be seen in figure 2.8 there are some side-product peaks present. One frequently found, was a peak with an added mass of 42 a.m.u. This is due to a S → O substitution in the thioacetyl peptide after the coupling reaction (probably due to oxidation). When this occurs the acetyl peptide cannot be converted by TFA into its thiazolinone derivative, this way blocking further generation of the sequence tag. On the other hand, when this happens frequently, the peaks with masses plus 42 a.m.u. can be helpful for mass spectrum interpretation.

The effects of reaction temperature, coupling time and the effectiveness of different volatile bases such as NEM, NMP, TMA, TEA and piperidine were tested in a systematic fashion. Changing the temperature over a range between 40 °C and 80 °C did not change the reactivity significantly, although at higher temperatures there seems to be more side reativity. Therefore, the reaction temperature was kept at 40 °C. Coupling time is rather short. Since the purpose of the reaction is to generate a ladder sequence, coupling should be about 30% efficient. It was experimentally found that a coupling time between 5 and 15 min. for TATG was optimal. Different volatile bases had no significant effect on the ladder generating chemistry. Therefore, for practical reasons the least odorous bases were selected, like NMP or NEM.

An important step in the ladder generating chemistry is the drying time after applying the sample and especially after the coupling reaction. TFA cleavage is not optimal when the system is not completely dry. Therefore, drying times of 10 - 20 min. were used in the experiments.

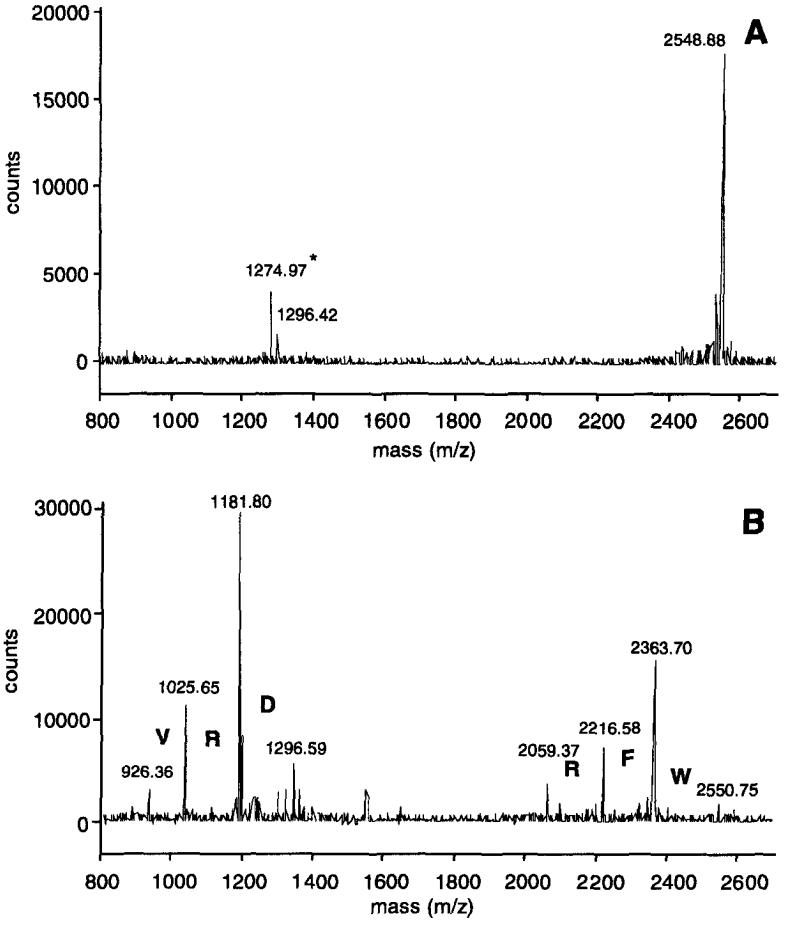


Figure 2.9 **Simultaneous ladder generation on the peptides C21W and angiotensin.** (A) Mass spectrum of C21W and angiotensin before the degradation cycles. The peak marked with an * is the double charged MH^{2+}-ion derived from C21W, giving exactly half of the mass of the single charged ion of C21W. Note that even though equal amounts of peptide were spotted on the MALDI target, intensities differ enormously, due to different ionization potentials of each individual peptide. (B) The mass spectrum after three reaction cycles with the synthesized TATG. Three cycles are observed for each peptide, resulting in N-terminal amino acid sequences WFR and DRV for C21W and angiotensin, respectively.

Since the main interest was in the generation of short sequence tags on protein digests, thus mixtures of peptides, experiments were carried out using TATG on either a mixture of the peptides C21W and angiotensin

(amino acid sequence DRVYI HPFHL, monoisotopic mass of 1295.68 a.m.u.) or on protein digests of myoglobin or calmodulin. Figure 2.9 shows the mass spectrum of the two peptides C21W and angiotensin before and after carrying out three cycles with the reagent TATG. The mass spectrum shows very clearly two sequence tags on both peptides. This is an important finding, because it is the first demonstration of the generation of short sequence tags on a mixture of peptides. In the past sequence ladders have been generated using amino- and carboxypeptidases, which works well with single peptides but is not so successful when analysing peptide mixtures. Similar results were obtained when tryptic digests from myoglobin or calmodulin were subjected to degradation with TATG (data not shown).

Figure 2.10 finally shows the generation of the sequence tag after four cycles as described in section 2.2.6. However, the theoretical distribution of each fragment (24.0% intact, 41.2%, 26.4%, 7.6% and 0.8 % respectively) after four cycles is not observed. This is partially due to the differences in ionization potential for each fragment and due to changes in kinetics after each cycle. But for the interpretation of the mass spectrum this is not so significant as long as good sequences can be read.

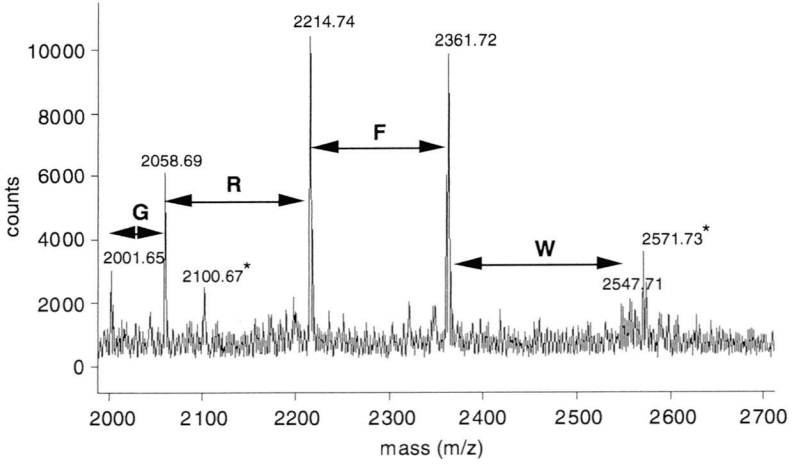


Figure 2.10 **Ladder generation on C21W as described in section 2.2.6.** The mass spectrum after four reaction cycles with the synthesized TATG. Four cycles are observed (residues WFRG). The peak marked with an * is a Na^+-ion adduct peak.

During the progress of optimizing the ladder generating chemistry other methods for immobilizing the peptides were tested. Initially, Nucleosil C_{18} and Poros 10 R2 were used to pack small columns. Both bind efficiently peptides, with Poros 10 R2 causing lower back pressure than Nucleosil C_{18}. Packing good columns is time consuming and therefore commercially available PTFE membranes with embedded C_{18} reversed-phase material were tested. Such membranes are in general designed for environmental reasons to remove organic compounds from (drinking) water, but it was found that peptides properly bind to these membranes too. Furthermore, it was shown that the ladder generating chemistry was not altered using these membranes (data not shown). In the later experiments only the C_{18} membranes were used to immobilize the peptides.

In section 2.3.2 an example of the generation of sequence tags on peptides from tryptic digests of ribosomal proteins from *E. coli* is presented to demonstrate that the method is capable for high throughput protein identification using the recently developed algorithm MassDynSearch (Korostensky et al., 1998).

2.3.2 Identification of ribosomal proteins from E. coli by partial N- terminal sequencing with TATG

In order to demonstrate that it is possible to generate short sequences on protein digests on a large scale, ribosomal proteins were isolated and purified by preparative reversed-phase HPLC. Figure 2.11 shows a typical HPLC chromatogram of such a separation. Each numbered fraction was subjected to tryptic digestion followed by the generation of a short N-terminal sequence using the thioacetylating reagent TATG. For this purpose the multiple sample loader (figure 2.2.5) was used to be able to process ten protein digests in parallel.

Independently, identification was done by MS/MS on the tryptic digests and by N-terminal Edman degradation on the intact proteins to confirm the results obtained by the ladder generating chemistry. First, each tryptic fraction corresponding to an HPLC fraction was subjected to MS/MS analysis followed by identification using SEQUEST as a control. Using this method all the proteins in each HPLC fraction could be identified (100% identification using SEQUEST). Each tryptic digest was then subjected to partial degradation using TATG. The results were used to search the database using MassDynSearch. The masses of the digests without partial

degradation were used to search the database using MassSearch. The experimentally obtained N-terminal sequence of the intact proteins were used to search the database using FASTA/TFASTA.

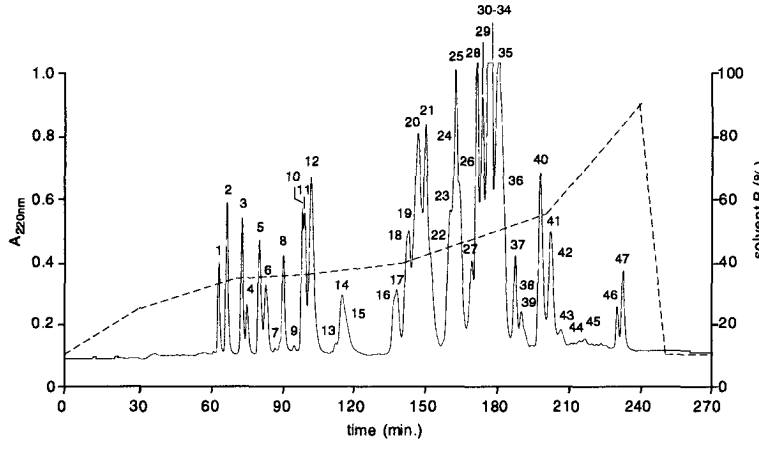


***Figure 2.11* HPLC chromatogram of ribosomal proteins from *E. coli*.** Ribosomal proteins were separated by HPLC on a preparative reversed-phase column, each individual peak numbered in the chromatogram was subjected to tryptic digestion and the obtained peptides were used for partial N-terminal sequencing. The dotted line shows the gradient of solvent B (80% acetonitrile/0.08% TFA).

Table 2.1 on the pages 74 - 89 summarizes the results obtained from all the experiments. For each HPLC fraction the protein identification obtained from SEQUEST, the calculated and experimentally determined protein mass, the N-terminal from the database and the experimentally determined N-terminal, the peptide masses from digestion by trypsin and the obtained chemical sequence tag(s) using TATG are given. Further indicated is which method (by using the programs MassSearch, MassDynSearch and FASTA/TFASTA) lead to positive identification. When using the data obtained from MassDynSearch, so to say the data obtained from the ladder sequencing, 50% of all proteins were identified, in contrast to 27% from the data from MassSearch alone. N-terminal analysis of the intact protein resulted only in the identification of 27% of the proteins, comparable with the results obtained after peptide fingerprinting, but one has to keep in mind that Edman degradation is a relative slow procedure. Each Edman degradation cycle, however, takes approximately 45 min. for a single protein at the time,

whereas, when to generate the short sequence tag using the multiple sample loader ten samples can be processed at the same time.

Since many HPLC fractions contain more than one protein the percentages of protein identification are not very high. This percentage of positive identifications using the search programs such as MassDynSearch will be more increased when applied on protein samples obtained from excised spots from 2-D gels. These spots usually contain only one single protein.

2.3.3 *Example of protein identification in fraction number 36*

For one fraction (fraction 36) the detailed protein identification process is shown. In figure 2.12 are shown the mass spectra of the peptides obtained from a tryptic digestion before (A) and after ladder sequencing (C), the mass spectrum of the intact proteins present (D) as well as an example of a MS/MS fragmentation spectrum (B).

The MS/MS spectrum in figure 2.12(B) shows the fragmentation of the peptide $m/z = 756$ from the tryptic digest. Fragmentation results in a series of b- and y-ions from which a short sequence can be read. From the MS/MS spectrum in figure 2.12(B) the sequence GLMPNPK could be assigned, resulting in a match of protein EC3874 using the program SEQUEST. MS/MS fragmentation of other peptides in this digest resulted in the identification of EC3873 and EC3259 in fraction 36 in a similar way (figure 2.13).

Using the peptide masses obtained from the digestion with trypsin (figure 2.12(A)) did not result in a unique identification with MassSearch. However, after performing partial N-terminal ladder sequencing three short sequences could be assigned. The tryptic peptide with $m/z = 1704.04$ shows a N-terminal tag FV matching EC3874, $m/z = 1512.58$ shows a N-terminal tag GLP matching EC3873 and finally $m/z = 2259.04$ shows a N-terminal tag S matching EC3254. The identification was done using a test version of the program MassDynSearch*.

* continue on page 92

Table 2.1 Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
1	EC3621, AC P02437 50S subunit L34	5380	5384	MKRTFQ	MKRTFQ (ragged)
2	EC1051, AC P02435 50S subunit L32	6446	6320 (- init M)	MAVQQN	MAVQQN
3	EC3556, AC P02436 50S subunit L33	6371	6259 (- init M, + mono-methylation)	MAKGIR	AKGIRE
4	EC3217, AC P21194 50S subunit L36	4364	4370	MKVRAS	MKVRAS
5	EC3107, AC P02427 50S subunit L27	9124	8995 (- init M)	MAHKKA	AHKKAG
6	EC3228, AC P02411 50S subunit L14	13541	13681 (conflict in sequence)	MIQEQT	Unreadable
	EC3260, AC P02367 30S subunit S12	13606	13681 (- init M)		
7	EC3834, AC P02432 50S subunit L31	7871	7872	MKKDIHPKYE	MKKDI
	EC3107, AC P02427 50S subunit L27	9124	8998 (- init M)	MAHKKA	No sequence
	EC3229, AC P02373 30S subunit S17	9704	Not detected	MTDKIR	No sequence
8	EC3834, AC P02432 50S subunit L31	7871	7874	MKKDIHPKYE	MKKDI
	EC3876, AC P02392 50S subunit L7/12	12295	Not detected	MSITKD	No sequence
9	EC2988, AC P02379 30S subunit S21	8500	8371 (- init M)	MPVIKV	Unreadable due to ragged N-terminus
10	EC3227, AC P02425 50S subunit L24	11316	11184 (- init M)	MAAKIR	No sequence
	EC3621, AC P02437 50S subunit L34	5380	Not detected	MKRTFQPSVLK	FQPSVL

Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag (MH^+ mono)[h]	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
919.5, 1075.5, 1232.5, 1474.6, 1707.5, 1942.2, 2210.9, 2239.0	919.59 **(T)**K matches EC3621		EC3621	EC3621
1041.3, 1232.3, 1244.2, 1261.0, 1277.2, 1543.3	1232.66 **(HH)**R matches EC1051		EC1051	EC1051
964.7, 1120.5, 1138.5, 1232.5, 1322.6, 1569.7, 1940.9, 2173.9, 2211.9, 2238.9	1568.46 **(LVS)**K *matches* EC3556		EC3556	EC3556
1018.7, 1138.6, 1207.6, 1234.7, 1362.7, 1404.8, 1569.7, 1689.9, 1779.8, 1940.8, 2239.1, 2296.9, 2314.1	None		Not applicable	EC3217
808.4, 831.5, 1085.5, 1386.7, 1404.7, 1435.6, 1576.7	1404.56 **(FGGE)**R matches EC3107		EC3107	EC3107
	1300.50 **(V)**R matches EC3107			
700.4, 785.3, 900.4, 936.4, 1131.3, 1321.5, 1404.6, 1492.8, 1871.7, 2121.7, 2304.5, 2983.9	1873.46 **(IG)**R matches EC3228 (chymotryptic cleavage)		EC3260	Not applicable
936.4, 1277.7, 1348.4, 1404.6, 1436.5, 1487.5, 1871.8, 1939.5, 1969.7, 1987.7	None		Not applicable	EC3834
				Not applicable
				Not applicable
936.6, 1003.5, 1150.6, 1189.8, 1348.7, 1390.7, 1436.7, 1487.7, 1583.8, 1754.9, 1805.0, 1870.0, 1970.0, 2133.9, 2200.9, 2421.2	None		Not applicable	EC3834
				Not applicable
814.5, 1189.5, 1299.6, 1403.5, 1487.3, 1940.8, 1969.7, 2210.9, 2237.9	1189.5 **(E)**R matches EC2988		EC2988	Not applicable
1083.8, 1171.7, 1189.7, 1244.8, 1542.9, 1768.0, 1804.1	1804.14 **(EAA)**K matches EC3227	EC3227	EC3227	Not applicable
				EC3621

Table 2.1 (continued) Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
11	EC3227, AC P02425 50S subunit L24	11316	11184 (- init M)	MAAKIR	No sequence
	EC2988, AC P02379 30S subunit S21	8500	8371 (- init M)	MPVIKV	No sequence
	EC1674, AC P07085 50S subunit L35	7289	7161 (- init M)	MPKIKTVRGAA	PKIKT
12	EC3225, AC P02370 30S subunit S14	11580	11449 (- init M)	MAKQSM	AKQSMK
	EC3557, AC P02428 50S subunit L28	9006	8876 (- init M)	MSRVCQ	SRVCQV
13	EC3215, AC P02366 30S subunit S11	13844	13880 (- init M, + mono-methylation)	MAKAPI	XXADR XXQPI XXV
14	EC3215, AC P02366 30S subunit S11	13844	13805 (- init M)	MAKAPI	Unreadable due to ragged N-terminus
	EC3234, AC P02375 30S subunit S19	10430	10446	MPRSLK	
15	EC3215, AC P02366 30S subunit S11	13844	13729 (- init M)	MAKAPI	No sequence
	EC189, AC P52098 yaeO; hypothetical	9698	Not detected	MSMNDT	MNDTYQ
16	EC23, AC P02378 30S subunit S20	9684	9556 (- init M)	MANIKS	ANIKSA
	EC3621, AC P02437 50S subunit L34	5380	Not detected	MKRTFQPSVLK RNR	VLKRNR
17	EC23, AC P02378 30S subunit S20	9684	9555 (- init M)	MANIKS	Unreadable
	EC4085, AC P02374 30S subunit S18	8986	8899 (- init M, + acetylation)	MARYFR	
	EC3227, AC P02425 50S subunit L24	11316	Not detected	MAAKIR	

Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag (MH^+ mono)[h]	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
1086.7, 1120.7, 1170.5, 1189.7, 1244.8, 1299.0, 1578.0, 1679.2, 1767.9, 1804.1, 1868.0, 1939.4, 1970.4, 2036.2, 2277.5, 2478.4, 2606.5	None	EC3227	EC3227	Not applicable
				Not applicable
				EC1674
924.6, 1086.7, 1120.7, 1404.9, 1539.7, 1539.8, 1580.0, 1666.7, 1703.1, 1860.2, 2036.3, 2477.3	1085.64 **(G)**R matches EC3557			EC3225
			EC3557	EC3557
1032.6, 1086.6, 1120.6, 1146.7, 1250.7, 1282.7, 1293.7, 1451.7, 1701.9, 1939.9, 2074.0, 2150.0, 2478.2	None		Not applicable	EC3215
819.4, 848.4, 868.4, 890.4, 1152.6, 1237.1, 1280.7, 1297.6, 1531.5, 1563.4, 1660.7, 1705.6, 1721.6, 1956.7, 2073.1, 2477.9, 2495.1, 2623.2	1280.60 **(KG)**K matches EC3234			Not applicable
		EC3234	EC3234	
1032.7, 1152.7, 1250.7, 1298.8, 1451.8, 1532.7, 1577.8, 1705.9, 1941.1, 2074.1, 2479.3, 2623.5	2478.04 **(pEVS)**R matches EC3215		EC3215	Not applicable
				EC189
1036.4, 1081.4, 1093.5, 1164.4, 1240.5, 1274.3, 1319.4, 1335.4, 1398.5, 1669.6, 1704.4, 1940.5, 2238.7, 2408.4, 2821.9	1319.81 **(AF)**R matches EC23		EC23	EC23
				EC3621
1036.7, 1096.7, 1164.7, 1240.8, 1285.8, 1319.8, 1399.8, 1442.6, 1669.9, 1697.1	1240.67 **(YLSL)**R matches EC4085			Not applicable
			EC4085	

