108

THE PRIMARY STRUCTURES AND THE EVOLUTIONARY CONSEQUENCES OF THE OLISTHODISCUS LUTEUS HISTONE PROTEINS

BY

ANTHONY SPIT

Submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy
in the
Faculty of Science,
University of Cape Town
May, 1993

ACKNOWLEDGEMENTS

I am grateful to the following people whom have assisted and supported me during the preparation of this work:

Assoc Prof Wolf Brandt, my supervisor, for his guidance during all phases of this project.

Prof Claus von Holt and Assoc Prof Trevor Sewell for their interest and encouragement, especially at the beginning of my studies.

Dr Jerry Rodrigues for is help and ideas.

Morvin Graham for her help with the electron micrographs.

Chloe Cheung Chak for her support during the last four years and for her help in preparing the manuscript.

Linda Cilliers for dealing with maverick commas while proof-reading the manuscript.

My Parents for their support and encouragement -as well as finanancial assistance.

My friends at Pace Dance Company and at Pancho's Mexican Kitchen, including Neil, Johann, Linda, Alan, Val, John and Jean. Also Stuart - I'm sorry there aren't more jokes.

CERTIFICATION OF SUPERVISOR

In terms of paragraph 8 of "General regulations for the degree of Ph.D", I, as supervisor of the candidate, A.F. Spit, certify that I approve of the incorporation in this thesis of material that has been submitted for publication.

Signed by candidate

WOLF BRANDT
ASSOCIATE PROFESSOR

TO CHLOE

ABBREVIATIONS

Ac	acetylated
ATZ	2-alkoxyl-5-thiazolinone
bp	base pairs
C- terminal	carboxy terminal
CnBr	cyanogen bromide
CTAB	cetyltrimethylammonium bromide
DMSO	dimethyl sulfoxide
DPTU	diphenylthiourea
ED	endoproteinase aspartic -N digestion
EDTA	ethylenediamine tetra-acetic acid
EGTA	ethylene glycol-bis-(ß-aminoethyl ether)-N,N,N',N'-tetraacetic acid
EK	endoproteinase lysine -C digestion
ER	endoproteinase arginine -C digestion
HF	heptafluorobutyric acid cleavage
HFBA	heptafluorobutyric acid
H*Ol*	Histone *Olisthodiscus luteus*
HPLC	high performance liquid chromotography
MDM	mutational difference matrix
N- terminal	amino terminal
O.luteus	*Olisthodiscus luteus*
PAGE	Polyacrylamide gel electrophoresis
pCMB	p-chloromercuribenzoate
PES	Provasoli's enriched sea water
PITC	phenylisothiocyanate
PMSF	phenylmethylsulfonyl fluoride
PTH	3-phenyl-2-thiohydantion
RuBisco	ribulose-1,5-biphosphate carboxylase
SDS	sodium dodecyl sulphate
TDG	thiodiglycol
TEMED	N,N,N',N'-tetramethylene diamine
TFA	trifluoroacetic acid
TLCK	N-α-tosyl-L-lyslchloromethylketone hydrochloride
TMA	trimethylamine
TPCK	N-tosyl phenylalanylchloromethane

Tris	Tris(hydroxymethyl) aminomethane
UEP	unit evolutionary period
V8	endoproteinase digestion with *Staphylococcus aureus* V8 at pH 4.4
V8H	endoproteinase digestion with *Staphylococcus aureus* V8 at pH 7.0

SUMMARY

During the course of this study, the histones of the algae *Olisthodiscus luteus* were isolated, purified and fractionated. Identification of the histones was achieved by partial primary structure analysis. The histones H1, H2A, H2B, H3 and H4 were found to be present in the *O.luteus* nucleus. The complete structure of H2A and H4 was determined. There is no evidence of the existence of the unique histone *H01* (Rizzo *et al.*, 1985). Construction of phylogenetic trees suggests that the alga *Olisthodiscus luteus* diverged from the animal line. By sequence comparison, the most closely related histone sequence to the algae was found to be that of the echinodermata. An endosymbiotic event between an echinodermata ancestor and a primitive unicellular alga is hypothesised in an attempt to explain the smilarity between the histones.