Table 2.1 (continued) Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
18	EC2134, AC P02426 50S subunit L25	10693	10694	MFTINAE	Unreadable
	EC3089, AC P02371 30S subunit S15	10268	10139 (- init M)	MSLSTE	
19	EC3235 AC P02387 50S subunit L2	29860	29712 (- init M)	MAVVKC	Unreadable
	EC3089, AC P02371 30S subunit S15	10268	10137 (- init M)	MSLSTE	
20	EC2542, AC P02420 50S subunit L19	13133	12999 (- init M)	MSNIIK	Unreadable
	EC3235, AC P02387 50S subunit L2	29860	Not detected	MAVVKC	
	EC3228, AC P02411 50S subunit L14	13541	13537	MIQEQT	
21	EC3235, AC P02387 50S subunit L2	29860	29678 (- init M)	MAVVKC	Unreadable
	EC3221, AC P02356 30S subunit S5	17603	Not detected	MAHIEK	
	EC3220, AC P02430 50S subunit L30	6541	6412 (- init M)	MAKTIK	
22	EC3220, AC P02430 50S subunit L30	6541	6413 (- init M)	MAKTIK	AKTIKI
	EC3235, AC P02387 50S subunit L2	29860	29707 (- init M)	MAVVKC	No sequence
23	EC3238, AC P02386 50S subunit L3	22243	22245	MIGLVG	MIGLVG
	EC3229, AC P02373 30S subunit S17	9704	9574 (- init M)	MTDKIR	TDKIRT
24	EC3154, AC P02410 50S subunit L13	16018	16017	MKTFTA	MKTFTA
	EC3238, AC P02386 50S subunit L3	22243	22252	MIGLVG	No sequence

Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag[h] (MH^+ mono)	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
922.6, 965.7, 1027.7, 1080.7, 1096.7, 1240.8, 1350.8, 1670.0, 1692.0, 2632.5, 3244.7	1080.7 **(M)**R matches EC2134	EC2134	EC2134	Not applicable
	1027.7 **(LQ)**R matches EC2134			
922.6, 965.6, 1027.6, 1080.6, 1096.6, 1164.7, 1669.9, 2632.3, 3244.6	2632 **(DAND)**H matches EC3089 (chymotryptic cleavage)			Not applicable
	922.34 **(Y)**R matches EC3089		EC3089	
1091.8, 1164.8, 1316.0, 1497.7, 1878.1, 2163.1, 2753.6, 3536.8, 4324.0	1315.99 **(LQA)**R matches EC2542		EC2542	Not applicable
		EC3235	EC3235	
894.5, 913.6, 955.6, 1091.7, 1164.7, 1249.7, 1315.8, 1497.6, 1688.8	1202.80 **(H)**R matches EC3220 (- Na^+)	EC3235	EC3235	Not applicable
			EC3220	
913.7, 955.6, 1091.7, 1164.8, 1249.8, 1315.9, 1497.7, 1688.9, 1808.3, 2188.3, 2238.3	913 **(AT)**R matches EC3220		EC3220	EC3220
		EC3235	EC3235	
886.6, 993.5, 1115.7, 1207.7, 2147.9, 2188.2, 2368.2	2187.72 **(IFT)**R matches EC3238	EC3238	EC3238	EC3238
				EC3229
771.5, 807.4, 839.4, 886.5, 993.5, 1115.6, 1153.5, 1207.6, 1315.8, 1481.6, 1535.6, 1805.9, 1919.8, 2103.9, 2188.0, 2233.9, 2368.0	2187.55 **(IFT)**R matches EC3238	EC3154	EC3154	EC3154
			EC3238	Not applicable

Table 2.1 (continued) Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
25	EC3154, AC P02410 50S subunit L13	16018	16019	MKTFTA	MKTFTA
	EC3212, AC P02416 50S subunit L17	14364	14365	MRHRKS	No sequence
26	EC3212, AC P02416 50S subunit L17	14364	14363	MRHRKS	MRHRKS
27	EC3222, AC P02419 50S subunit L18	12769	12769	MDKKSA	MDKKSA
28	EC3214, AC P02354 30S subunit S4	23469	23240 (- init M, conflict in sequence)	MARYLG	Unreadable
	EC3233, AC P02423 50S subunit L22	12226	12226	METIAK	
	EC3232, AC P02352 30S subunit S3	25983	25828 (- init M)	MGQKVH	
	EC3222, AC P02419 50S subunit L18	12769	12767	MDKKSA	
29	EC3216, AC P02369 30S subunit S13	13099	12965 (- init M)	MARIAGI	Unreadable
	EC3239, AC P02364 30S subunit S10	11735	11733	MQNQRI	
	EC3233, AC P02423 50S subunit L22	12226	12224	METIAK	

Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag (MH^+ mono)[h]	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
945.7, 978.7, 1013.6, 1121.7, 1153.7, 1185.7, 1315.9, 1481.8, 1535.8, 1589.1, 1634.9, 1663.9, 1806.1, 1920.2, 2104.3, 2188.4, 2234.3, 2256.3, 2368.4	2233.40 **(AE)**K matches EC3154	EC3154	EC3154	EC3154
	2104 **(VY)**I matches EC3154 (C-term. peptide)	EC3212	EC3212	Not applicable
945.3, 978.4, 1017.4, 1588.8, 1634.4, 1650.4, 2102.6, 2187.7, 2233.5, 2367.6	1634.43 **(AGDN)**R matches EC3212	EC3212	EC3212	EC3212
887.4, 914.5, 951.4, 978.5, 1094.3, 1123.6, 1160.6, 1213.5, 1289.6, 1650.6, 1781.8, 1834.8, 1934.5, 2187.9, 2237.8, 2396.9, 2425.9, 2606.0	951.51 **(SG)**R matches EC3222	EC3222	EC3222	EC3222
951.4, 1000.5, 1050.5, 1163.5, 1213.6, 1331.7, 1456.6, 1631.8, 1667.7, 1830.7, 1875.7, 1959.8, 2045.8, 2153.8, 2185.9, 2396.9, 2578.9	1456.6 **(GNT)**R matches EC3214	EC3214	EC3214	Not applicable
934.5, 1137.3, 1213.6, 1239.6, 1255.6, 1370.6, 1428.7, 1460.6, 1716.8, 2209.8, 2578.9	1213.78 **(TS)**R matches EC3233 (C-term. peptide)			Not applicable
		EC3233	EC3233	

Table 2.1 (continued) Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
30	EC3233, AC P02423 50S subunit L22	12226	12223	METIAK	Unreadable
	EC3216, AC P02369 30S subunit S13	13099	12967 (- init M)	MARIAGI	
	EC3236, AC P02424 50S subunit L23	11199	11197	MIREER	
	EC3231, AC P02414 50S subunit L16	15281	15312 (mono-methylation)	MLQPKR	
	EC3224, AC P02361 30S subunit S8	14126	13990 (- init M)	MSMQDP	
	EC877, AC P02349 30S subunit S1	61158	61508	MTESFA	
	EC2988, AC P02379 30S subunit S21	8500	Not detected	MPVIKV	
31	EC3233, AC P02423 50S subunit L22	12226	Not detected	METIAK	Unreadable
	EC2545, AC P02372 30S subunit S16	9190	9187	MVTIRL	
	EC3236, AC P02424 50S subunit L23	11199	11193	MIREER	
	EC2988, AC P02379 30S subunit S21	8500	Not detected	MPVIKV	
	EC3224, AC P02361 30S subunit S8	14126	13988 (- init M)	MSMQDP	
32	EC3233, AC P02423 50S subunit L22	12226	Not detected	METIAK	Unreadable
	EC3230, AC P02429 50S subunit L29	7273	7276	MKAKEL	
	EC3231, AC P02414 50S subunit L16	15281	15319 (mono-methylation)	MLQPKR	
	EC1673, AC P02421 50S subunit L20	13496	Not detected	MARVKR	
	EC3234, AC P02375 30S subunit S19	10430	Not detected	MPRSLK	

Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag (MH^+ mono)[h]	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
934.5, 1036.5, 1142.6, 1213.6, 1241.7, 1280.7, 1332.7, 1348.7, 1377.6, 1422.7, 1460.7, 1542.6, 1570.6, 1620.7, 1716.9, 1785.9, 1926.1, 2073.9, 2178.9, 2260.9	1521.71 **(L)**Y matches EC2988 (C-term. peptide)		EC2988	Not applicable
1004.5, 1036.5, 1152.6, 1169.6, 1213.6, 1228.6, 1285.5, 1438.5, 1573.6, 1620.7, 1716.8, 1785.8, 2177.9, 2685.0, 2782.3, 3536.9, 3699.4, 3776.3	1521.71 **(LA)**Y matches EC2988 (C-term. peptide)	EC3233	EC3233 EC2988	Not applicable
806.4, 851.4, 968.4, 1004.5, 1033.5, 1049.5, 1152.5, 1169.5, 1213.5, 1228.5, 1268.5, 1438.4, 1470.4, 1605.6, 1642.7, 1716.7	1685.02 **(KVE)**R matches EC3234	EC3233	EC3233 EC3234	Not applicable

Table 2.1 (continued) Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
33	EC3230, AC P02429 50S subunit L29	7273	7276	MKAKEL	MKAKEL
	EC3219, AC P02413 50S subunit L15	14980	14974	MRLNTL	No sequence
34	EC3219, AC P02413 50S subunit L15	14980	14974	MRLNTL	Unreadable
	EC4083, AC P02358 30S subunit S6	15703	15763	MRHYEI	
	EC3153, AC P02363 30S subunit S9	14856	14974 (- init M, conflict in sequence)	MAENQY	
	EC3223, AC P02390 50S subunit L6	18903	Not detected	MSRVAK	
	EC3221, AC P02356 30S subunit S5	17603	17507 (- init M, + acetylation)	MAHIEK	
35	EC3874, AC P02384 50S subunit L1	24729	24575 (- init M)	MAKLTK	Unreadable
	EC3873, AC P02409 50S subunit L11	14875	14855 (- init M, + methylation)	MAKKVQ	
	EC3221, AC P02356 30S subunit S5	17603	17486 (- init M, + acetylation)	MAHIEK	
36	EC3874, AC P02384 50S subunit L1	24729	24586 (- init M)	MAKLTK	Unreadable
	EC3873, AC P02409 50S subunit L11	14875	14867 (- init M, + methylation)	MAKKVQ	
	EC3259, AC P02359 30S subunit S7	20019	19880 (- init M)	MPRRRV	
37	EC3226, AC P02389 50S subunit L5	20301	20162 (- init M)	MAKLHD	AKLHDY

Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag (MH^+ mono)[h]	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
1004.7, 1013.8, 1033.7, 1131.8, 1152.8, 1169.8, 1213.8, 1228.8, 1253.5, 1283.8, 1387.9, 1455.0, 1470.7, 1512.9, 1574.0, 1621.1, 1643.2, 1717.2, 1833.1, 1990.2, 2032.2	1643.21 **(SV)**R matches EC3230		EC3230	EC3230
		EC3219	EC3219	
964.4, 999.4, 1013.5, 1043.4, 1131.5, 1158.3, 1226.5, 1254.4, 1283.5, 1299.5, 1387.6, 1454.6, 1512.4, 1642.6, 1703.6, 1751.6, 1892.6, 1989.6, 2120.5, 2332.8, 2495.7, 2640.7, 2703.8	1989 **(LNT)**R matches EC3219	EC3219	EC3219	Not applicable
		EC3153	EC3153	
844.5, 964.6, 1131.6, 1222.7, 1387.7, 1454.7, 1512.6, 1654.7, 1703.8, 1751.7, 1879.8, 2704.0	1704.04 **(FV)**R matches EC3874	EC3874	EC3874	Not applicable
844.5, 1029.2, 1130.7, 1170.7, 1267.7, 1384.5, 1512.6, 1605.8, 1703.8, 1751.8, 1838.9, 1879.9, 2259.1	1704.04 **(FV)**R matches EC3874		EC3874	Not applicable
	1512.58 **(GLP)**R matches EC3873		EC3873	
	2259.06 **(S)**K matches EC3259		EC3259	
882.0, 1030.6, 1043.9, 1091.9, 1107.9, 1133.1, 1267.2, 1476.3, 1510.2, 1693.4, 1828.5, 2099.8	1267.66 **(ALLA)**R matches EC3226	EC3226	EC3226	EC3226

Table 2.1 (continued) Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
38	EC169, AC P02351 30S subunit S2	26743	Not detected	MATVSM	ATVSM
	EC2516, AC P21507 RNA helicase	49914	Not detected	MTVTTFSE	
	EC3219, AC P02413 50S subunit L15	14980	Not detected	MRLNTL	
	EC2815, AC P07012 Peptide chain release factor 2	41250	Not detected	MFEINP	
39	EC169, AC P02351 30S subunit S2	26743	Not detected	MATVSM	ATVSM
	EC2546, AC P07019 GTP binding export factor	49787	Not detected	MFEINPVN	
40	EC1673, AC P02421 50S subunit L20	13496	13359 (- init M)	MARVKR	ARVKRG
	EC169, AC P02351 30S subunit S2	26743	26712	MATVSM	No sequence
41	EC3875, AC P02408 50S subunit L10	17711	17569 (- init M)	MALNLQ	Unreadable
	EC3237, AC P02388 50S subunit L4	22086	22079	MELVLK	
42	EC3875, AC P02408 50S subunit L10	17711	17571 (- init M)	MALNLQ	Unreadable
	EC3237, AC P02388 50S subunit L4	22086	22061	MELVLK	
43	EC3875, AC P02408 50S subunit L10	17711	17574 (- init M)	MALNLQ	Unreadable
	EC2215, AC unassigned	11305	11171	MSNQFG	
44	EC3875, AC P02408 50S subunit L10	17711	17574 (- init M)	MALNLQ	Unreadable

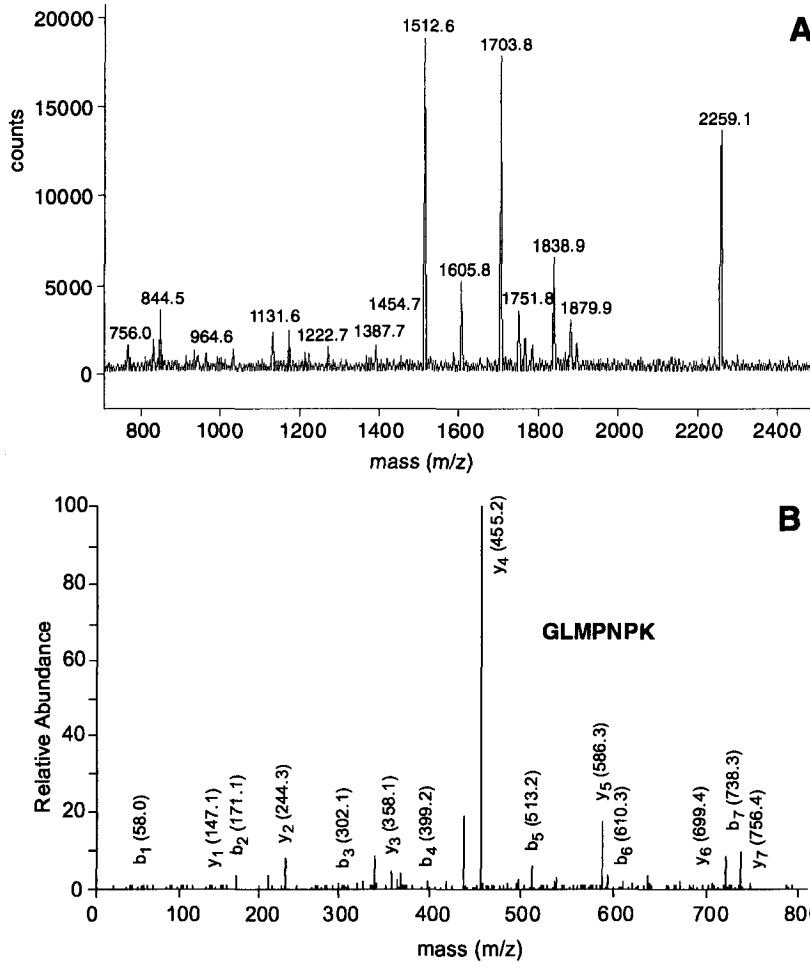
Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag (MH^+ mono)[h]	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
777.3, 857.3, 882.4, 965.4, 1267.4, 1404.5, 1576.3, 1622.4, 1889.3, 2392.4, 2514.5, 2529.3	None	EC169	Not applicable	EC169
777.5, 857.6, 882.6, 1036.6, 1267.7, 1285.6, 1603.8, 1703.9, 2393.1, 2577.1	None		Not applicable	EC169
977.6, 1020.6, 1104.7, 1267.8, 1421.9, 1549.8, 1670.1, 1704.1	1104 **(IL)K** matches EC1673		EC1673	EC1673
				Not applicable
1020.7, 1036.6, 1045.7, 1192.9, 1226.8, 1243.8, 1267.8, 1447.8, 1603.9, 1687.9, 1971.3, 2030.2, 2215.1, 2384.2, 2516.5, 2561.4	1616.90 **(AA)**R matches EC3875 (chymotryptic cleavage)	EC3875	EC3875	Not applicable
	1446.72 **(LAT)**R matches EC3875	EC3237	EC3237	
916.7, 1020.7, 1045.7, 1193.7, 1226.8, 1417.8, 1447.8, 1603.9, 1688.0, 2216.1, 2384.3, 2561.4, 2577.4	2561.66 **(DAF)**K matches EC3875	EC3875	EC3875	Not applicable
	1688 **(AAA)**R matches EC3875	EC3237	EC3237	
913.7, 1022.7, 1036.7, 1052.7, 1125.9, 1163.8, 1267.9, 1417.9, 1447.9, 1604.1, 1670.2, 1704.2, 1704.2, 1965.3, 1974.3, 2278.6, 2310.6, 2465.5, 2561.7, 2577.6, 3017.9, 3049.9	None	EC3875	Not applicable	Not applicable
1133.6, 1704.1, 1757.9, 1835.1, 1973.3, 2212.2, 2238.2, 2379.3	None		Not applicable	Not applicable

Table 2.1 (continued) Identification of ribosomal proteins.

Fraction number[a]	Protein ID[b]	Calc. mass (av.)[c]	Exp. mass (av.)[d]	N-term DB[e]	N-term (exp)[f]
45	EC3876, AC P02392 50S subunit L7/L12	12295	12297 (- init M)	MSITKD	SITKD
46	EC3876, AC P02392 50S subunit L7/L12	12295	12158 (- init M)	MSITKD	MSITKD
47	EC3876, AC P02392 50S subunit L7/L12	12295	12199 (- init M)	MSITKD	MSITKD

Masses tryptic digest (MH^+ mono)[g]	N-terminal chemical tag (MH^+ mono)[h]	Protein identified by MassSearch[i]	Protein identified by MassDynSearch[j]	Protein identified by FASTA[k]
1025.6, 1070.8, 1115.8, 1277.9, 1390.8, 1603.9, 1681.9, 1704.1, 1831.1, 1938.3, 2112.3, 2211.4, 2240.4, 2379.3	None		Not applicable	EC3876
964.6, 1025.6, 1113.7, 1244.7, 1267.7, 1482.8, 1634.0, 1679.9, 1695.9, 1826.9, 1842.8, 2015.1, 2037.1, 2721.3, 2769.4	2015.23 **(FG)**K matches EC3876 964 **(T)**K matches EC3876	EC3876	EC3876	EC3876
964.6, 986.6, 1025.6, 1200.7, 1244.7, 1263.0, 1482.7, 1633.9, 1679.9, 1695.9, 1733.8, 1810.9, 2015.1, 2037.1	2014.93 **(FGV)**K matches EC3876 1244.55 **(A)**K matches EC3876		EC3876	EC3876

[a] The fraction number corresponding to the peaks of the HPLC chromatogram shown in figure 2.11.
[b] Protein ID shows the *Escherichia coli* identification number (EC), the SwissProt accession number (AC) and a description of the corresponding identified protein. As control, all proteins were identified by MS/MS using the SEQUEST matches of two peptides, this was then confirmed by manual interpretation.
[c] The calculated average intact protein mass (in a.m.u.) from the database.
[d] The experimentally determined average intact protein mass, given as average MH^+ (in a.m.u.). The protein samples were measured by MALDI-MS in linear mode with α-CHC as matrix using an accelerating voltage of 25 kV, a pulse delay time of 150 ns, a grid voltage of 91% and a guide wire voltage of 0.1%. In the case where (- init M) is mentioned this means that the initially synthesized methionine residue is cleaved off, also it is stated whether proteins are post-translationally modified by methylation or acetylation.
[e] Representation of the N-termini of the protein according to the database.
[f] The experimentally determined N-termini of the protein (by Edman degradation). When no sequence was found or the sequence was unreadable, this is stated.
[g] The experimentally obtained monoisotopic masses of the obtained peptides after digestion with the protease trypsin. The peptides were measured by MALDI-MS in delayed extraction reflector mode with α-CHC as matrix using an accelerating voltage of 20 kV, a pulse delay time of 75 ns, a grid voltage of 55% and a guide wire voltage of 0.05%.
[h] Representation of the N-terminal chemical tags (printed in bold) generated by partial N-terminal sequencing through thioacetylation with TATG. The C-terminal amino acid residue of the tryptic peptide is shown as well. When no ladder sequence could be generated this is stated.
[i] Protein identification from the data of the tryptic peptide masses by searching SwissProt and trEMBL using the program MassSearch (James et al., 1993). When there is no entry the score was below the best 20 of the output and therefore not significant.
[j] Protein identification from the data of the N-terminal chemical tag by by searching SwissProt and trEMBL using the program MassDynSearch (Korostensky et al., 1998). Note that this includes the MassSearch masses too.
[k] Protein identification from the data obtained from protein N-terminal degradation according to Edman by searching SwissProt and trEMBL using the program FASTA/TFASTA (Pearson and Lipman, 1988).


A
counts
mass (m/z)
1512.6
1703.8
2259.1
1838.9
1605.8
1751.8
1879.9
1454.7
1131.6
844.5
756.0
964.6
1222.7
1387.7
B
Relative Abundance
GLMPNPK
y4 (455.2)
y5 (586.3)
b7 (738.3)
y2 (244.3)
b3 (302.1)
y3 (358.1)
b4 (399.2)
b5 (513.2)
b6 (610.3)
y6 (699.4)
y7 (756.4)
y1 (147.1)
b2 (171.1)
b1 (58.0)

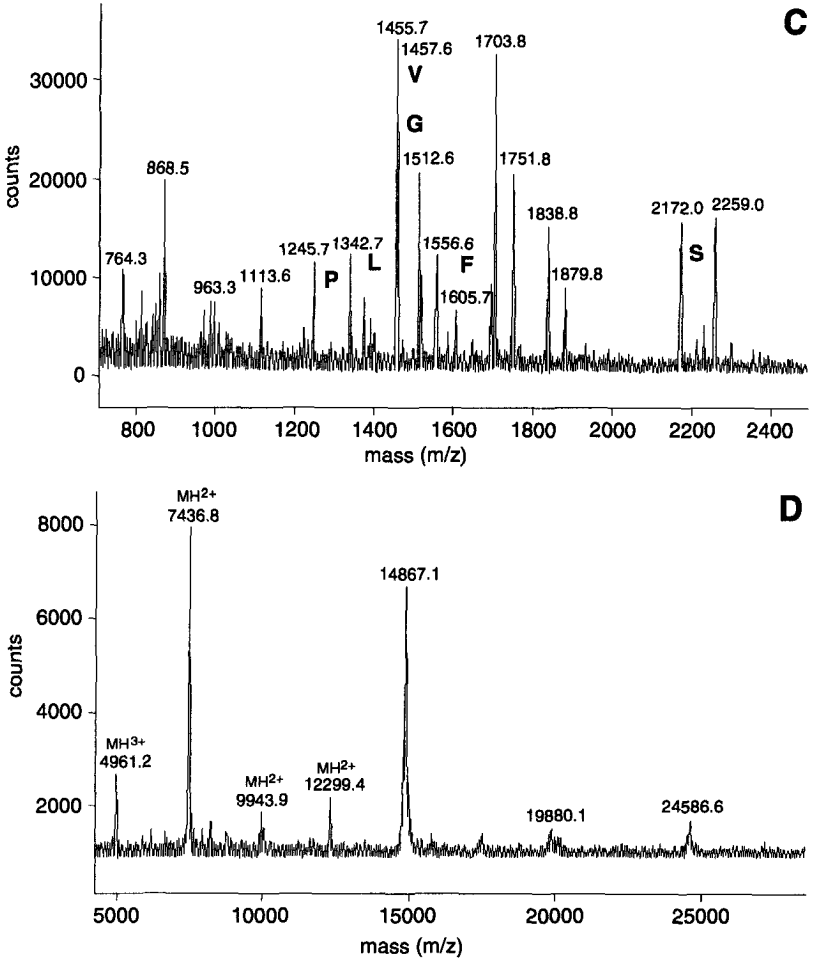


***Figure 2.12* Example of protein identification by MS.** The figure shows the mass spectra of the ribosomal proteins isolated from *E. coli* in fraction 36 from the HPLC. (A) Mass spectrum of the peptides obtained after digestion with trypsin. (B) MS/MS fragmentation spectrum from peptide m/z = 756.0 from the tryptic digest. The series of b- and y-ions formed after fragmentation is indicated, resulting in the sequence GLMPNPK. (C) as (A) but after partial N-terminal ladder sequencing. Tryptic peptide m/z = 2259.06 gives the tag S, *m/z = 1704.04 gives the tag FV and m/z = 1512.58 gives the tag GLP.* (D) Mass spectrum of the intact proteins present. For detailed explanation see the text and table 2.1.

The experimentally determined masses for the intact proteins match very well with the calculated masses, however in these three proteins the initially synthesized methionine residue is cleaved off, a process that is seen very often in prokaryotes. In this fraction N-terminal Edman sequencing on the intact proteins did not result in readable sequences, probably due to the complexity of the protein mixture. In other fractions however, in some cases it is possible to read multiple sequences after Edman degradation. This example shows that it is possible with the novel methods to identify three proteins in a single fraction.

```
SUMMARY2HTML v.0 (for SEQUEST22 output), Copyright 1996
Molecular Biotechnology, Univ. of Washington, J.Eng/J.Yates
Licensed to Finnigan MAT
09/23/98, 04:55 PM, C:\LCQ\database\Ecoli.txt, MONO

  #  File*              MassI   MassA  Xcorr*  DelCn*   Sp    RSp   Ions    Reference*   Sequence**
---  -----------------  ------  ------  ------  -----  ------  ---  ------  ----------  ---------------
  1  rib36.0020.0025.1   756.0   756.4  1.5091  0.456  195.4    1   7/ 12  ECOLI_3874  (R)GLMPNPK
  2  rib36.0043.0048.1   775.0   775.3  1.0002  0.131  366.6    1   8/ 12  ECOLI_3770  (K)NMNPDGK
  3  rib36.0068.0072.2  1393.0  1392.8  1.4254  0.036  100.6   17   7/ 24  ECOLI_691   (K)SDALLVALPPIQ
  4  rib36.0095.0099.1   676.0   675.4  1.3370  0.242   70.8    4   5/ 10  ECOLI_3259  (R)NALAMR
  5  rib36.0120.0124.1  1013.0  1013.6  0.9452  0.077  400.8    1   8/ 14  ECOLI_530   (K)KWLRLAAR
  6  rib36.0144.0149.1  1274.0  1274.6  2.4954  0.466  524.6    1  14/ 22  ECOLI_3259  (R)LANELSDAAENK
  7  rib36.0164.0168.1  1513.0  1512.9  2.1540  0.473  283.7    1  13/ 26  ECOLI_3873  (K)DLPIPVVITVYA

  1. ECOLI_3259          20 (2,0,0,0,0) { 4 6 }
  2. ECOLI_3874          10 (1,0,0,0,0) { 1 }
  3. ECOLI_530           10 (1,0,0,0,0) { 5 }
  4. ECOLI_3770          10 (1,0,0,0,0) { 2 }
  5. ECOLI_3873          10 (1,0,0,0,0) { 7 }
  6. ECOLI_691           10 (1,0,0,0,0) { 3 }
  7. ECOLI_1569           8 (0,1,0,0,0) { }
  8. ECOLI_2733           8 (0,1,0,0,0) { }
  9. ECOLI_1799           8 (0,1,0,0,0) { }
 10. ECOLI_735            8 (0,1,0,0,0) { }
```

***Figure 2.13* Output of SEQUEST.** The figure shows a list of some of the generated MS/MS spectra of the tryptic digest of fraction 36 with the assigned sequence. A second list shows the top score of proteins that were identified.

2.4 Conclusions

In order to make proteome analysis feasible, high throughput methods for the identification of proteins have to be developed. A novel method for the rapid identification of proteins has been described. After digestion of a protein with a specific protease a short ladder sequence is generated through thioacetylation under basic conditions of the N-termini of the peptides, while immobilized on reversed-phase material. This minimizes sample handling steps and thus sample losses. Subsequent cleavage of the N-termini of the thioacetylated peptides is achieved by hydrolysis with acid. The method works very well on pure peptides and on mixtures of peptides. The method was applied to identify a set of isolated and HPLC-purified ribosomal proteins from *E. coli*. All the protein fractions from the HPLC were subjected to digestion with the protease trypsin. Using peptide fragment fingerprinting data obtained by MS/MS and the program SEQUEST all the proteins in each HPLC fraction could be identified (100%). By the novel partial N-terminal tag sequence method, 50% of the proteins could be identified (MassDynSearch), compared to only 27% when the sequence tag was not present (MassSearch). It is a relative fast and easy method for protein identification. A single cycle takes approximately 40 min. like for the Edman degradation chemistry, but a multi sample loader was developed, which allows to process ten samples at the same time. In the near future this could be extended to fifty or even hundred samples to be processed in parallel. Since the chemistry is repetitive the method is potentially useful to full automation. Another strong argument in favour of this developed method is the high sensitivity. With the novel MS techniques sensitivities in the low femtomole and even attomole range can be reached.

As a conclusion the scheme outlined in figure 1.14 is a good strategy for proteome analysis. Large amounts of protein spots can be cut out from the 2-D gels with the aid of a robot, concentrated and finally digested by a specific protease. One part of the protein digest is then subjected to direct MALDI-MS measurement, whereas the other part is subjected to partially N-terminal sequencing before MALDI-MS measurement. Database searching can then be performed using both MassSearch and MassDynSearch. All these steps can be performed in an

automated fashion. This minimizes sample handling and thus protein losses and allows the identification of large numbers of proteins in a relative short time, making proteome analysis feasible.

3. Part B: 2-D Gel Analysis of Oocytes from Starfish

This section describes a study on the effect of the natural hormone 1-methyladenine (1-MA) on oocytes from two different species, *Asterina pectinifera* and *Astropecten aurantiacus* at the protein level. Two-dimensional gels of control oocytes and oocytes treated with the hormone 1-MA were compared and proteins that were expressed differently were analysed, using the techniques described before. It is an example of a subtractive proteome analysis, only the protein spots that change (in this case upon adding the hormone 1-MA) are analysed. The powerful tools of proteomics, described in the previous sections were applied in this study.

3.1 Introduction

3.1.1 Regulation of the cell cycle

The regulation of the cell division cycle is a complex process which involves kinase cascades, protease action, production of second messengers and many other operations. The cell division cycle of eukaryotic cells is conventionally divided into four stages, G_1, S, G_2 and M, where M indicates both mitosis, which can be further divided into four steps (prophase, metaphase, anaphase and telophase) and cytokinesis, which is the process of cell division (figure 3.1). DNA is replicated during the S phase, whereas histones and other proteins are synthesized during the G_1 phase. Synthesis of proteins continues in the G_2 phase, increasing the cell mass significantly. At the end of the G_2 phase cells enter the M phase, where they undergo mitosis or meiosis. The duration of the S, G_2 and M phases is relatively constant, whereas that of the G_1 phase is variable. A point within G_1 defines a time at which cells become obligatory committed to proceed to the S phase. This point is referred to as the restriction point R. Before the restriction point however, alternatives to the progression through the cell cycle are possible, for example differentiation or death. A fifth phase, termed G_0, defines a quiescent state in which cells may enter after having reached G_1. During this phase cells stop growing and their protein synthesis is decreased markedly (Alberts et al., 1994).

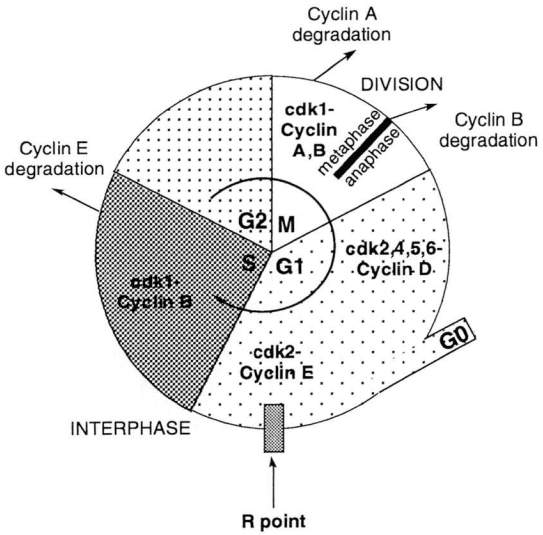


***Figure 3.1* The cell cycle.** A schematic representation of the cell cycle with the points of actions of the cdk-cyclin complexes and the positions of degradation of some cyclins. S is the period of DNA replication, G_1 and G_2 are gaps separating successive periods of DNA replication, M is mitosis, the R point is a restriction point within G_1, which defines a time when the cells become irreversibly committed to divide, G_0 is a quiescent state, in which the cells stop growing and depress protein synthesis. The times of S, G_2 and M are relatively constant whereas the time of the G_1 phase is variable (adapted from Santella et al., 1998).

The crucial features in the cell cycle are the existence of two transition control points, one at the G_2/M boundary and one during the G_1 phase. The M phase is characterized by the activation of a kinase, which is described by various names, depending on whether its activity has been assayed by stimulation of *Xenopus* oocytes to undergo meiosis (MPF) or by phosphorylation of histone H1 (H1 kinase). Originally named maturation-promoting factor (MPF) for its ability to release *Xenopus* oocytes from meiotic block, MPF is now considered to stand for M phase promoting factor. One of the targets for phosphorylation is histone H1 and the activity described as H1 kinase is now taken to be equivalent, or at least broadly overlapping with MPF (Nurse, 1990). MPF consists of two cytoplasmic components, with molecular weights of approximately 34 kDa and 45 kDa. The protein coded by the cell-division-cycle (cdc) gene of the fission yeast *Schizosaccharomyces pombe* codes for a homologue to the *Xenopus* p34 subunit. These components were initially purified by Lohka et al. (1988) and

represent two different types of subunits. The first component is the protein kinase catalytic subunit $p34^{cdc2}$ (Dunphy et al., 1988; Gautier et al., 1988) and the second component is the regulatory cyclin B protein (Draetta et al., 1989). Comparative sequence analysis of $p34^{cdc2}$ shows an extensive degree of homology to known serine/threonine protein kinases. In particular the highly conserved ATP-binding site of previously characterized protein kinases is highly conserved (Hindley and Phear, 1984). Nowadays it is well accepted that the main regulators in the initiation and progression of the cell cycle are these cyclin dependent protein kinases (cdk's). The various cyclins are synthesized and degraded during specific phases of the cell cycle, thereby controlling the activation and inactivation of MPF (Evans et al., 1983). The degradation of the cyclins is as important for exit from mitosis as its synthesis is for entry. The cyclins are degraded by proteolysis at the metaphase-anaphase transition. This process requires a signal sequence in the cyclin polypeptide chain that targets it for degradation by providing a site for attachment of ubiquitin. Thus, the association of the kinase subunit with the cyclin subunit is essential to activate the MPF. However, the mechanism leading to the activation of the cdk's is not completely understood yet. The activity of the complex might be modulated by phosphorylation performed by different (tyrosine) kinases (Lewin, 1990; Norbury and Nurse, 1992). The time taken to reactivate MPF after exit from mitosis is just long enough to allow one round of DNA replication.

3.1.2 *Proteolysis of cyclins in the control of the cell cycle*

The amount of cdk's remains in general constant, whereas the cyclins are synthesized and hydrolysed by proteases during specific phases of the cell cycle (King et al., 1996). It is clear nowadays that proteases play an important role in the regulation of the cell cycle. For example, the ubiquitin-targeted destruction of a number of cdk inhibitors, the breakdown of the nuclear envelope at the mitosis stage, the disassembly of the microtubules of the mitotic spindle, the degradation of actin binding proteins and the intermediate filament components and the segregation of chromosomes, which is initiated by the degradation of an inhibitor.

Cyclins are degraded in the proteasomes (large multi-subunit protease complexes, that selectively degrade proteins) following ubiquitin targeting. All cyclins, up to now eight types are known, share a domain of sequence homology called the cyclin box (see figure 3.2), which is the part of the

molecule responsible for the interaction with the cdk (Hunt, 1991). The mitotic cyclins A and B, which control progression through mitosis and meiosis, also contain a conserved destruction box that is assumed to serve as a signal for ubiquitin. However, the box is not sufficient as a signal for ubiquitin-dependent proteolysis, since it does not contain a conserved lysine to accept ubiquitin (King et al., 1996).

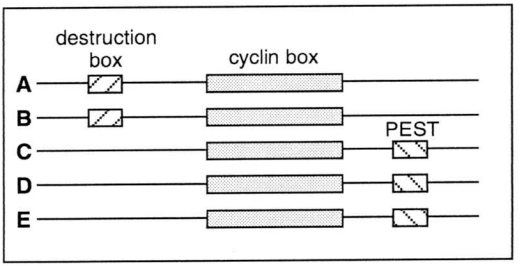


Figure 3.2 **Domain organization of some cyclins.** The cyclins whose role in the cell cycle has been established are shown. The percentage similarity in the cyclin box varies between 40 - 60%. The destruction box is a conserved nine residue sequence loacted 40 - 50 amino acids from the N-terminus. The PEST domain is typically found in short-lived proteins and could indicate proteolysis by calpain (adapted from Santella et al., 1998).

The destruction box is absent from the G_1 cyclins C, D and E. These G_1 cyclins contain a PEST domain. This is a domain rich in proline, glutamic acid, aspartic acid and serine, which is typically found in short-lived proteins. Although PEST sequences are not absolute indicators of calpain proteolysis, they are routinely present in calpain substrates (figure 3.2), and have been suggested to mediate the attack by this Ca^{2+}-dependent protease. Cyclin A and B become abruptly degraded at the end of mitosis, cyclin E in mid-to-late S-phase. The D-cyclins appear during G_1, shortly before the point at which the cell becomes committed to the S-phase (figure 3.1). They remain elevated through the remainder of the cycle, but in the absence of growth factors they disappear rapidly (Pines and Hunter, 1995).

3.1.3 The role of calcium in the regulation of the cell cycle

Second messengers have been suggested to play a role in triggering the protein kinase cascades involved in the cell cycle, that could result in their activation. Among the second messengers calcium ions and its binding protein calmodulin (CaM) appear to be of special importance

(Santella, 1998). The resumption and progression of the cell cycle, both mitotic and meiotic, are accompanied by transient increases in cytosolic Ca^{2+}. They have been observed at the G_1/S transition, during the S-phase and at the exit from mitosis. These Ca^{2+}-transients may lead to the activation of the ubiquitous Ca^{2+}-binding protein CaM. Activation of CaM leads to the stimulation of protein kinases and of the CaM-dependent protein phosphatase calcineurin. For example, one of the kinases that activate the cdk/cyclin complex is controlled by calmodulin (CaM-kinase II) (Baitinger et al., 1990). The G_1/S transition, the progression from G_2 to M and the metaphase/ anaphase transition are specific points of intervention of CaM-kinase II. As mentioned in the previous section, another possibility for the role of Ca^{2+} in the cycle is through the activation of the Ca^{2+}-dependent protease calpain (see section 3.1.5).

3.1.4 Calcium signalling in the cell nucleus

Calcium signalling is one of the most intensively studied areas of cell biology. An essential prerequisite for a second messenger, such as Ca^{2+} is that its intracellular concentration can be rapidly modified. The concentration of intracellular Ca^{2+} in resting cells is kept low, at 50 - 100 nM, whereas the extracellular Ca^{2+} concentration is 1 - 3 mM. Biological membranes are relatively impermeable to Ca^{2+}-ions and maintain this steep gradient until extracellular signals lead to a Ca^{2+}-influx into the cytosol. Due to the large Ca^{2+} gradient across the membrane, even minor changes in membrane permeability will lead to significant variations in the cytosolic Ca^{2+} concentration. Several calcium transporting systems in the membrane have been studied extensively, such as the Ca^{2+}-channel (for a review see Reuter, 1984), the Na^+/Ca^{2+}-exchanger (for a review see Philipson, 1985), the plasma membrane Ca^{2+}-ATPase (for a review see Carafoli, 1991a, 1991b) and the sarco(endo)plasmic reticulum Ca^{2+}-ATPase (for a review see MacLennan et al., 1997). These transport systems have different kinetic properties, to serve the different requirements of cells during the functional cycle. There will be situations where Ca^{2+} must be regulated in the cytosol, or in other cell compartments, very rapidly and with high precision. Other situations may require slower movements of bulk amounts of Ca^{2+}. In general, the ATPases are selected for the high affinity transport and fine tuning of the Ca^{2+}, whereas the channels and exchangers are all low-Ca^{2+}-affinity systems (Carafoli, 1987).