CONTENTS

Acknowledgements		i
Certificaton of supervisor		ii
Abbreviations		v
Summary		vii

CHAPTER 1

	INTRODUCTION	1
1.1	Histones and the early Eukaryotes	1
1.2	Scope of this Thesis	7
1.3	Olisthodiscus luteus	8
1.3.1	Morphology	8
1.3.2	Classification of *Olisthodiscus luteus*	10
1.3.3	*O.luteus* Cell Cycle	11
1.3.4	*O.luteus* Chromatin Structure	11
1.3.5	*Olisthodiscus luteus* Chloroplasts	12
1.4	Protein Primary Structure Determination	14

CHAPTER 2

	RESULTS and DISCUSSION	
2.1	Culturing of Olisthodiscus luteus	17
2.2	Nuclei Isolation	18
2.3	Analysis of Nuclear Proteins	20
2.4	Histone Extraction	23
2.4.1	Sodium Chloride Histone Extraction	23
2.4.2	Sulphuric Acid Histone Extraction	23
2.4.3	Hydrochloric Acid Histone Extraction	24
2.4.4	Protamine Histone Displacement	27
2.4.5	Conclusions from Protein Extraction Protocols	31
2.5	Proteolysis Inhibitor Assay	33
2.6	Fractionation of Olisthodiscus luteus Basic Proteins	36
2.6.1	Column Chromatography on a Bio-Gel P60 Column	36

2.6.2	Fractionation of *O.luteus* Nuclear Proteins by HPLC	40
2.7	**Elucidation of the Partial Primary Structures of Proteins A1 and A2**	44
2.7.1	Electroelution of Proteins A1 and A2	44
2.7.2	Partial Primary Structure Determination of Protein A1 and Protein A2	44
2.8	**Elucidation of the Partial Primary Structure of Protein B**	47
2.8.1	Peptide Production from Protein B	47
2.8.2	Partial Primary Protein Structure of Protein B	49
2.9	**Elucidation of the Partial Primary Structure of Protein C1**	51
2.9.1	Peptide Production from Protein C1	51
2.9.2	Primary Structure Determination of Peptides generated from Protein C1	51
2.10	**Elucidation of the Complete Primary Structure of Protein C2**	55
2.10.1.1	Digestion of Protein C2 with *Staphylococcus aureus* V8	55
2.10.1.2	Digestion of Protein C2 with Endoproteinase Arg-C	55
2.10.1.3	Digestion of Protein C2 with Endoproteinase Asp-N	55
2.10.1.4	Chemical Cleavage of Protein C2 with Cyanogen Bromide	60
2.10.2	**Determination of the Complete Primary Structure of Protein C2**	60
2.11.	**Elucidation of the Complete Structure of Protein D**	64
2.11.1	Production of Peptides from Protein D	64
2.11.1.1	Digestion of Protein D with *Staphylococcus aureus* V8	64
2.11.1.2	Separation of Peptide D.V8.4	64
2.11.1.3	Fragmentation of Peptide D.V8.3	64
2.11.1.4	Chemical Cleavage of Protein D with Heptafluorobutyric Acid	70
2.11.1.5	Digestion of Protein D with Endo-Lys Protease	70
2.11.2	The Complete Primary Structure of Protein D	70
2.12	**Table of Sequence Results**	75

CHAPTER 3

	CONCLUSIONS and DISCUSSIONS	101
3.1	**Genetic Computer Group Multiple Sequence Analysis**	101
3.1. 1	Data Banks	101
3.1.2	Multiple Sequence Analysis	101

3.2	Olisthodiscus luteus Histone H1	104
3.3	Olisthodiscus luteus Histone H2A	112
3.4	Olisthodiscus luteus Histone H2B	117
3.5	Olisthodiscus luteus Histone H3	124
3.6	Olisthodiscus luteus Histone H4	129
3.7	Evolution of the Olisthodiscus luteus Histones	134