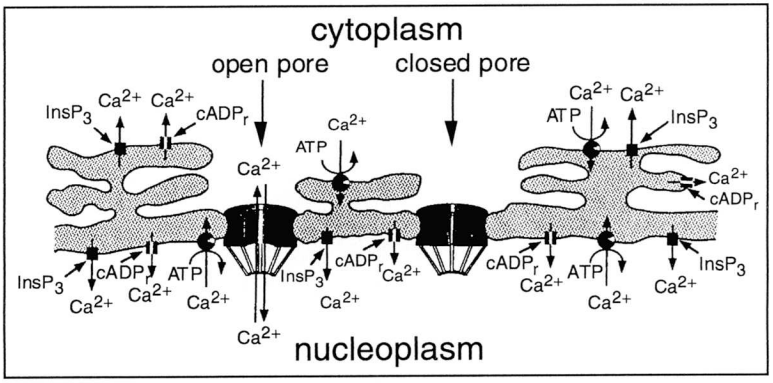


Figure 3.3 **The nuclear envelope in the regulation of calcium.** Pore complexes are represented in both calcium permeable and calcium sealed states. The latter state is indicated graphically by the adjoining of the pore subunits. The model shows the calcium transporting systems of the envelope: the Ca^{2+}-ATPase, the $InsP_3$-sensitive calcium release channels and the cADPr (ryanodine) sensitive channel, which are visualized on both the inner and outer membranes of the envelope (Santella, 1996).

Whereas Ca^{2+}-signalling in the cytoplasm has been studied extensively, much less is known about Ca^{2+}-signalling in the nucleus. The issue of nuclear calcium has been a topic of discussion for a long time and still is. In general, the nuclear envelope is seen as a structure that restricts only the permeability of molecules larger than about 20 kDa and therefore small ions such as Ca^{2+}-ions should in principle diffuse freely in and out of the nucleus.

On the other hand, the nuclear envelope is seen as an effective sieve that limits the traffic of Ca^{2+}-ions between the nucleus and the cytosolic compartment. Different experimental approaches have contributed to these conflicting conclusions (Bachs et al., 1992; Williams et al., 1987; Neylon et al., 1990; Himpens et al., 1994). It is clear, that in cell systems where the cytoplasmic compartment is spatially minor with respect to the nuclear one significantly different results by Ca^{2+} responses are observed with respect to those on cells in which large cytoplasmic and nuclear compartments coexist (in for example oocytes). Recent studies have shown that the nuclear Ca^{2+} can be regulated independently of cytosolic Ca^{2+} (Badminton et al., 1996). Figure 3.3 shows how Ca^{2+} can be regulated in the nucleus. Depending on the cell type, cytosolic Ca^{2+} either equilibrates almost instantaneously with the nuclear pool or is apparently filtered at the envelope level. The problem

of the control of the pore gating is still unsolved. Mechanisms based on the state of depletion of the cell Ca^{2+} stores (endoplasmic reticulum and lumen of the envelope) have been proposed. Apart from pore gating mechanisms, Ca^{2+}-transporters have been described in the nuclear envelope like Ca^{2+}-pumps as well as inositol 1,4,5-triphosphate ($InsP_3$) and cyclic ADP ribose-modulated (cADPr) Ca^{2+}-channels. A plausible picture on the Ca^{2+}-traffic in and out of the nucleus would have both (gated) nuclear pores and Ca^{2+}-transporters acting together to control the fluxes of Ca^{2+} and eventually control its homeostasis in the nucleoplasm (figure 3.3). Therefore, Ca^{2+} is not only important for regulating cytoplasmic events, but may play an important role in the nucleus as well. Certainly, it modulates important processes in the nucleus such as regulation of expression of certain genes, and participates in the control of the cell cycle (Whitaker and Patel, 1990; Santella and Carafoli, 1997).

3.1.5 The role of calcium dependent proteases in the cell cycle

It is interesting to notice that there is a role for Ca^{2+}-dependent proteases in the regulation of the cell cycle. Recent studies have shown that Ca^{2+} is related to the activation of the proteasome. Both proteolysis of cyclin B that promotes exit from mitosis and that of a proto-oncogene that releases the inhibition of cyclin proteolysis appear to be controlled by Ca^{2+}, the first through the activation of the ubiquitin-proteasome, the latter through the activation of calpain, a cysteine protease. The overall effects of calpain were studied in PtK1 cells, monitoring the redistribution of calpain during the mitotic process by immuno-fluorescence. The protease was found to relocate from a plasma membrane location in interphase to mitotic chromosomes during early and late anaphase. In telophase and cytokinesis, calpain then migrated to a bilaterally symmetrical location at the cell periphery and to an association with the cytoplasmic bridge. These findings were strongly indicative of the intervention of calpain in mitosis (Schollmeyer, 1988). The role of calpain in the mitotic and meiotic cycle thus seems very likely. Interestingly, when calpain was injected into these cells together with Ca^{2+}, the levels of the latter necessary to promote the metaphase/anaphase transition dropped about hundred fold from a value of 10 μM which is normally considered necessary to a value of about 0.1 μM. Several studies have shown effects on the cell cycle by injecting calpain to cells and these effects were inhibited by the co-injection of calpastatin, the natural specific

protein inhibitor of calpain, adding weight to the results obtained (March et al., 1993; Lane et al., 1992).

3.1.6 *Effect of calpain on the resumption of meiosis in starfish oocytes*

The cell cycle is frequently studied in eggs and oocytes, because these cells contain large nuclei, which make them accessible to a variety of experiments. Prophase arrested starfish oocytes resume meiosis, this means the large germinal vesicle (nucleus) breaks down, after the addition of the natural hormone 1-methyladenine (1-MA) (Kanatani et al., 1969). In many oocyte species this is accompanied by a cytoplasmic Ca^{2+} transient (Moreau et al., 1978). Mobilization of Ca^{2+} from intracellular pools occurs in oocytes, eggs and zygotes of other species also during fertilization and the first cell cycle, indicating a general role for Ca^{2+} in these processes (Means, 1994). It was also shown in prophase arrested oocytes that Ca^{2+} transients follow after cytoplasmic injection of $InsP_3$, ryanodine or caffeine (Chiba et al., 1990), indicating that two types of intracellular channels mediate the emptying of Ca^{2+} stores in most cells, those sensitive to $InsP_3$ and those sensitive to ryanodine and caffeine (Berridge, 1993). Addition of 1-MA induces three distinct Ca^{2+} spikes in starfish oocytes, two in the cytoplasm and one in the nucleus and is followed by the resumption of meiosis. Addition of modulators of the intracellular Ca^{2+} channels, $InsP_3$ or cADPr into the nucleus of starfish oocytes in the absence of the hormone 1-MA showed a similar effect on Ca^{2+} transients indicating an important role for Ca^{2+} in the mitotic and meiotic cell cycle (Santella and Kyozuka, 1994, 1997). This hormonal effect on the Ca^{2+} elevation could therefore also activate calpain. Experiments have shown that after injecting calpain directly into the oocytes from two starfish species, *Asterina pectinifera* and *Astropecten auranciacus* resumption of meiosis was observed (Santella et al., 1998). An important question arises which cellular substrates are attacked by calpain. Cytoskeletal proteins and microtubule associated proteins, that are preferred calpain substrates would be attractive candidates. In this study some proteins that are indeed related to the cytoskeleton could be identified that are involved in the process of resumption of meiosis after adding the hormone 1-MA.

3.2 Materials and Methods

3.2.1 Materials

Oocytes, stored in natural filtered sea water, from the starfishes *Asterina pectinifera* from the Mutsu Bay near Asamushi Marine Biological Station (Japan) or from *Astropecten auranciacus* from the Bay of Naples (Italy) were a kind gift of Dr. Luigia Santella (Stazione Zoologica "Anton Dohrn", Naples, Italy).

Ammonium persulphate (MicroSelect), dithiothreitol (MicroSelect), formaldehyde, formic acid, glycerol, glycine, iodoacetamide, SDS, silicon oil (DC 200, 20 mPa.s), silver nitrate and urea (MicroSelect) were purchased from Fluka AG (Buchs, Switzerland). Bromophenol blue and α-cyano-4-hydroxycinnamic acid were purchased from Aldrich GmbH (Buchs, Switzerland). Agarose, CHAPS, Nonidet P-40 and Trizma base were purchased from Sigma Chemical Co. (Buchs, Switzerland). Ammonium hydrogen carbonate and sodium carbonate were purchased from Merck AG (Darmstadt, Germany) Acetic acid, HPLC grade acetonitrile and HPLC grade methanol were purchased from Riedel-de Haën AG (Seelze, Germany). Acrylamide, N,N'-methylene bisacrylamide, carrier ampholytes and TEMED were purchased from BDH Laboratory Supplies (Poole, England). Immobiline strips (pH 3 - 10, linear) were purchased from Pharmacia (Uppsala, Sweden). Coomassie Brilliant Blue (Serva Blue G) was purchased from Serva (Heidelberg, Germany). Sequencing grade modified trypsin was purchased from Promega (Zürich, Switzerland). μCalpain was purchased from Calbiochem-Novabiochem AG (Läufenfingen, Switzerland). All other chemicals were of the highest purity commercially available.

3.2.2 Sample preparation for two-dimensional gel electrophoresis

Samples of oocytes from the two species *A. pectinifera* and *A. auranciacus* were prepared in the standard high pH urea/NP-40 "pink" mix (9 M urea, 4% NP-40 detergent, 2% ampholyte (pH 8 - 10), 1% DTT in water). A small trace of bromophenol blue was added as tracking dye. Usually samples were mixed in a ratio of 1:2 in respect to the solubilizing reagent. After solubilization the samples were centrifuged in an Eppendorf centrifuge at high speed for 10 min. to remove any undissolved particles

(Anderson, 1988). It was found that the amount of undissolved material in the samples of *A. auranciacus* was considerably larger than in the samples of *A. pectinifera*. The supernatants (approximately 75 µL) were loaded in the sample cups mounted on the immobilized pH gradient strips (IPG's) for the first dimension.

3.2.3 *CANP digestion on oocytes from A. pectinifera*

Before subjecting the oocytes to digestion with µCANP, an amount of oocytes in a buffer containing 10 mM Tris-HCl, pH 7.4, 100 mM KCl, 20 mM NaCl and 3 mM $MgCl_2$ were disrupted using a Branson small-tip sonicater. To one part of the disrupted cells was added EGTA (2 mM), to one part was added µCANP (120 units) and Ca^{2+} (0.5 mM) and to one part was added the specific inhibitor of calpain Cbz-Leu-Leu-Tyr-CHN_2 (0.3 mM). The samples were incubated for 60 min. at room temperature prior to solubilization in the "pink" mix and further processed as described in section 3.2.2.

Samples of isolated nuclei from the oocytes were treated in the same way, except that the sonication step was omitted.

3.2.4 *Two-dimensional gel electrophoresis*

(a) The first dimension. Prior to isoelectric focusing IPG strips were rehydrated overnight in a solution containing 8 M urea, 2% CHAPS, 10 mM DTT, 2% ampholyte (pH 4 - 8) with a trace of bromophenol blue. The samples were loaded on each Immobiline strip using the Pharmacia cup system. The strips were run parallel in batches of twenty on a Multiphor II system. The proteins were focused at 300 V to allow the samples to enter the gel, then the voltage was slowly increased to 3,500 V during 6 h and run at 3,500 V for 24 h. The total Vxh was always between 70,000 and 80,000.

(b) The second dimension. After the focusing was completed the strips were equilibrated for 20 min. in a solution containing 50 mM Tris-HCl, pH 6.8, 6 M urea, 25% glycerol, 2% SDS and 100 mM DTT followed by incubation for 5 min. in a solution containing 50 mM Tris-HCl, pH 6.8, 6 M urea, 25% glycerol, 2% SDS and 100 mM iodoacetamide. This was done to reduce and alkylate Cys residues. The strips were then transferred to 12% polyacrylamide gels (Laemmli, 1970). The Iso-Dalt apparatus from Hoefer (San Francisco, CA, USA) was used to run twenty gels in parallel. Samples were always run in multiple batches to insure reproducibility of the protein

pattern. The gels were run in a 40 L tank using a running buffer containing 25 mM Tris, 192 mM glycine and 0.1% SDS, pH 8.3 at 12 - 15 °C at a constant current of 400 mA for 16 - 24 h (Anderson, 1988).

Gel fixing and staining were done as described previously by Schägger and von Jagow (1987) since it results in a very clear background, suitable for scanning densitometry. Gels were fixed overnight in 50% methanol and 10% acetic acid in water, stained in 0.025% Serva Blue G in 10% acetic acid for 3 h and destained in 10% acetic acid until a clear background was obtained. Silver staining was done as described by Doucet and Trifaro (1988). Gels were fixed overnight in 40% ethanol and 10% acetic acid in water, then washed in water (3 x 20 min.) before sensitization with a solution of 5 mg/L DTT for 30 min. Gels were incubated in a 0.1% $AgNO_3$ solution for 30 min., the excess $AgNO_3$ solution washed off with water for 30 s before the gels were developed in a 3% Na_2CO_3 and 0.11% formaldehyde solution. Gels were placed in a 1% acetic acid solution to stop development.

Wet gels were scanned using a Personnel Densitometer from Molecular Dynamics (Sunnyvale, CA, USA) before analysis.

3.2.5 In-gel digestion of proteins

Spots of interest were excised from the 2-D gel and washed in water to remove excess acid (3 x 30 min.). Destaining of the gel pieces was done by adding 300 µL of 100 mM NH_4HCO_3 in 50% acetonitrile and incubating for 30 min. This was repeated until the blue dye was completely removed from the gel pieces. A small amount of acetonitrile was added to dehydrate the gel pieces followed by drying the pieces in a speed-vac. The gel pieces were rehydrated in a trypsin solution in 100 mM NH_4HCO_3 by very slowly adding the solution. About 1 µg of trypsin was added per protein sample. For further rehydration small portions of the 100 mM NH_4HCO_3 buffer were added until the gel pieces were fully covered with buffer. Digestion was performed overnight at room temperature.

Tryptic peptides were extracted from the gel by incubation for 3 h after adding 50 µL of 100 mM NH_4HCO_3 buffer to the sample. The peptides were collected and on the remaining gel pieces three further extractions were done with 50 µL of 50% acetonitrile and 5% formic acid for 1 h each (Rosenfeld et al., 1992). The fractions were pooled and dried in the speed-vac. The peptides were redissolved in 6 - 8 µL of 1% TFA prior to MALDI analysis.

3.2.6 MALDI-TOF mass spectrometry

The tryptic peptides were spotted onto the 100-position MALDI sample target. For each sample approximately 0.5 µL was required and the same amount of a saturated matrix solution was added (10 mg α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 1.25% TFA in water) and allowed to dry at ambient temperature. To obtain better spectra the spots on the MALDI target were washed by pipetting ice-cold water on the target spot, which was then removed by suction using a fine pipette after 5 s.

Mass spectra were recorded using a Voyager Elite MALDI-TOF mass spectrometer (Perseptive Biosystems, Framingham, MA, USA). Samples were analysed in delayed extraction reflector mode using an accelerating voltage of 20 kV, a pulse delay time of 75 ns, a grid voltage of 55% and a guide wire voltage of 0.05%. Spectra were accumulated for 32 or 64 laser shots.

3.2.7 nanoESI-MS/MS sequencing

Tryptic digests (sub-picomole amounts) from the proteins excised from the 2-D gels were desalted on the C_{18} membranes and redissolved in 50% methanol and 0.5% acetic acid before introducing into the mass spectrometer at a flowrate of 30 - 50 nL/min. using a nanoES capillary tip (Protana, Odense, Denmark). MS/MS sequencing was performed on a Finnigan MAT TSQ 700 mass spectrometer (San Jose, CA, USA). The instrument monitors the masses of the peptides, using the first quadrupole Q1 to scan the mass range from 500 to 2000 a.m.u. The peaks of interest were then selected in the first quadrupole Q1 with a mass window of ± 1.5 a.m.u. and the selected ions were subsequently subjected to fragmentation in the second quadrupole Q2, which is filled with argon as collision gas (at 2 mTorr pressure). The resulting fragments were analysed in the third quadrupole Q3, scanning the mass range from 50 to 2000 a.m.u.

The obtained MS/MS fragmentation spectra were used to search the protein and nucleic acid databases using the program SEQUEST in a fully automated mode (Eng et al., 1994).

3.2.8 Peptide mass fingerprinting using MassSearch

MassSearch searches the SwissProt and trEMBL databases by peptide masses after specific digestion of the protein. The algorithm was developed in collaboration with the group of Prof. Gaston Gonnet (Computation Biology Research Group, ETH Zürich). A complete description of the algorithm is described elsewhere (Gonnet, 1994). Appendix 5 shows an example of the MassSearch page on the WWW, as shown with Netscape Navigator.

3.3 Results and Discussion

3.3.1 Two-dimensional gel analysis of oocytes from A. pectinifera

In this section the analysis of 2-D gels of oocytes from the starfish *A. pectinifera* is described. Oocytes from *A. pectinifera* were from the Mutsu Bay (Japan). 2-D gels were performed as described in section 3.2.2 and 3.2.4. Figures 3.4 and 3.5 show examples of 2-D gels of oocytes from two sets of experiments. All the gels shown in this section are stained with the dye Coomassie Blue. After gel imaging, the manual analysis of the protein pattern before and after treatment with the hormone 1-MA was done. After the addition of the hormone 1-MA meiosis is resumed. This means that the nucleus breaks down. Spots that either change dramatically in intensity, disappear or appear after treatment of 1-MA are assigned and selected for further analysis in order to identify these proteins. Oocytes from *A. auranciacus* from the Bay of Naples (Italy) were applied too, but for some reason not understood yet, separation and visualization of the proteins was not of enough high quality that analysis on the spots could be performed. It was observed that after solubilization of the Italian species the amount of undissolved material was much larger than for the Japanese species.

Tables 3.1 and 3.2 summarize the results of the MassSearch output. One might find the identification output low (only 5 out of 37 spots that were analysed give a reliable output) but one has to keep in mind that the genome of *A. pectinifera* is not known at this stage and as a consequence only *de novo* sequenced proteins are available in the database. In figure 3.6 the mass spectrum for the tryptic digest for spot 13 from gel 3.4(A) is shown and in table 3.3 an example of an excerpt of a MassSearch output is shown with respect to the identification of this spot (tubulin alpha) from gel 3.4(A). This identification is reliable since the score is very high and the output shows all the same proteins, but from different species. For dynamin, lamin and two kinds of actins from gel 3.5(A) this is similar, a high score, combined with the same type of protein (high confidence level). All the five proteins disappear upon adding the natural hormone 1-MA. As mentioned before, after the addition of the hormone 1-MA prophase arrested oocytes resume meiosis, thus the nucleus breaks down. This breakdown is associated with the degradation of certain proteins. The proteins are degraded through the action of the hormone. These are all proteins related

to the cytoskeleton of the cell. This is an interesting finding since cytoskeletal proteins play major roles in mitosis and meiosis. Furthermore, they are potential targets as substrates for calpain. The hormone 1-MA that induces maturation does induce Ca^{2+} elevation in the cell and could therefore provide the ideal conditions for activation of calpain. The role of calpain and its activator Ca^{2+} in the mitotic (and meotic) seems thus very likely. Still not clear is the cellular locus of calpain action, in the cytoplasm or in the nucleus. It was found in previous studies that calpain is optimally activated by millimolar Ca^{2+} levels, but interestingly however, in the presence of DNA the optimal Ca^{2+} requirement dropped to the micromolar level (Mellgren et al., 1993) This could imply that the locus of action of calpain, as it activates the mitotic process, is in the nucleus.

Experiments have been carried out to determine whether the proteins that are subjected to degradation are nuclear proteins. In order to prove this, 2-D gel electrophoresis with nuclei, isolated from the oocytes, were performed (see section 3.3.2).

Some other tryptic digest fingerprints matched to proteins that could play a role as well in the breakdown of the nucleus. However, these had to be confirmed by other techniques such as nanoESI-MS/MS, since identification by mass fingerprinting alone did not resulted in unambiguously data output. But nanoESI-MS/MS did not result in more proteins that could be identified, although the MS/MS data confirmed the data on the already identified proteins. There are no homologous proteins in the database since starfish proteins and genes have not been sequenced much.

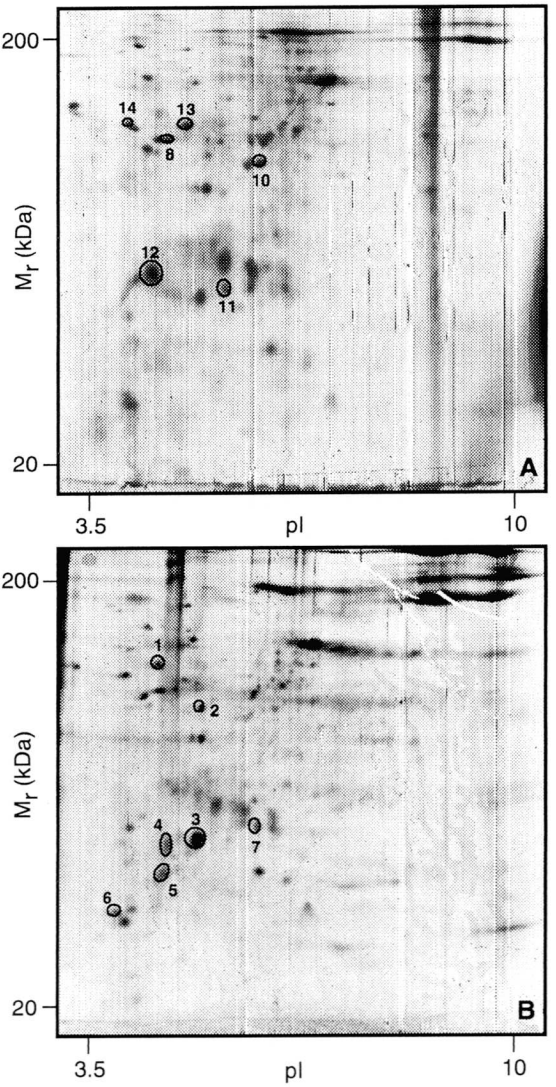


Figure 3.4 **Two-dimensional gels of oocytes from *A. pectinifera*.** (A) Control oocytes and (B) oocytes treated with the hormone 1-MA after 70 min. The spots that have changed upon treatment with 1-MA are numbered and selected for further analysis.

Table 3.1 Analysis of proteins from oocytes from *A. pectinifera* from the 2-D gels in figure 3.4. The masses are monoisotopic. Identification is performed using MassSearch.

Spot No	Masses of tryptic digest (Da)	Identification
1	1053.9, 1130.9, 1160.0, 1202.2, 1246.0, 1315.3, 1344.1, 1496.3, 1524.4, 1603.4, 1608.4, 1894.6, 1941.5, 2022.9	tubulin beta
2	1118.8, 1246.9, 1342.9, 1418.0, 1478.8, 1712.1, 1740.2, 1893.3	unknown
3	1263.9, 1751.9, 1813.9, 1830.0, 1917.1, 1921.4, 1951.1, 2008.2, 2024.1, 2328.4, 2443.5	inhibitor of apoptosis
4	1060.2, 1249.4, 1270.2, 1317.9, 1362.0, 1404.8, 1450.1, 1494.1, 1538.2, 1582.3, 1626.2, 1670.3, 1714.4	actinin
5	1672.3, 1923.1, 1941.4, 2009.5, 2083.4	unknown
6	935.8, 1044.5, 1122.9, 1180.0, 1234.9, 1278.2, 1309.2, 1317.1, 1392.2, 1714.5, 1994.5	unknown
7	981.8, 1022.9, 1045.9, 1486.3	unknown
8	1054.1, 1077.9, 1131.1, 1160.1, 1246.1, 1259.2, 1288.3, 1329.3, 1510.4, 1539.6, 1603.5, 1618.5, 1887.6, 1959.7	unknown
10	1401.2, 1721.4, 1753.4, 1812.4, 1828.3, 1844.2, 1861.3, 1964.6, 2366.7	unknown
11	882.8, 1180.1, 1202.2, 1278.3, 1321.0, 1486.3, 1587.4, 1708.4	unknown
12	1264.2, 1297.2, 1752.6, 1798.5, 1814.4, 1869.3, 1874.6, 1903.4, 1920.5, 1960.6, 1975.6	nuclear migration protein
13	1099.9, 1393.1, 1410.2, 1457.3, 1687.3, 1718.3, 1756.4, 1783.5, 1824.3, 1993.4, 2007.4, 2408.8	tubulin alpha
14	1069.9, 1202.1, 1216.9, 1354.1, 1377.9, 1405.1, 1528.2, 1573.2, 1868.4, 1967.4, 2016.3	unknown

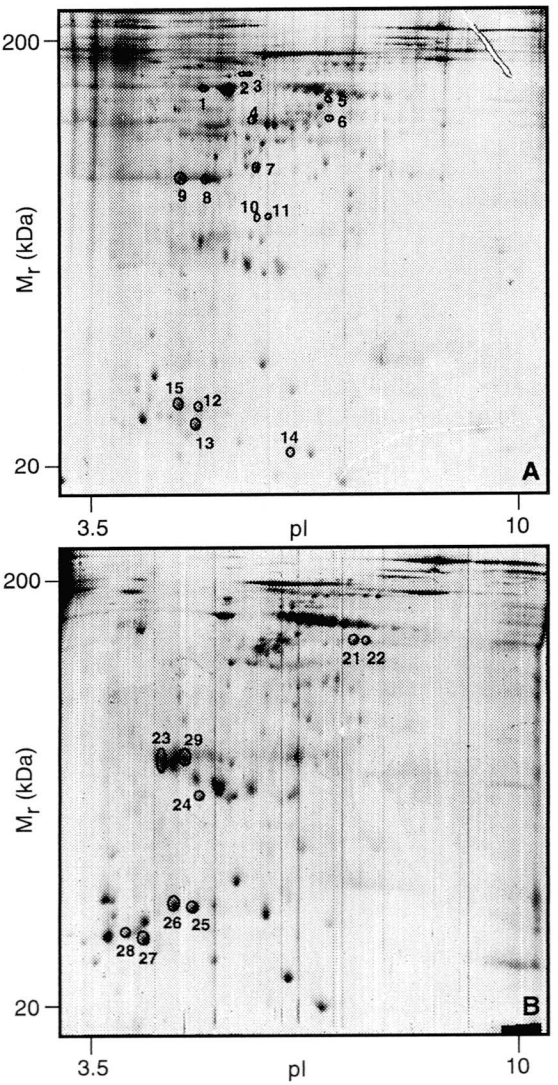


Figure 3.5 **Two-dimensional gels of oocytes from *A. pectinifera*.** (A) Control oocytes and (B) oocytes treated with the hormone 1-MA after 70 min. The spots that have changed upon treatment with 1-MA are numbered and selected for further analysis.

Table 3.2 Analysis of proteins from oocytes from *A. pectinifera* from the 2-D gels in figure 3.5. The masses are monoisotopic. Identification is performed using MassSearch.

Spot No	Masses of tryptic digest (Da)	Identification
1	1084.3, 1253.3, 1375.2, 1430.3, 1486.2, 1641.3, 1787.8, 1922.5, 1995.4, 2011.2	dynamin
2	1058.9, 1193.6, 1277.6, 1345.5, 1457.8, 1588.5, 1623.7, 1917.4, 2149.7	lamin
3	1375.8, 1697.9, 2159.9, 2504.4, 2553.9, 2780.7, 2914.2	unknown
4	1254.8, 1365.8, 1377.9, 1419.8, 1565.8, 1638.9, 1753.2, 1847.1, 2013.2, 2418.9, 2781.1	unknown
5	1059.3, 1265.4, 1277.9, 1316.9, 1362.8, 1388.9, 1413.8, 1524.7, 1618.9, 1685.1, 1955.9, 1985.1	unknown
6	1024.6, 1076.7, 1271.8, 1281.7, 1496.6, 1579.9, 1609.9, 1854.1, 1892.1, 2037.2, 2079.9	replication protein
7	936.6, 1028.5, 1051.6, 1075.3, 1640.8, 1712.8, 1812.9, 1828.8, 1844.9, 1862.0, 1965.0	cell division and growth cycle
8	976.5, 1055.3, 1060.7, 1198.7, 1225.2, 1325.4, 1352.6, 1515.6, 1547.7, 1790.9, 1954.1, 1960.9, 1976.9, 2008.9, 2185.1	actin
9	945.6, 976.5, 1059.3, 1130.3, 1198.8, 1515.5, 1547.7, 1790.9, 1924.9, 1960.8, 1972.7	actin
10	2082.8, 2194.5, 2522.1, 2914.1	unknown
11	1059.1, 1479.3, 1825.4, 2082.5, 2417.9, 2521.8	unknown
12	1160.4, 1276.4, 1320.6, 1471.2, 1492.5, 1634.7, 1677.5, 1724.7, 1756.6, 2055.8	unknown
13	1059.3, 1259.6, 1521.5, 1538.6, 2080.7, 2812.0	unknown
14	1036.7, 1354.9, 1369.8, 1385.9, 1426.8, 1438.8, 1994.9, 2048.8, 2080.2, 2135.9	unknown

Table 3.2 (continued) Analysis of proteins from oocytes from *A. pectinifera* from the 2-D gels in figure 3.5. The masses are monoisotopic. Identification is performed using MassSearch.

Spot No	Masses of tryptic digest (Da)	Identification
15	1059.4, 1259.8, 1521.7, 1538.8, 1994.2, 2083.1, 2153.1	unknown
21	984.6, 994.3, 1059.5, 1073.2, 1133.5, 1164.9, 1226.4, 1264.5, 1324.9, 1352.8, 1621.1, 2083.1	unknown
22	1059.4, 1233.8, 1336.8, 1501.4, 1691.0, 1830.3, 1922.2, 2012.1, 2082.3	unknown
23	1059.4, 1233.8, 1336.8, 1501.4, 1691.0, 1706.9, 1924.1, 2083.0	unknown
24	1059.3, 1075.3, 1132.4, 1275.8, 1324.6, 1353.7, 1397.9, 1420.0, 1485.6, 1708.9, 1751.9, 1789.7, 1838.9, 1922.9	unknown
25	1059.3, 1133.3, 1264.2, 1424.0, 1574.6, 1638.9, 1753.9, 1798.9, 1814.8, 1836.8, 1920.7, 2008.9	unknown
26	1059.2, 1076.3, 1123.9, 1133.1, 1750.8, 1798.7, 2008.8, 2082.7	unknown
27	903.3, 1059.3, 1133.3, 1260.5, 1324.6, 1351.6, 1510.5, 1524.5, 1923.9, 2082.9	unknown
28	1057.7, 1277.8, 1320.7, 1325.7, 1393.9, 1475.8, 1517.7, 1548.8, 1670.9, 1707.9 1791.8, 1924.2, 1994.1, 2083.1	unknown
29	964.5, 1001.3, 1059.4, 1072.7, 1132.4, 1277.9, 1324.7, 1355.9, 1493.8, 1706.9, 1923.1, 2083.1	

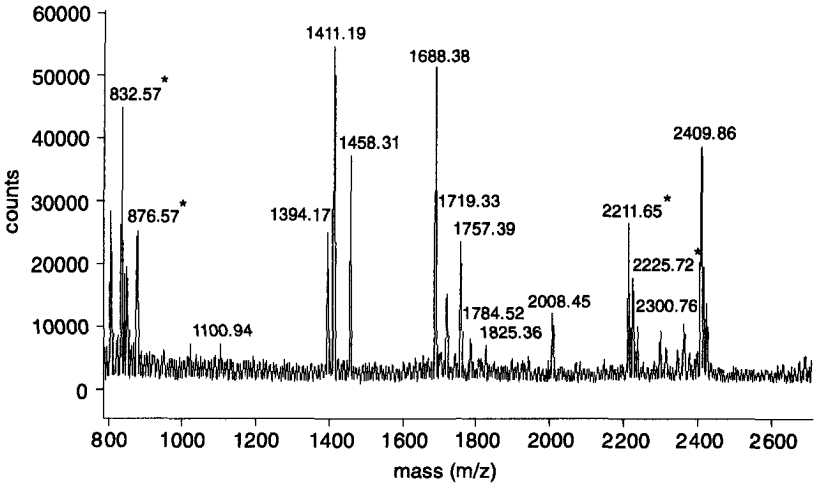


***Figure 3.6* MALDI mass spectrum of the tryptic digest of spot 13.** A typical mass spectrum for a tryptic digest of a protein is shown. The peaks appearing in the mass range that is shown in this figure are useful for database searching, the lower masses show mainly matrix associated peaks and higher masses correspond to incomplete digestion products. The peaks marked with an * are products from auto proteolysis of trypsin and are not entered in the MassSearch program.

Table 3.3 Excerpt of the output of the computer program MassSearch. The score is the quality of the match between the given masses and a protein in the database, the higher the score the better the match, n shows the theoretical number of tryptic digests produced for a protein between the minimum and maximum mass values used in the search, k indicates the number of masses that are successfully matched against those of the theoretical digestion. AC, DE and OS are parameters from the SwissProt database and stand for ACcession number (of the database), DEscription (name and type of the protein) and Origin of Species (species name where protein originates from with the name in Latin and if available in English), respectively. In the last line the unmatched masses are given.

score	n	k	AC	DE	OS
109.5	10	7	P05214	Tubulin alpha-3 and alpha-7 chain	Mus musculus (mouse)
Unmatched weights				[1392.1, 1717.3, 1782.5, 1992.4]	
108.9	10	7	P41383	tubulin alpha-2 and alpha-4 chain	Patella vulgata (common limpet)
Unmatched weights				[1392.1, 1717.3, 1782.5, 1992.4]	
106.7	11	7	P18258	Tubulin alpha-1 chain	Paracentrotus lividus (common sea urchin
Unmatched weights				[1392.1, 1717.3, 1782.5, 1992.4]	

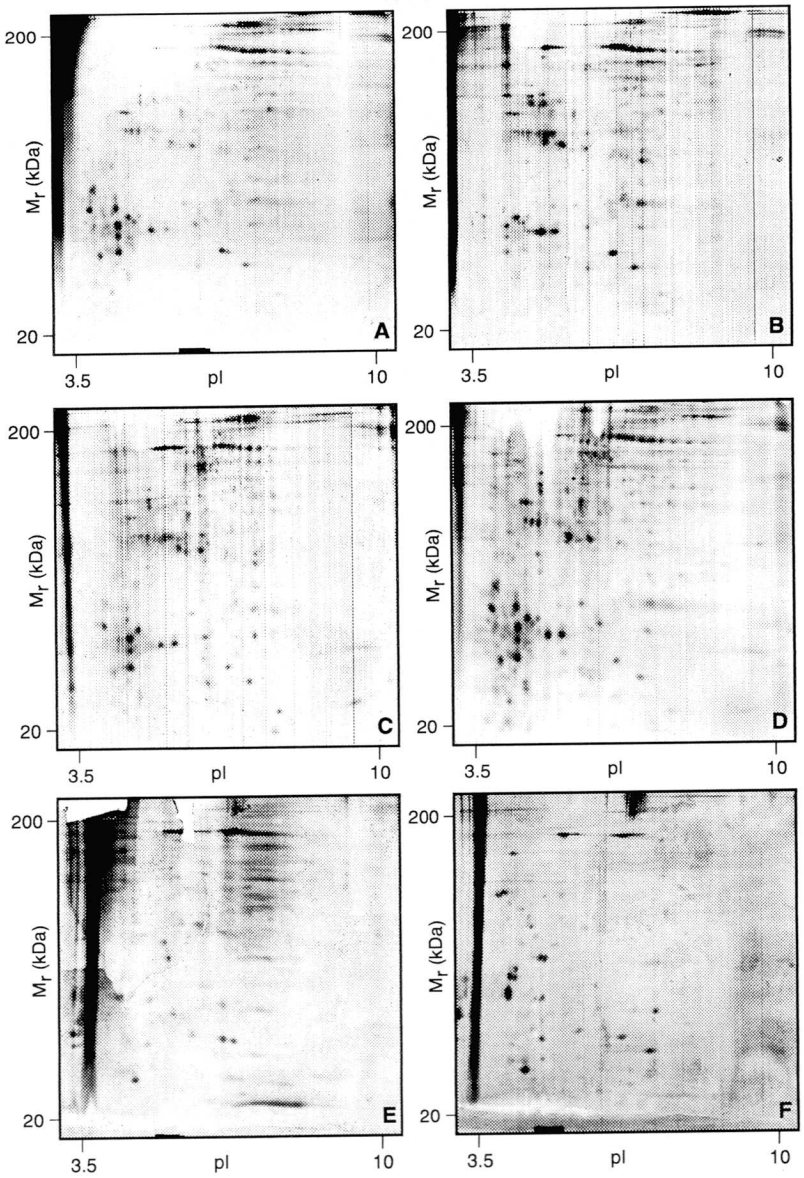
Table 3.3 (continued) Excerpt of the output of the computer program MassSearch.

score	n	k	AC	DE	OS
100.1	9	7	P06603	Tubulin alpha-1 chain	Drosophila melanogaster (fruit fly)
Unmatched weights				[1392.1, 1717.3, 1755.4, 2006.4]	
100.1	9	7	P06605	Tubulin alpha-3 chain	Drosophila melanogaster (fruit fly)
Unmatched weights				[1392.1, 1717.3, 1755.4, 2006.4]	
96.9	10	7	P18288	Tubulin alpha chain, testis-specific	Oncorhynchus mykiss (rainbow trout)
Unmatched weights				[1392.1, 1717.3, 1755.4, 1992.4]	
89.2	12	6	P04687	Tubulin alpha-1 chain	Homo sapiens (human)
Unmatched weights				[1392.1, 1686.3, 1717.3, 1782.5, 1992.4]	
89.2	12	6	P02551 P05212 P05210	Tubulin alpha-1 chain	Rattus norvegicus (rat) Mus musculus (mouse) Cricetulus griseus (Chinese hamster)
Unmatched weights				[1392.1, 1686.3, 1717.3, 1755.4, 1782.5, 1992.4]	
86.1	12	6	P30436	Tubulin alpha chain	Oncorhynchus keta (chum salmon)
Unmatched weights				[1392.1, 1686.3, 1755.4, 1782.5, 1992.4, 2006.4]	

3.3.2 *The effect of calpain on the degradation of the nucleus*

In figure 3.7 the effect of μCANP on the breakdown of the nucleus of *A. pectinifera* is shown. As mentioned before, if the degradation of the proteins, involved in the breakdown of the nucleus, is due to the activation of the Ca^{2+}-dependent protease calpain, this could imply a more important role for Ca^{2+} in the cell cycle and the nucleus.

Figure 3.7 **The effect of calpain on the degradation of the nucleus.** (A) control oocytes (B) oocytes treated with the specific CANP inhibitor Cbz-Leu-Leu-Tyr-CHN_2 (C) oocytes treated with μCANP/Ca^{2+} (D) oocytes treated with 1-MA after 70 min. (E) control nuclei and (F) nuclei treated with μCANP/Ca^{2+}. The gels in figures (E) and (F) are silver stained.


200
20
M_r (kDa)
A
3.5
pI
10
200
20
M_r (kDa)
B
3.5
pI
10
200
20
M_r (kDa)
C
3.5
pI
10
200
20
M_r (kDa)
D
3.5
pI
10
200
20
M_r (kDa)
E
3.5
pI
10
200
20
M_r (kDa)
F
3.5
pI
10

In figure 3.7(A) a 2-D gel with control oocytes is shown as well as in figure 3.7(B). However, the difference between the samples is the presence of the specific calpain inhibitor Cbz-Leu-Leu-Tyr-CHN_2 in the latter one. In fact, this gel is the real control, since the gel in figure 3.7(A) shows a protein pattern that could be of the breakdown products of endogenous calpain. Figure 3.7(C) shows the gel of the μCANP treated oocytes and shows many degradation products. As a comparison a 2-D gel of the oocytes treated with the natural hormone 1-MA after 70 min. is show in figure 3.7(D). Many of the same breakdown products can be seen in both of these gels. Finally, in figures 3.7(E) and 3.7(F) silver stained gels of isolated control nuclei and nuclei treated with μCANP are shown, respectively. It is clear that some of the degraded proteins origin from the nucleus, although one has to take care when comparing Coomassie Blue stained gels with silver stained gels.

It can be concluded that CANP has an effect on the breakdown of some (cytoskeletal) proteins. This effect of CANP seems to be of the same order as the effect that is induced upon adding the hormone 1-MA. Four protein spots were excised and in-gel digested for further analysis. But peptide mass fingerprinting and peptide fragment mass fingerprinting did not result in unambiguous identification of the proteins. Further, some of the degraded proteins are nuclear proteins. This is an important finding and could imply a more important role for Ca^{2+} as the second messenger in the nucleus.

3.3.3 *Time course on the effect of 1-methyladenine on A. pectinifera*

In order to see the time range of the process of the breakdown of the nucleus 2-D gels representing a time course on the effect of 1-MA on oocytes are shown in figure 3.8. The time of 20 min. is chosen since it is known that this is the time before the nucleus starts to break down. Although these gels are difficult to analyse it can be seen that the protein pattern reaches actually after these 20 min. already a stable state. The time of 70 min. is the time where the proteins were analysed (section 3.3.1). This time resembles the stage after the first meiotic division. The time point at 116 min. resembles the stage after the second meiotic division. But as can be seen from the 2-D gels, the pattern does not really change after 20 min.

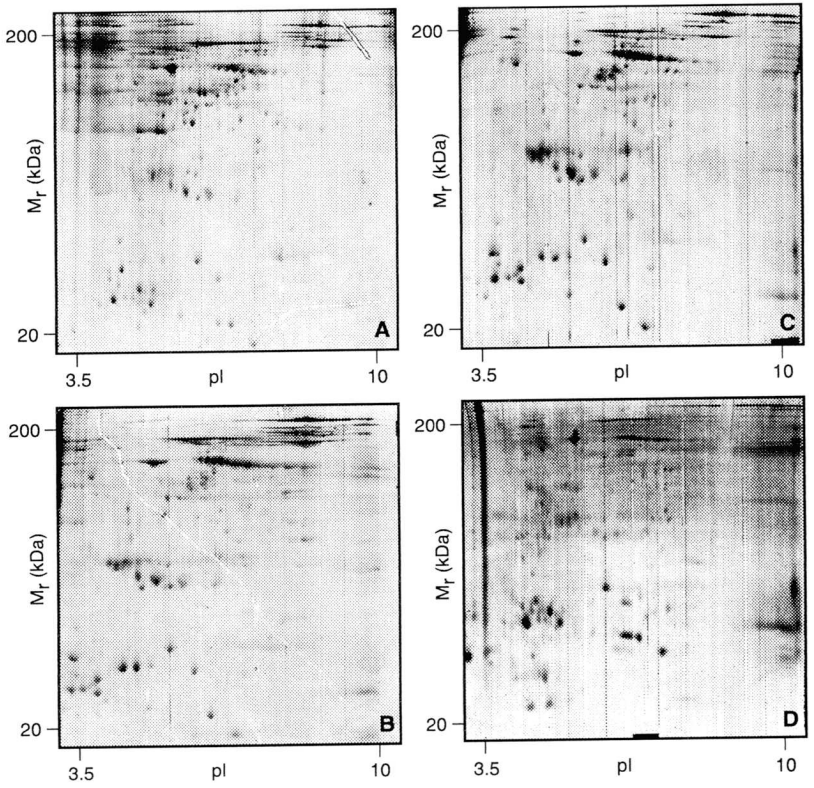


***Figure 3.8* Time course on the effect of the hormone 1-methyladenine on oocytes from *A. pectinifera*.** (A) control oocytes (B) oocytes treated with the hormone 1-MA after 20 min. (C) 70 min. and (D) 116 min.

3.3.4 *Comparison of oocytes from A. pectinifera and A. auranciacus*

In this experiment oocytes from two different species of starfish were compared. Figures 3.9(A) and 3.9(B) show 2-D gels of *A. pectinifera* before and after treatment with 1-MA, whereas figure 3.9(C) and 3.9(D) show 2-D gels of *A. auranciacus* before and after treatment with 1-MA. Even though the times after treatment with the hormone are not the same, they resemble equal states in the cell cycle of these types of oocytes. It is clear from this figure that the two species are completely distinct. The protein pattern of oocytes from *A. auranciacus* is much less resolved than that one of *A.*

pectinifera. There is no good explanation for the fact that the oocytes of *A. auranciacus* do not give such a clear protein pattern. Different methods of solubilization have been tried without improvement. Actually, one should not expect the same protein pattern, since the two species of starfish differ completely. In general, one has to be extremely careful when comparing 2-D gels, especially with eukaryotic cells. It is known that 2-D gels, even with proteins from the same species, can differ to some extent, due to different post-translational modifications of proteins.

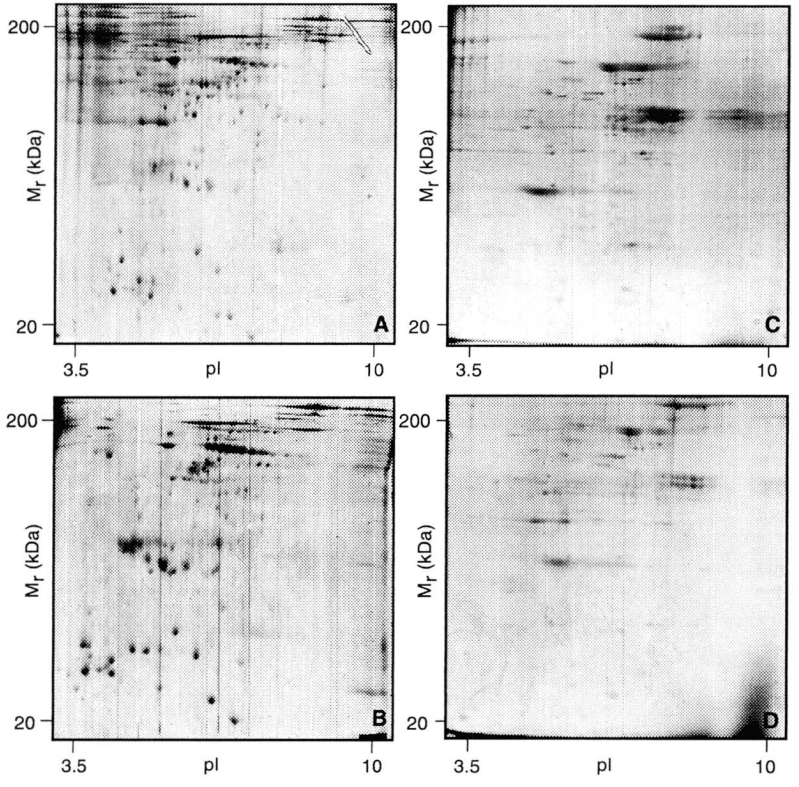


***Figure 3.9* Comparison of oocytes from *A. pectinifera* and *A. auranciacus*.** (A) control oocytes of *A. pectinifera* (B) oocytes from *A. pectinifera* treated with the hormone 1-MA after 70 min. (C) control oocytes of *A. auranciacus* and (D) oocytes from *A. auranciacus* treated with the hormone 1-MA after 110 min.

3.4. Conclusions

Two-dimensional gel electrophoresis is currently the only method available to separate complex protein mixtures from whole cells or tissue. In this study the effect of the natural hormone 1-MA on prophase arrested oocytes from two species, *A. pectinifera* and *A. auranciacus* at the protein level is studied. After adding the hormone meiosis is resumed and this is accompanied with the breakdown of the nucleus. After careful analysis of the different 2-D gels of the oocytes from *A. pectinifera* 37 protein spots were excised and in-gel digested with the protease trypsin. The resulting peptide masses were measured by MALDI-MS and protein identification was done by peptide mass fingerprinting using the program MassSearch. Out of the 37 protein spots the following proteins tubulin alpha, dynamin, lamin and two types of actin were matched with a high score. The identification was confirmed by nanoESI-MS/MS fragmentation analysis. Unfortunately, the other mass fingerprints did not match with high scores to significant proteins. The output of the identification might seem low (14%), but one has to keep in mind that the genome of the starfish *A. pectinifera* is unknown, in fact only a few proteins from this species and related species are sequenced directly. The identified proteins are all related to the cytoskeleton and more interestingly they are potential targets for the Ca^{2+}-dependent protease calpain. Some of the proteins are indeed degraded by calpain and originate from the nucleus. This could imply a more important role for Ca^{2+} in the cell cycle and in the nucleus.

Oocytes from the starfish *A. auranciacus* showed a protein separation pattern at a much lower resolution and therefore analysis could not be done on this species.

The time course shows that the degradation of proteins involved in the breakdown of the nucleus is rather fast. It is seen that after a period of 20 min. after adding the hormone the protein pattern seems stable. The analysis was done 70 min. after adding the hormone. This is at the stage corresponding to the first meiotic division.

The two species were also directly compared. Comparison of 2-D gels is difficult since a minor change in the protein properties might have a dramatic effect on its behaviour on a 2-D gel. Especially, when comparing proteins from different species, even if the species are closely related, like in the case with *A. pectinifera* and *A. auranciacus*, large differences in the

protein pattern can be observed. This is clearly the case in this comparison, even though the separation of the oocytes from *A. auranciacus* is not so good.

Finally, it must be mentioned that the methods that were applied in this study for protein identification are very powerful. Methods for improving the solubilization of the oocyte samples might result in even better separation of the proteins.

4. References

Abola, E.E., Manning, N.O., Prilusky, J., Stampf, D.R. and Sussman, J.L. (1996) The Protein Data Bank: current status and future challenges. *J. Res. Natl. Inst. Stand. Technol.* **101**, 231 - 241.

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651 - 1656.

Adams, M.D. (1996) Serial analysis of gene expression: EST's get smaller. *BioEssays* **18**, 261 - 262.

Aebersold, R.H., Teplow, D.B., Hood, L.E. and Kent, S.B. (1986) Electroblotting onto activated glass. High efficiency preparation of proteins from analytical SDS-PAGE gels for direct sequence analysis. *J. Biol. Chem.* **261**, 4229 - 4238.

Aebersold, R.H., Leavitt, J., Saavedra, R.A., Hood, L.E. and Kent, S.B. (1987) Internal amino acid sequence analysis of proteins separated by 1- or 2-D gel electrophoresis after *in situ* digestion on nitrocellulose. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6970 - 6974.

Aebersold, R. (1990) High sensitivity sequence analysis of proteins separated by polyacrylamide gel electrophoresis. In: Chrambach, A., Dunn, M.J. and Ladula, B.J. (eds.) *Advances in Electrophoresis.* VCH-Verlag, Weinheim.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson J.D. (1994) *Molecular Biology of the Cell (3rd. edition).* Garland Publishing Inc., New York.

Anderson, N.L., Taylor, J., Scandora, A.E., Coulter, B.P. and Anderson, N.G. (1981) The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clin. Chem.* **27**, 1807 - 1820.

Anderson, N.G. and Anderson, N.L. (1982) The human protein index. *Clin. Chem.* **28**, 739 - 748.

Anderson, L. (1988) *Two-Dimensional Electrophoresis. Operation of the ISO-DALT System.* Large Scale Biology Press, Washington D.C.

Anderson, L. and Seilhammer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533 - 537.

Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts and new words. *Electrophoresis* **19**, 1853 - 1861.

Appel, R.D., Hochstrasser, D.F., Funk, M., Vargas, J.R., Pellegrini, C., Muller, A.F. and Scherrer, J.R. (1991) The MELANIE project: from a biopsy to automatic protein map interpretation by computer. *Electrophoresis* **12**, 722 - 735.

Appel, R.D., Sanchez, J.C., Bairoch, A., Golaz, O., Miu, M., Pasquali, C., Vargas, R.J., Hughes, G. and Hochstrasser, D.F. (1993) SWISS-2DPAGE: a database of two-dimensional gel electrophoresis images. *Electrophoresis* **14**, 1232 - 1238.

Appel, R.D., Bairoch, A. and Hochstrasser, D.F. (1994) A new generation of information retrieval tool for biologists: the example of the expasy WWW server. *Trends Biochem. Sci.* **19**, 258 - 260.

Appel, R.D., Bairoch, A., Sanchez, J.C., Vargas, J.R., Golaz, O., Pasquali, C. and Hochstrasser, D.F. (1996) Federated two-dimensional electrophoresis database: a simple means of publishing two-dimensional electrophoresis data. *Electrophoresis* **17**, 540 - 546.

Bachs, O., Agell, N. and Carafoli, E. (1992) Calcium and calmodulin function in the cell nucleus. *Biochim. Biophys. Acta* **1113**, 259 - 270.

Badminton, M., Campbell, A.K. and Rembolds, C. (1996) Differential regulation of nuclear and cytosolic Ca^{2+} in HeLa cells. *J. Biol. Chem.* **271**, 31210 - 31214.

Bailey, J.M. and Shively, J.E. (1990) Carboxy-terminal sequencing: formation and hydrolysis of C-terminal peptidylthiohydantoins. *Biochemistry* **29**, 3145 - 3156.

Bailey, J.M., Nikfarjam, F., Shenoy, N.R. and Shively, J.E. (1992) Automated carboxy-terminal sequence analysis of peptides and proteins using diphenylphosphoro-isothiocyanatidate. *Protein Sci.* **1**, 1622 - 1633.

Bailey, J.M., Tu, O., Issai, G., Ha, A. and Shively, J.E. (1995) Automated carboxy-terminal sequence analysis of polypeptides containing C-terminal proline. *Anal. Biochem.* **224**, 588 - 596.

Bairoch, A. and Apweiler, R. (1997) The SWISS-PROT protein sequence data bank and its new supplement trEMBL. *Nucleic Acids Res.* **25**, 31 - 36.

Baitinger, C., Alderton, J., Poeni, M., Schulman, H. and Steinhardt, R.A. (1990) Multifunctional Ca^{2+}/calmodulin-dependent protein kinases is necessary for nuclear envelope breakdown. *J. Cell Biol.* **111**, 1763 - 1773.

Barrett, G.C. (1967) Cleavage of N-thiobenzoyl-dipeptides with trifluoroacetic acid: the basis of a new stepwise degradation of polypeptides. *Chem. Commun.* 487 - 488.

Bartlet-Jones, M., Jeffery, W.A., Hansen, H.F. and Pappin, D.J. (1994) Peptide ladder sequencing by mass spectrometry using a novel, volatile degradation reagent. *Rapid Comm. Mass Spec.* **8**, 737 - 742.

Bauw, G., De Loose, M., Inze, D., Van Montagu, M. and Vandekerckhove, J. (1987) Alterations in the phenotype of plant cells studied by amino terminal sequence analysis of proteins electroblotted from 2D gels. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4806 - 4810.

Beadle, G.W. and Tatum, E.L. (1941) Genetic control of biochemical reactions in neurospora. *Proc. Natl. Acad. Sci. U.S.A.* **27**, 499 - 506.

Berridge, M.J. (1993) Inositol triphosphate and calcium signalling. *Nature* **361**, 315 - 325.

Biemann, K. (1988) Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spec.* **16**, 99 - 100.

Biemann, K. (1990) Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Meth. Enzym.* **193**, 886 - 887.

Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Görg, A., Westermeier, R. and Postel, W. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem. Biophys. Meth.* **6**, 317 - 339.

Bjellqvist, B., Sanchez, J.C., Pasquali, C., Ravier, F., Paquet, N, Frutiger, S., Hughes, G.J. and Hochstrasser, D.F. (1993a) Micropreparative 2-D electrophoresis allowing the separation of milligram amounts of proteins. *Electrophoresis* **14**, 1375 - 1378.

Bjellqvist, B., Hughes, G., Pasquali, C., Paquet, N, Ravier, F., Sanchez, J.C., Frutiger, S. and Hochstrasser, D.F. (1993b) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequence. *Electrophoresis* **14**, 1023 - 1031.

Bjellqvist, B., Basse, B., Olsen, E. and Celis, J.E. (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* **15**, 529 - 539.

Blattner, F.R., Plunkett, G. , Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453 - 1474.

Boyd, V.L., Bozzini, M., Zon, G., Noble, R.L. and Mattaliano, R.J. (1992) Sequencing of peptides and proteins from the carboxy terminus. *Anal. Biochem.* **206**, 344 - 352.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., Fitzgerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelly, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts, K.M., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O. and Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococccus jannaschii*. *Science* **273**, 1058 - 1073.

Burkhart, W.A., Moyer, M.B., Bailey, J.M. and Miller, C.G. (1996) Electroblotting proteins to teflon tape and membranes for N- and C-terminal sequence analysis. *Anal. Biochem.* **236**, 364 - 367.

Campbell, K.P., MacLennan, D.H. and Jorgensen, A.O. (1983) Staining of the Ca^{2+} binding proteins, calsequesterin, calmodulin, troponin C and S-100, with the cationic dye "Stains-all". *J. Biol. Chem.* **258**, 11267 - 11273.

Carafoli, E. (1987) Intracellular calcium homeostasis. *Annu. Rev. Biochem.* **56**, 395 - 433.

Carafoli, E. (1991a) The calcium pumping ATPase of the plasma membrane. *Annu. Rev. Physiol.* **53**, 531 - 547.

Carafoli, E. (1991b) Calcium pump of the plasma membrane. *Physiol. Rev.* **71**, 129 - 153.

Cash, P. (1991) The application of two-dimensional polyacrylamide gel electrophoresis to medical microbiology: molecular epidemiology of viruses and bacteria. *Electrophoresis* **12**, 592 - 604.

Celis, J.E., Rasmussen, H.H., Olsen, E., Madsen, P., Leffers, H., Honore, B., Dejgaard, K., Gromov, P., Hoffmann, H.J., Nielsen, M., Vassilev, A., Vintermyr, O., Hao, J., Celis, A., Basse, B., Lauridsen, J., Ratz, G.P., Andersen, A.H., Walbum, E., Kjaergaard, I., Puype, M., Van Damme, J. and Vandekerckhove, J. (1993) The human keratinocyte two-dimensional database: update 1993. *Electrophoresis* **14**, 1091 - 1198.

Chait, B.T., Wang, R., Beavis, R.C. and Kent, S.B. (1993) Protein ladder sequencing. *Science* **262**, 89 - 92.

Chiba, K., Kado, R.T. and Jaffe, L.A. (1990) Development of calcium release mechanism during starfish oocyte maturation. *Dev. Biol.* **140**, 300 - 306.

Chothia, C. and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823 - 826.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. and Barell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537 - 544.

Colston, M.J. and Davis, E.O. (1994) The ins and outs of protein splicing elements. *Mol. Microbiol.* **12**, 359 - 363.

Cooper, A.A. and Stevens, T.H. (1995) Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends Biochem. Sci.* **20**, 351 - 356.

Dole, M., Mack, L.L., Hines, R.L., Mobley, R.C., Ferguson, L.D. and Alice, M.B. (1968) Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240 - 2249.

Doucet, J.P. and Trifaro, J.M. (1988) A discontinuous and highly porous sodium dodecyl sulphate-polyacrylamide slab gel system of high resolution. *Anal. Biochem.* **168**, 265 - 271.

Draetta, G., Luca, F., Westendorf, J., Brizuela,L., Ruderman, J. and Beach, D. (1989) Cdc2 protein kinase is complexed with both cyclinA and B: evidence for proteolytic inactivation of MPF. *Cell* **56**, 829 - 838.

Dunphy, W.G., Brizuela, L., Beach, D. and Newport, J. (1988) The Xenopus *cdc2* protein is a component of MPF, a cytoplasmic regulator of mitosis. *Cell* **54**, 423 - 431.

Eckershorn, C., Jungblut, P., Mewes, W., Klose, J. and Lottspeich, F. (1988) Identification of mouse brain proteins after two-dimensional electrophoresis and electroblotting by microsequence analysis and amino acid composition analysis. *Electrophoresis* **9**, 830 - 838.

Edman, P. (1949) A method for the determination of the amino acid sequence of peptides. *Arch. Biochem. Biophys.* **22**, 475 - 483.

Edman, P. and Begg, G. (1967) A protein sequenator. *Eur. J. Biochem.* **1**, 80 - 91.

Eng, J.K., McCormack, A.L. and Yates, J.R. (1994) Correlating tandem mass spectral data of peptides to sequences in a protein database. *J. Am. Soc. Mass. Spec.* **5**, 976 - 989.

Evans, T., Rosenthal, E.T., Youngblom, J., Distel, D. and Hunt, T. (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* **33**, 389 - 396.

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

Figeys, D., Ning, Y. and Aebersold, R. (1997) A microfabricated device for rapid protein identification by microelectrospray ion trap mass spectrometry. *Anal. Chem.* **69**, 3153 - 3160.

Figeys, D., Gygi, S.P., Zhang, Y., Watts, J., Gu, M. and Aebersold, R. (1998a) Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes. *Electrophoresis* **19**, 1811 - 1818.

Figeys, D., Gygi, S.P., McKinnon, G. and Aebersold, R. (1998b) An integrated microfluidics-tandem mass spectrometry system for automated protein analysis. *Anal. Chem.* **70**, 3728 - 3734.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influnzae* Rd. *Science* **269**, 496 - 512.

Franzen, B., Iwabuchi, H., Kato, H., Lindholm, J. and Auer, G. (1991) Two-dimensional polyacrylamide gel electrophoresis of human lung cancer: qualitative aspects of tissue preparation in relation to histopathology. *Electrophoresis* **12**, 509 - 515.

Franzen, B., Linder, S., Okuzawa, K., Kato, H. and Auer, G. (1993) Nonenzymatic extraction of cells from clinical tumor material for analysis of gene expression by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **14**, 1045 - 1053.

Franzen, B., Hirano, T., Okuzawa, K., Uryu, K., Alaiya, A.A., Linder, S. and Auer, G. (1995) Sample preparation of human tumors prior to two-dimensional electrophoresis of proteins. *Electrophoresis* **16**, 1087 - 1089.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchinson, C.A. and Venter J.C. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397 - 403.

Gale, D.C. and Smith, R.D. (1993) Small volume and low flow-rate electrospray ionization mass spectrometry of aqueous samples. *Rapid Comm. Mass Spec.* **7**, 1017 - 1021.

Garrels, J.I. (1989) The QUEST system for quantitative analysis of two-dimensional gels. *J. Biol. Chem.* **264**, 5269 - 5282.

Gautier, J., Norbury, C., Lohka, M., Nurse, P. and Maller, J. (1988) Purified maturation-promoting factor contains the product of a Xenopus homolog of the fission yeast cell cycle control gene cdc^{2+}. *Cell* **54**, 433 - 439.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Life with 6,000 genes. *Science* **274**, 546 - 567.

Goldberg, H.A., Domenicucci, C., Pringle, G.A. and Sodek, J. (1988) Mineral-binding proteoglycans of fetal porcine calvarial bone. *J. Biol. Chem.* **263**, 12092 - 12101.

Gevaert, K., Verschelde, J.L., Puype, M., Van Damme, J., Goethals, M., De Boeck, S. and Vandekerckhove, J. (1996) Structural analysis and identification of gel-purified proteins, available in the femtomole range, using a novel computer program for peptide sequence assignment, by matrix-assisted laser desorption ionization-reflection time-of-flight mass spectrometry. *Electrophoresis* **17**, 918 - 924.

Gonnet, G.H. (1994) *A Tutorial Introduction to Computational Biochemistry using Darwin.* ETH Zürich Report, Zürich.

Gooley, A.A., Ou, K., Russell, J., Wilkins, M.R., Sanchez, J.C., Hochstrasser, D.F. and Williams, K.L. (1997) A role for Edman degradation in proteome studies. *Electrophoresis* **18**, 1068 - 1072.

Görg, A., Postel, W. and Günther, S. (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**, 531 - 546.

Görg, A., Boguth, G., Obermaier, C. and Weiss, W. (1998) Two-dimensional electrophoresis of proteins in an immobilized pH 4 - 12 gradient. *Electrophoresis* **19**, 1516 - 1519.

Griffin, P.R., MacCoss, M.J., Eng, J.K., Blevins, R.A., Aaronson, J.S. and Yates, J.R. (1995) Direct database searching with MALDI-PSD spectra of peptides. *Rapid Comm. Mass Spec.* **9**, 1546 - 1551.

Hardy, S.J.S., Kurland, C.G., Voynow, P. and Mora, G. (1969) The ribosomal proteins of *Escherichia coli.* I. Purification of the 30S ribosomal proteins. *Biochemistry* **8**, 2897 - 2905.

Hartwell, L.H. and Kastan, M.B. (1994) Cell cycle control and cancer. *Science* **266**, 1821 - 1828.

Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C. and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5011 - 5015.

Herbert, B.R., Molloy, M.P., Gooley, A.A., Walsh, B.J., Bryson, W.G. and Williams, K.L. (1998) Improved protein solubility in 2-D electrophoresis using tributyl phosphine. *Electrophoresis* **19**, 845 - 851.

Himpens, B., De Smedt, H. and Casteels, R. (1994) Relationship between [Ca^{2+}] changes in nucleus and control. *Cell Calcium* **16**, 239 - 246.

Hindley, J. and Phear, G.A. (1984) Sequence of the cell division gene CDC2 from *Schizosaccharomyces pombe;* patterns of splicing and homology to protein kinases. *Gene* **31**, 129 - 134.

Hochstrasser, D.F. (1997) Clinical and biomedical applications of proteomics. In: Wilkins, M.R., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (eds.) *Proteome research: new frontiers in functional genomics.* Springer-Verlag, Berlin.

Honore, B., Leffers, H., Madsen, P. and Celis, J.E. (1993) Interferon-gamma up-regulates a unique set of proteins in human keratinocytes. Molecular cloning and expression of the cDNA encoding the RGD-sequence containing protein IGUP I-5111. *Eur. J. Biochem.* **218**, 421 - 430.

Hoogland, C., Sanchez, J.C., Tonella, L., Bairoch, A., Hochstrasser, D.F. and Appel, R.D. (1998) Current status of the SWISS-2DPAGE database. *Nucl. Acids Res.* **26**, 332 - 333.

Hughes, J.H., Mack, K. and Hamparian, V.V. (1988) India ink staining of proteins on nylon and hydrophobic membranes. *Anal. Biochem.* **173**, 18 - 25.

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

Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E. and Engelhard, V.H. (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261 - 1263.

Hunt, T. (1991) Cyclins and their partners: from a single idea to complicated reality. *Semin. Cell Biol.* **2**, 213 - 222.

Inglis, A.S., Duncan, M.W., Adams, P. and Tseng, A. (1992) Formation of proline thiohydantoin with ammonium thiocyanate: progress towards a viable C-terminal amino acid sequencing procedure. *J. Biochem. Biophys. Meth.* **25**, 163 - 171.

James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1993) Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Comm.* **195**, 58 - 64.

James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1994) Protein identification in DNA databases by peptide mass fingerprinting. *Protein Sci.* **3**, 1347 - 1350.

James, P. (1997a) Of genomes and proteomes. *Biochem. Biophys. Res. Comm.* **231**, 1 - 6.

James, P. (1997b) Protein identification in the post-genome era: the rapid rise of proteomics. *Q. Rev. Biophys.* **30**, 279 - 331.

Jensen, O.N., Podtelejnikov, V. and Mann, M. (1996) Delayed extraction improves specificity in database searches by matrix assisted laser desorption/ionization peptide maps. *Rapid Comm. Mass Spec.* **10**, 1371 - 1378.

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

Kamp, R.M. and Wittmann-Liebold, B. (1984) Purification of *Escherichia coli* 50S ribosomal proteins by high performance liquid chromatography. *FEBS Lett.* **167**, 59 - 63.

Kamp, R.M., Bosserhoff, A., Kamp, D. and Wittmann-Liebold, B. (1984) Application of high performance liquid chromatographic techniques to the separation of ribosomal proteins of different organisms. *J. Chromatogr.* **317**, 181 - 192.

Kanatani, H., Shirai, H., Nakanishi, K. and Kurosawa, T. (1969) Isolation and identification of a meiosis inducing substance in starfish *Asterina amurensis*. *Nature* **221**, 273 - 274.

Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. and Stevens, T.H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase. *Science* **250**, 651 - 657.

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

Kaufmann, R., Spengler, B. and Lützenkirchen, F. (1993) Mass spectrometric sequencing of linear peptides by product ion analysis using MALDI-TOF-MS. *Rapid Comm. Mass Spec.* **7**, 902 - 910.

Kenner, G.W. and Khorana, H.G. (1952) Peptides part II. Selective degradation by removal of the terminal amino acid bearing a free amino group. The use of alkyl alkoxydithioformates (dialkyl xanthates). *J. Chem. Soc.*, 2076 - 2081.

Kenrick, K.G. and Margolis, J. (1970) Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique. *Anal. Biochem.* **33**, 204 - 207.

Kertesz, M.A., Leisinger, T. and Cook, A.M. (1993) Proteins induced by sulphate limitation in *Escherichia coli, Pseudomonas putida or Staphylococcus aureus*. *J. Bacteriol.* **175**, 1187 - 1190.

King, R.W., Deshales, R.J., Peters, J.M. and Kirschner, M.W. (1996) How proteolysis drives the cell cycle. *Science* **274**, 1652 - 1658.

Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis in mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231 - 243.

Korostensky, C., Staudenmann, W., Dainese, P., Hoving, S., Gonnet, G. and James, P. (1998) An algorithm for the identification of proteins using peptides with ragged N- or C-termini generated by sequential endo- and exopeptidase digestions. *Electrophoresis* **19**, 1933 - 1940.

Kovarova, H., Stulik, J., Macela, A., Lefkovits, I. and Skrabkova, Z (1992) Using two-dimensional gel electrophoresis to study immune response against intracellular bacterial infection. *Electrophoresis* **13**, 741 - 742.

Kriger, M.S., Cook, K.D. and Ramsey, R.S. (1995) Durable gold-coated fused silica capillaries for use in electrospray mass spectrometry. *Anal. Chem.* **67**, 385 - 389.

Kunst, F., Ogasarawa, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, V., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Düsterhöft, A., Ehrlich, S.D., Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249 - 256.

Laemmli, U.K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680 - 685.

Lane, R.D., Allan, D.M. and Mellgren, R.L. (1992) A comparison of the intracellular distribution of μ-calpain, m-calpain and calpastatin in proliferating human A431 cells. *Exp. Cell Res.* **203**, 5 - 16.

Latham, K.E., Garrels, J.I. and Solter, D. (1993) Two-dimensional analysis of protein synthesis. *Meth. Enzym.* **255**, 473 - 489.

Latter, G.I., Metz, E., Burbeck, S. and Leavitt, J. (1983) Measurement of amino acid composition of proteins by computerized microdensitometry of two-dimensional electrophoresis gels. *Electrophoresis* **4**, 122 - 126.

Lemkin, P.F. and Lipkin, L.E. (1981) GELLAB: a computer system for two-dimensional gel electrophoresis. III. Multiple two-dimensional gel analysis. *Comp. Biomed. Res.* **14**, 407 - 446.

Lewin, B. (1990) Driving the cell cycle: M phase kinase, its partners, and substrates. *Cell* **61**, 743 - 752.

Li, K.W., Geraerts, W.P., Van Elk, R. and Koose, J. (1989) Quantification of proteins in the subnanogram and nanogram range: comparison of the AuroDye, FerriDye, and India ink staining methods. *Anal. Biochem.* **182**, 44 - 47.

Linnekin, D., Evans, G., Michiel, D. and Farrar, W.L. (1992) Characterization of a 97-kDa phosphotyrosyl protein regulated by multiple cytokines. *J. Biol. Chem.* **267**, 23993 - 23998.

Lohka, M.L., Hayes, M.K. and Maller, J. (1988) Purification of maturation-promoting factor, an intracellular regulator of mitotic events. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3009 - 3013.

MacLennan, D.H., Rice, W.J. and Green, NM. (1997) The mechanism of Ca^{2+} transport by sarco(endo)plasmic reticulum Ca^{2+}-ATPases. *J. Biol. Chem.* **272**, 28815 - 28818.

Manabe, T., Oda, O. and Okuyama, T. (1982) Amino acid microanalysis of proteins extracted from spots of fixed, stained 2-dimensional gels. *J. Chrom.* **241**, 361 - 370.

Mann, M., Hojrup, P. and Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spec.* **22**, 338 - 345.

Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390 - 4399.

Manz, A. (1997) Ultimate speed and sample volumes in electrophoresis. *Biochem. Soc. Trans.* **25**, 278 - 281.

March, K.L., Wilenskey, R.L., Roeske, R.W. and Hatahaway, D.R. (1993) Effects of thiol protease inhibitor on cell cycle and proliferation of vascular smooth muscle cells in culture. *Circ. Res.* **72**, 413 - 423.

Markwell, M.A., Haas, S.M., Bieber, L.L. and Tolbert, N.E. (1978) Modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal. Biochem.* **87**, 206 - 210.

Matsudaira, P. (1987) Sequence of picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**, 10035 - 10038.

McCormack, A.L., Schieltz, D.M., Goode, B., Yang, S., Barnes, G., Drubin, X. and Yates, J.R. (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**, 767 - 776.

Means, A.R. (1994) Calcium, calmodulin and cell cycle regulation. *FEBS Lett.* **347**, 1 - 4.

Mellgren, R.L., Song, K. and Mericle, M.T. (1993) m-Calpain requires DNA for activity on nuclear proteins at low concentrations. *J. Biol. Chem.* **268**, 653 - 657.

Michaelis, L. and Menten, M.L. (1913) Die Kinetik der Ivertinwerkung. *Biochem. Z.* **49**, 333 - 369.

Miller, P.E. and Denton, M.B. (1986) The quadrupole mass filter: basic operating concepts. *J. Chem. Ed.* **63**, 617 - 622.

Moreau, M., Guerrier, P., Dorée, M. and Ashley, C.C. (1978) Hormone-induced release of intracellular Ca^{2+} triggers meiosis in starfish oocytes. *Nature* **272**, 251 - 253.

Mross, G.A. and Doolittle, R.F. (1977) Step-wise degradation of peptides attached to solid supports. In: Needleman, S.B. (ed.) *Advanced methods in protein sequence determination.* Springer-Verlag, Berlin.

Neylon, C.B., Hoyland, J., Mason, W.T. and Irvine, R.F. (1990) Spatial dynamics of intracellular calcium in agonist-stimulated vascular smooth muscle cells. *Am. J. Physiol.* **259**, C675 - C686.

Nikodem, V. and Fresco, J.R. (1979) Protein fingerprinting by SDS-gel electrophoresis after partial fragmentation with CNBr. *Anal. Biochem.* **97**, 382 - 386.

Norbury, C. and Nurse, P. (1992) Animal cell cycles and their control. *Annu. Rev. Biochem.* **61**, 441 - 470.

Nurse, P. (1990) Universal control mechanism regulating onset of M-phase. *Nature* **344**, 503 - 508.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007 - 4021.

Ortiz, M.L., Calero, M., Fernandez-Patron, C., 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 electroblotted proteins. *FEBS Lett.* **296**, 300 - 304.

Papayannopoulos, I.A. (1995) The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spec. Rev.* **14**, 49 - 73.

Pappin, D., Hojrup, P. and Bleasby, A. (1993) Rapid identification of proteins by peptide mass fingerprinting. *Current Biology* **3**, 327 - 332.

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

Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444 - 2448.

Pederson, S., Bloch, P.L., Reeh, S. and Neidhart, F.C. (1978) Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14**, 179 - 190.

Peitsch, M.C. and Guex, N. (1997) Large-Scale Comparative Protein Modelling. In: Wilkins, M.R., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (eds.) *Proteome research: new frontiers in functional genomics.* Springer-Verlag, Berlin.

Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, J.N., Thorner, J. and Belfort, M. (1994) Protein splicing elements: Inteins and exteins - A definition of terms and recommended nomenclature. *Nucl. Acids Res.* **22**, 1125 - 1127.

Perler, F.B. (1998) Protein splicing of inteins and Hedgehog autoproteolysis: structure, function and evolution. *Cell*, **92**, 1 - 4.

Persidis, A. (1998) High-troughput screening. Advances in robotics and miniturization continue to accelerate drug lead identification. *Nature Biotechnol.* **16**, 488 - 489.

Philipson, K.D., (1985) Sodium-calcium exchange in plasma membrane vesicles. *Annu. Rev. Physiol.* **47**, 561 - 571.

Pines, J. and Hunter, T. (1995) Cyclin-dependent kinases: an embarrassment of riches? In: Hutchison, C. and Glover, D.M. (eds.) *Cell Cycle Control.* IRL Press, New York.

Praxmayer, C., Murach, K.F., Baumgartner, B., Aberger, F., Schlegel, E. and Illmensee, K. (1992) Protein synthesis in murine organs during post-implantation development detected by two-dimensional gel electrophoresis. *Electrophoresis* **13**, 720 - 722.

Quadroni, M., Staudenmann, W., Kertesz, M. and James, P. (1996) Analysis of global responses by protein and peptide fingerprinting of proteins isolated by two-dimensional gel electrophoresis: application to the sulphate starvation response of *Escherichia coli*. *Eur. J. Biochem.* **239**, 773 - 781.

Rabilloud, T., Hubert, M. and Tarroux, P. (1986) Procedures for two-dimensional electrophoretic analysis of nuclear proteins. *J. Chromatogr.* **351**, 77 - 89.

Rabilloud, T., Valette, C. and Lawrence, J.J. (1994) Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis* **15**, 1552 - 1558.

Rabilloud, T. (1996) Solubilization of proteins for electrophoretic analyses. *Electrophoresis* **17**, 813 - 829.

Rabilloud, T., Adessi, C., Giraudel, A. and Lunardi, J. (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **18**, 307 - 316.

Rademaker, G.J. (1996) *Mass spectrometry: a modern approach to solving biological structural problems.* PhD thesis, University of Utrecht, The Netherlands.

Reuter, H. (1984) Ion channels in cardiac cell membranes. *Annu. Rev. Physiol.* **46**, 473 - 484.

Righetti, P.G. (1990) Immobilized pH gradients: theory and methodology. In: Burdon, R.H. and Van Knippenberg, P.H. (eds.) *Laboratory techniques in biochemistry and molecular biology*. Elsevier, Amsterdam.

Roepstorff, P. and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Environ. Mass Spectrom.* **11**, 601 - 603.

Roepstorff, P. (1997) Mass spectrometry in protein studies from genome to function. *Curr. Opin. Biotechnol.* **8**, 6 - 13.

Rosenfeld, J., Capdevielle, J., Guillemot, J.C. and Ferrara, P. (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional electrophoresis gels. *Anal. Biochem.* **203**, 173 - 179.

Sanchez, J.C., Ravier, F., Pasquali, C., Frutiger, S., Paquet, N., Bjellqvist, B., Hochstrasser, D.F. and Hughes, G.J. (1992) Improving the detection of proteins after transfer to polyvinylidene difluoride membranes. *Electrophoresis* **13**, 715 - 717.

Sanchez, J.C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M.R. and Hochstrasser, D.F. (1997) Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. *Electrophoresis* **18**, 324 - 327.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchinson, C.A., Slocombe, P.M. and Smith, M. (1977) Nucleotide sequence of bacteriophage ϕ174 DNA. *Nature* **265**, 687 - 695.

Santella, L. and Kyozuka, K. (1994) Reinitiation of meiosis in starfish oocytes requires an increase in nuclear Ca^{2+}. *Biochem. Biophys. Res. Comm.* **203**, 674 - 680.

Santella, L. (1996) The cell nucleus: an eldorado to future calcium research? *J. Membr. Biol.* **153**, 83 - 92.

Santella, L. and Carafoli, E. (1997) Calcium signalling in the cell nucleus. *FASEB J.* **11**, 1091 - 1109.

Santella, L. and Kyozuka, K. (1997) Effects of 1-methyladenine on nuclear Ca^{2+} transients and meiosis resumption in starfish oocytes are mimicked by the nuclear injection of inositol 1,4,5-triphosphate and cADP-ribose. *Cell Calcium* **22**, 11- 20.

Santella, L. (1998) The role of calcium in the cell cycle: facts and hypotheses. *Biochem. Biophys. Res. Comm.* **244**, 317 - 324.

Santella, L., Kyozuka, K., De Riso, L. and Carafoli, E. (1998) Calcium, protease action, and the regulation of the cell cycle. *Cell Calcium* **23**, 123 - 130.

Schägger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulphate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**, 368 - 379.

Scheele, G.A. (1975) Two-dimensional gel analysis of soluble proteins. *J. Biol. Chem.* **250**, 5375 - 5385.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467 - 470.

Schena, M. (1996) Genome analysis with gene expression microarrays. *BioEssays* **18**, 427 - 431.

Schlack, P. and Kumpf, W. (1926) Über eine neue Methode zur Ermittlung der Konstitution von Peptiden. *Z. Physiol. Chem.* **154**, 125 - 130.

Schollmeyer, J.E. (1988) Calpain II involvement in mitosis. *Science* **240**, 911 - 913.

Shaw, G. (1993) Rapid identification of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5138 - 5142.

Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two-dimensional gels. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14440 - 14445.

Shevchenko, A., Chernushevich, I., Ens, W., Standing, K.G., Thomson, B., Wilm, M. and Mann, M. (1997) Rapid "de novo" peptide sequencing by a combination of nanoelectrospray, isotopic labelling and a quadrupole/time-of-flight mass spectrometer. *Rapid Comm. Mass Spec.* **11**, 1015 - 1024.

Silhavy, T.J., Berman, M.L. and Enquist L.W. (1984) *Experiments with Gene Fusions.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Stark, G.R. (1968) Sequential degradation of peptides from their carboxyl termini with ammonium thiocyanate and acetic anhydride. *Biochemistry* **7**, 1796 - 1807.

Staudenmann, W., Dainese-Hatt, P., Hoving, S., Lehmann, A., Kertesz, M. and James, P. (1998) Sample handling for proteome analysis. *Electrophoresis* **19**, 901 - 908.

Stryer, L. (1988) *Biochemistry (3rd. edition)*. W.H. Freeman and Co., New York.

Tew, K.D., Monks, A., Barone, L., Rosser, D., Akerman, G., Montali, J.A., Wheatly, J.B. and Schmidt Jr., D.E. (1996) Glutathione-associated enzymes in the human cell lines of the National Cancer Institute Drug Screening Program. *Mol. Pharmacol.* **50**, 149 - 159.

Thiede, B., Salnikow, J. and Wittmann-Liebold, B. (1997) C-terminal ladder sequencing by an approach combining chemical degradation with analysis by matrix-assisted-laser-desorption ionization mass spectrometry. *Eur. J. Biochem.* **244**, 750 - 754.

Traub, P., Mizushima, S., Lowry, C.V. and Nomura, M. (1979) Reconstitution of ribosomes from subribosomal components. *Meth. Enzym.* **49**, 391 - 407.

Valaskovic, G.A., Kelleher, N.L. and McLafferty, F.W. (1996) Attomole protein characterization by capillary electrophoresis-mass spectrometry. *Science* **273**, 1199 - 1202.

VanBogelen, R.A., Sanker, P., Clark, R.L., Bogan, J.A. and Neidhardt, F.C. (1992) The gene-protein database of *Escherichia coli*: edition 5. *Electrophoresis* **13**, 1014 - 1054.

Vanfleteren, J.R., Raymackers, J.G., Van Bun, S.M. and Mehus, L.A. (1992) Peptide mapping and microsequencing of proteins separated by SDS-PAGE after limited *in situ* hydrolysis. *BioTechniques* **12**, 550 - 557.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* **270**, 484 - 487.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B. and Kinzler, K.W. (1997) Characterization of the yeast transcriptome. *Cell* **88**, 243 - 251.

Verentchikov, A.N., Ens, W. and Standing, K.G. (1994) Reflecting time-of-flight mass spectrometer with an electrospray source and orthogonal extraction. *Anal. Chem.* **66**, 126 - 133.

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

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.

Wallace, A. and Saluz, H.P. (1992) Beyond silver staining. *Nature* **357**, 608 - 609.

Whitaker, M. and Patel, R. (1990) Calcium and cell cycle control. *Development* **108**, 525 - 542.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F. and Williams, K.L. (1995) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotech. Gen. Eng. Rev.* **13**, 19 - 50.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L. and Hochstrasser, D.F. (1996a) From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analyses. *Bio/Technology* **14**, 61 - 65.

Wilkins, M.R., Gasteiger, E., Sanchez, J.C., Appel, R.D. and Hochstrasser, D.F. (1996b) Protein identification with sequence tags. *Curr. Biol.* **6**, 1543 - 1544.

Wilkins, M.R., Ou, K., Appel, R.D., Sanchez, J.C., Yan, J.X., Golaz, O., Farnsworth, V., Cartier, P., Hochstrasser, D.F., Williams, K.L. and Gooley, A.A. (1996c) Rapid protein identification using N-terminal "sequence tag" and amino acid analysis. *Biochem. Biophys. Res. Comm.* **221**, 609 - 613.

Wilkins, M.R., Gasteiger, E., Tonella, L., Ou, K., Tyler, M., Sanchez, J.C., Gooley, A.A., Walsh, B.J., Bairoch, A., Appel, R.D., Williams, K.L. and Hochstrasser, D.F. (1998) Protein identification with N- and C-terminal sequence tags in proteome projects. *J. Mol. Biol.* **278**, 599 - 608.

Williams, D.A., Becker, P.L. and Fay, F.S. (1987) Regional changes in calcium underlying contraction of single smooth muscle cells. *Science* **235**, 1644 - 1648.

Williams, K.L. and Hochstrasser, D.F. (1997) Introduction to the proteome. In: Wilkins, M.R., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (eds.) *Proteome research: new frontiers in functional genomics.* Springer-Verlag, Berlin.

Wilm, M. and Mann, M. (1996) Analytical properties of the nano electrospray ion source. *Anal. Chem.* **66**, 1 - 8.

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.

Yamawaki, T. and Toyoda, N. (1994) Human chorionic gonadotropin secretion and protein phosphorylation in chorionic tissue. *Endocr. J.* **41**, 509 - 516.

Yates, J.R., Speicher, S., Griffin, P.R. and Hunkapiller, T. (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214**, 397 - 408.

Yates, J.R., Eng, J.K., McCormack, A.L. and Schieltz, D. (1995a) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**, 1426 - 1436.

Yates, J.R., Eng, J.K. and McCormack, A.L. (1995b) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal. Chem.* **67**, 3202 - 3210.

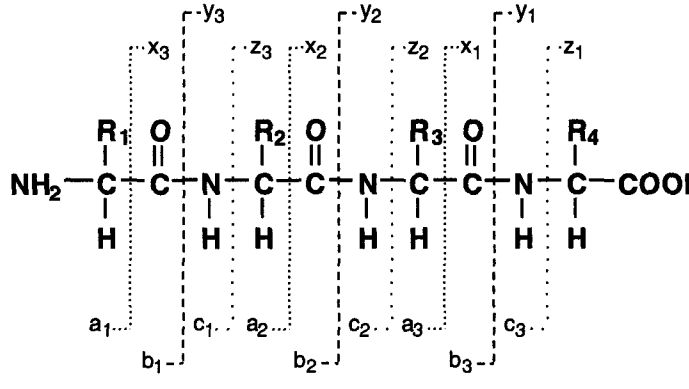
Yates, J.R., Eng, J.K., Clauser, K.R. and Burlinghame, A.L. (1996) Searching sequence databases with uninterpreted high-energy CID spectra of peptides. *J. Am. Soc. Mass Spectrom.* **7**, 1089 - 1098.

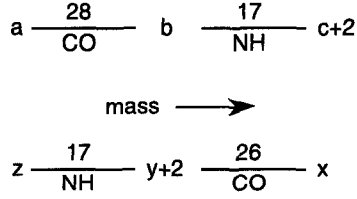
Yates, J.R. (1998) Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **33**, 1 - 19.

Yolken, R., Cervenak, J., Shore, A.D., Torrey, E.F. and Johnston, N. (1997) Multivariate analysis of RNA levels from postmortem human brains as measured by three different methods of RT-PCR. Stanley Neuropathology Consortium. *J. Neurosci. Methods* **77**, 83 - 92.

Zubay, G.L., Parson, W.W. and Vance, D.E. (1995) *Principles of Biochemistry (1st. edition)*. Wm. C. Brown Publisher, Oxford, England.

Appendix 1





Fragmentation nomenclature according to Roepstorff and Fohlman (1984) and Biemann (1988, 1990). The low energy collisions induced by the gas in the TSQ or IT mass spectrometers cause mainly fragmentation along the peptide backbone, with some neutral losses such as water and ammonia.

Here is shown how a peptide can fragment into a series of daughter ions. Fragmentation at the peptide bonds results in b-ions if the positive charge remains at the N-terminal, or in y-ions if the positive charge remains at the C-terminal. The charged portion of the fragmented peptide is detected in the mass spectrometer. The b- and y-ion fragmentation produces a ladder of peptides, the mass difference giving sequence information. Other fragmentation can also occur along the peptide backbone. This generates a- and c-ions if the positive charge remains at the N-terminal, or x- and z-ions if the positive charge remains at the C-terminal.

Appendix 2

TYPE OF PROGRAM	INTERNET URL ADDRESS
peptide mass fingerprinting and sequence tag data	
MassSearch	http://cbrg.inf.ethz.ch/MassSearch
MassDynSearch	http://cbrg.inf.ethz.ch/MassDynSearch
PeptideSearch	http://www. mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html
Sequest	http://thompson.mbt.washington/edu/sequest/
TagIdent	http://www.expasy.ch/www/guess-prot. html
other programs that assist interpretation of analytical data	
MultiIdent	http://www.expasy.ch/sprot/multiident.html
PeptideMass	http://www.expasy.ch/sprot/peptide-mass.html
Compute pI/MW	http://www.expasy.ch/ch2d/pi_tool.html
PropSearch	http://www.embl-heidelberg.de/aaa.html
MS-digest	http://www.ludwig.ucl.ac.uk/msdigest.html

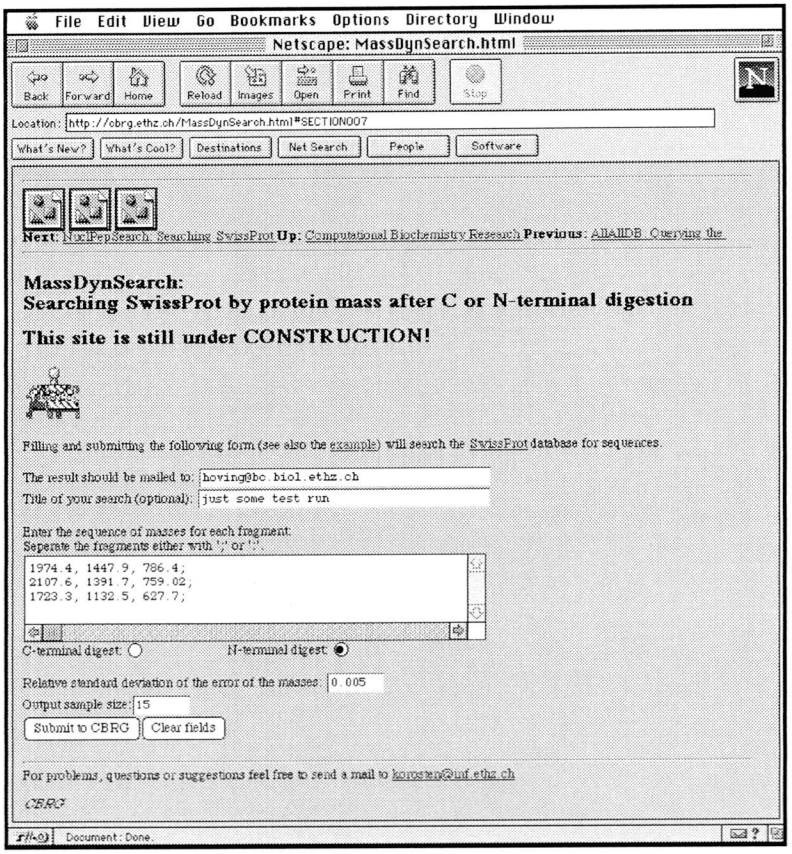
Some protein identification and characterization programs available on the World-Wide Web with their URL addresses are listed here. An up-to-date list is maintained at http://www.expasy.ch/tools.html. An updated list of WWW addresses for protein sequence databases and related programs is maintained at http://www.expasy.ch/users/springer97_table51.html.

Appendix 3

Name residue	*Symbols*		*Monoisotopic Mass*	*Average Mass*
Glycine	Gly	G	57.02146	57.0520
Alanine	Ala	A	71.03711	71.0788
Serine	Ser	S	87.03203	87.0782
Proline	Pro	P	97.05276	97.1167
Valine	Val	V	99.06841	99.1326
Threonine	Thr	T	101.04768	101.1051
Cysteine	Cys	C	103.00919	103.1448
Isoleucine	Ile	I	113.08406	113.1595
Leucine	Leu	L	113.08406	113.1595
Asparagine	Asn	N	114.04293	114.1039
Aspartic Acid	Asp	D	115.02694	115.0886
Glutamine	Gln	Q	128.05858	128.1308
Lysine	Lys	K	128.05858	128.1308
Glutamic Acid	Glu	E	129.04259	129.1155
Methionine	Met	M	131.04049	131.1986
Histidine	His	H	137.05891	137.1412
Phenylalanine	Phe	F	147.06841	147.1766
Arginine	Arg	R	156.10111	156.1876
Tyrosine	Tyr	Y	163.06333	163.1760
Tryptophan	Trp	W	186.07931	186.2133

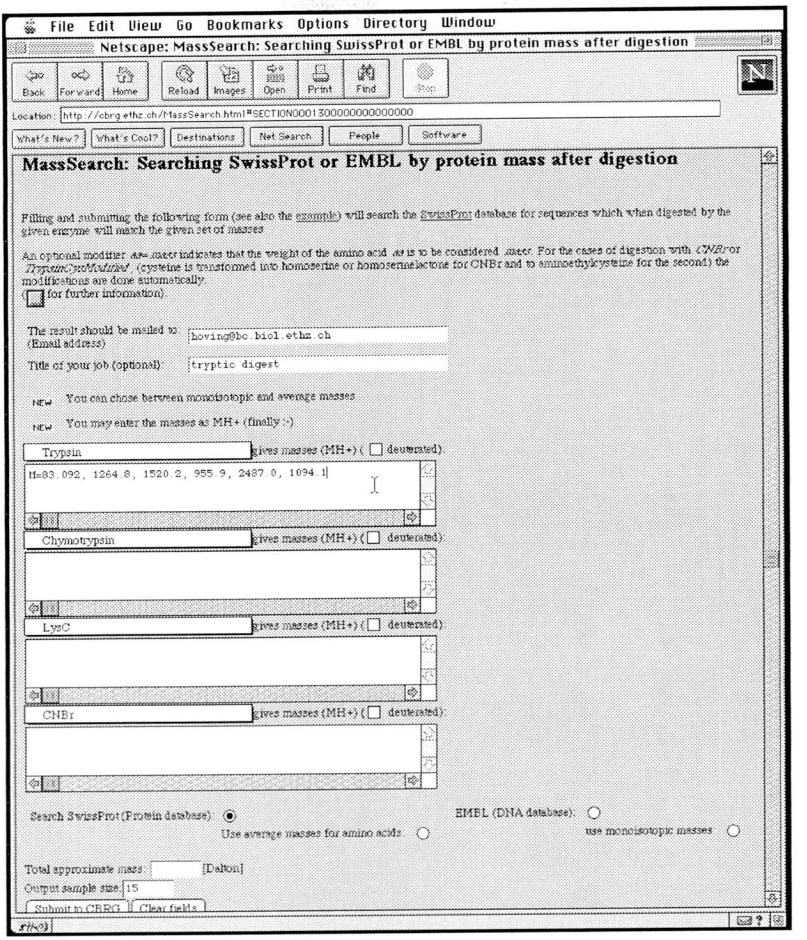
The molecular mass of a normally terminated and unmodified peptide or protein may be calculated by summing the masses of the appropriate amino acid residues from the table and adding the masses of H (1.00782 for monoisotopic, 1.0079 for average mass) and OH (17.00274 for monoisotopic, 17.0073 for average mass) for the N- and C-termini respectively.

Appendix 4



An example of the MassDynSearch page on the WWW, as shown with Netscape Navigator. Note that this site is still under construction. The mass of each intact peptide together with its correspondent masses of the generated tag are entered in the main white box. It is possible to select whether a C-terminal or a N-terminal sequence tag data set is entered. After submitting the data, the search results are send back by electronic mail to the user.

Appendix 5



An example of the MassSearch page on the WWW, as shown with Netscape Navigator. The masses of the peptides of each protein digest are entered in the main white boxes. It is possible to select which protease or which chemical method was used for digestion. The masses can either be entered as monoisotopic or as average masses. After submitting the data, the search results are send back by electronic mail to the user.

Curriculum vitae

Name: Sjouke Hoving

Date of birth: January 9, 1966
Place of birth: Vlagtwedde, The Netherlands
Nationality: Dutch

Education and Research:

1984 - 1990: Hogeschool Drenthe, Emmen, The Netherlands
Higher Laboratory Education
Ing. (B.Sc.) in Chemistry

1989 - 1990: University of Nijmegen, Nijmegen, The Netherlands
Final bench training period under supervision of
Prof. Dr. Jan Joep H.H.M. De Pont with project on
"Effect of free fatty acids and detergents on H,K-ATPase"

1990 - 1993: The Feinberg Graduate School
The Weizmann Institute of Science, Rehovot, Israel
M.Sc. in Life Sciences under supervision of
Prof. Dr. Steven J.D. Karlish with final thesis on
"Characterization of isothiouronium derivatives as high affinity cation analogues for the Na,K-ATPase"

1993 - 1995: The Netherlands Cancer Institute/AvL and
Vrije Universiteit, Amsterdam, The Netherlands
Technical assistant under supervision of
Prof. Dr. Hans V. Westerhoff with project on
"Quantitative analysis of the control and regulation of the glutamine synthetase cascade in *Escherichia coli*"

1995 - 1998: Swiss Federal Institute of Technology,
Zürich, Switzerland
Ph.D. in Natural Sciences under supervision of
Prof. Dr. Ernesto Carafoli with final dissertation on
"A novel method for N-terminal ladder sequencing and subtractive proteome analysis of oocytes from *Asterina pectinifera* after resumption of meiosis"

List of Publications

Korostensky, C., Staudenmann, W., Dainese, P., Hoving, S., Gonnet, G. and James, P. (1998) An algorithm for the identification of proteins using peptides with ragged N- or C-termini generated by sequential endo- and exopeptidase digestions. *Electrophoresis* **19**, 1933 - 1940.

Staudenmann, W., Dainese-Hatt, P., Hoving, S., Lehmann, A., Kertesz, M. and James, P. (1998) Sample handling for proteome analysis. *Electrophoresis* **19**, 901 - 908.

Van Heeswijk, W.C., Hoving, S., Molenaar, D., Stegeman, B., Kahn, D. and Westerhoff, H.V. (1996) An alternative P_{II} protein in the regulation of glutamine synthetase in *Escherichia coli*. *Mol. Microbiol.* **21**, 133 - 146.

Hoving, S., Bar-Shimon, M., Tijmes, J.J., Goldshleger, R., Tal , D.M. and Karlish, S.J.D. (1995) Novel aromatic isothiouronium derivatives which act as high affinity competitive antagonists of alkali metal cations on Na,K-ATPase. *J. Biol. Chem.* **270**, 29788 - 29793.

Van Heeswijk, W.C., Stegeman, B., Hoving, S., Molenaar, D., Kahn, D. and Westerhoff, H.V. (1995) An additional P_{II} in *Escherichia coli:* a new regulatory protein in the glutamine synthetase cascade. *FEMS Microbiol. Lett.* **132**, 153 - 157.

Wijker, J.E., Jensen, P.R., Snoep, J.L., Vaz Gomes, A., Guiral, M., Jongsma, A.J.P., de Waal, A., Hoving, S., van Dooren, S., van der Weijden, C.C., van Workum, M., van Heeswijk, W.C., Molenaar, D., Wielinga, P., Richard, P., Diderich, J., Bakker, B.M., Teusink, B., Hemker, M., Rohwer, J.M., van der Gugten, A.A., Kholodenko, B.N. and Westerhoff, H.V. (1995) Energy, control and DNA structure in the living cell. *Biophys. Chem.* **55**, 153 - 165.

Goldshleger, R., Tal, D.M., Shainskaya, A., Or, E., Hoving, S. and Karlish, S.J.D. (1994) Organization of the cation binding domain of the Na,K-pump. In: Bamberg E. and Schoner W. (eds.) *The Sodium Pump: structure mechanism, hormonal control and its role in disease.* Springer-Verlag, New York.

Karlish, S.J.D., Goldshleger, R., Tal, D.M., Capasso, J.M., Hoving, S. and Stein, W.D. (1992) Identification of the cation binding domain of Na,K-ATPase. *Acta Physiol. Scand.* **146**, 69 - 76.

Capasso, J.M., Hoving, S., Tal, D.M., Goldshleger, R. and Karlish, S.J.D. (1992) Extensive digestion of Na,K-ATPase by specific and non-specific proteases with preservation of cation occlusion sites. *J. Biol. Chem.* **267**, 1150 - 1158.

Swarts, H.G.P., Van Uem, T.J.F., Hoving, S., Fransen, J.A.M. and De Pont, J.J.H.H.M. (1991) Effect of free fatty acids and detergents on H,K-ATPase. The steady-state ATP phosphorylation level and the orientation of the enzyme in membrane preparations. *Biochim. Biophys. Acta* **1070**, 283 - 292.

Acknowledgements

The last part of the thesis is usually not the easiest to write. After three years of "Living and working in Switzerland" (see for more details a book by David Hampshire) I would like to thank a lot of people for their kindness and hospitality.

First of all, I would like to thank Prof. Ernesto Carafoli for acting as examiner of my thesis work. Especially, the discussions and comments on Part B of the thesis were unforgettable.

I am grateful to Prof. Josef Brunner for acting as co-examiner. Also, I had the privilege to do some chemistry with the famous TID cross-linkers in your laboratory.

Of course, without Dr. Peter James this work would never have been finished. Thank you, Peter, for everything you did during my stay in the ETH. It was also you who persuaded me to join the group in the protein chemistry laboratory and who decided to change the project drastically after the first year.

Further, I want to mention the colleagues in the A-floor. It is a fact that a lot of people come and go in this lab. I am grateful to Werner Staudenmann for all the discussions, not only about the work, but also about Italy, computer games and much more. Almost everything I know about the mass spectrometer I learned from you. Paola Dainese, later extended to Dainese-Hatt, thank you for always creating a enjoyable atmosphere in the lab, for the help on the oocyte project and the experience of an Italian wedding. I thank Martin Münchbach for many helpful suggestions on the work and the translation of the summary in German (now it is your turn on the C-terminal) and Manfredo Quadroni for help with the MS/MS measurements. Chantal Corti, Estelle Leclerc-L'Hostis, Luca Signor (not only for NMR measurements and translation of the summary in Italian), Carlotta Zamparelli, Suzanna Martinovic and Alla Shainskaya all were present at some stage and contributed to the nice atmosphere in the A-floor, thank you.

I want to thank Tony Lehmann for the construction of the apparatus shown in figure 2.5 and for all the quick technical assistance that was needed when tubes or valves were broken due to the aggressive TFA.

Dr. Luigia Santella from the Stazione Zoologica "Anton Dohrn" in Naples (Italy) is acknowledged for the preparation of the oocyte samples.

In the beginning of my thesis work, three weeks were spent in the Umeą University (Sweden) in the laboratory of Dr. Peter Brodin. It is not due to your excellent guidance in the world of molecular biology that none of that work appeared in this thesis. But the time spend in Umeą with temperatures reaching -20 °C at night was interesting and educative.

Furthermore, I would like to thank all the people in the Institute of Biochemistry, especially Silvio Calderara, Lynda Dowse (for administrative help), Jim Dyer, Paolo Gazzotti, Armando Genazzani, Danilo Guerini, Fabrizio Guidi, Mark Hilge, Sasha Kraev, Joachim Krebs, Lei Li, Marcel Menz, Rolf Moser (for computer related problems), Timo Scheller, Holger Schmid, Sonia Vadrucci, Xiao-Quan Wang and Alessia Zecca.

I also would like to thank Rob Klompmaker (mountain-biking in Wallis), Carola Knuppe, Annette Langedijk (Corso di Italiano), Edwin Meijerink, Janet Storm and Ana Vaz Gomes for their friendship and for the good times we always have, either here in Zürich or elsewhere.

At last I would like to thank my family for all the support and confidence in me all these years.

Thank you all.....!!!

Zürich, December 1998

Sjouke