CHAPTER 4

	MATERIALS and METHODS	154
4.1	Solutions and Buffers	154
4.1.1	Buffers	154
4.1.2	Enzymes	155
4.2	Isolation of Olisthodiscus luteus Basic Nuclear Proteins	155
4.2.1	Culturing of *Olisthodiscus luteus*	155
4.2.2	Cell Harvest	156
4.2.3	*Olisthodiscus luteus* Nuclei Isolation	158
4.3	Basic Nuclear Protein Extraction	159
4.3.1	2 M NaCl Extraction	159
4.3.2	Hydrochloric Acid Extraction	159
4.3.3	Sulphuric Acid Extraction	160
4.3.4	Protamine Displacement of Basic Nuclear Proteins	160
4.4	Analysis of Extracted Nuclear Proteins by SDS-PAGE	161
4.4.1	Solutions for SDS-PAGE and Gel Casting	161
4.4.2	Sample Preparation	162
4.4.3	Electrophoresis	162
4.4.4	Staining and Destaining of Gels	162
4.5	Electroelution of Electrophoresed Proteins	162
4.6	Protease Inhibitor Assay	164
4.6.1	Protease Inhibitor Assay Stock Solutions	164
4.6.2	The Protease Inhibitor Assay	164
4.7	Fractionation of Extracted Olisthodiscus luteus Nuclear Proteins	167
4.7.1	Molecular Exclusion Chromatography in the presence of NaCl	167
4.7.2	Reverse Phase High Performance Liquid Chromatography (HPLC)	167
4.7.2.1	HPLC Apparatus	168

3.2	Olisthodiscus luteus Histone H1	104
3.3	Olisthodiscus luteus Histone H2A	112
3.4	Olisthodiscus luteus Histone H2B	117
3.5	Olisthodiscus luteus Histone H3	124
3.6	Olisthodiscus luteus Histone H4	129
3.7	Evolution of the Olisthodiscus luteus Histones	134

CHAPTER 4

	MATERIALS and METHODS	154
4.1	Solutions and Buffers	154
4.1.1	Buffers	154
4.1.2	Enzymes	155
4.2	Isolation of Olisthodiscus luteus Basic Nuclear Proteins	155
4.2.1	Culturing of *Olisthodiscus luteus*	155
4.2.2	Cell Harvest	156
4.2.3	*Olisthodiscus luteus* Nuclei Isolation	158
4.3	Basic Nuclear Protein Extraction	159
4.3.1	2 M NaCl Extraction	159
4.3.2	Hydrochloric Acid Extraction	159
4.3.3	Sulphuric Acid Extraction	160
4.3.4	Protamine Displacement of Basic Nuclear Proteins	160
4.4	Analysis of Extracted Nuclear Proteins by SDS-PAGE	161
4.4.1	Solutions for SDS-PAGE and Gel Casting	161
4.4.2	Sample Preparation	162
4.4.3	Electrophoresis	162
4.4.4	Staining and Destaining of Gels	162
4.5	Electroelution of Electrophoresed Proteins	162
4.6	Protease Inhibitor Assay	164
4.6.1	Protease Inhibitor Assay Stock Solutions	164
4.6.2	The Protease Inhibitor Assay	164
4.7	Fractionation of Extracted Olisthodiscus luteus Nuclear Proteins	167
4.7.1	Molecular Exclusion Chromatography in the presence of NaCl	167
4.7.2	Reverse Phase High Performance Liquid Chromatography (HPLC)	167
4.7.2.1	HPLC Apparatus	168

4.7.2.2	HPLC Column	168
4.7.2.3	Solvent Preparation	169
4.7.2.4	Solvent Gradient Programme	169
4.7.2.5	Sample Preparation	169
4.8	**Production of Peptides**	170
4.8.1	Enzyme Cleavages of Proteins	170
4.8.1.1	*Staphylococcus aureus* V8 Protease Digestion	170
4.8.1.2	Endoproteinase Asp-N Digestion	170
4.8.1.3	Endoproteinase Arg-C Digestion	171
4.8.1.4	Endoproteinase Lys-C Digestion	171
4.8.2	Chemical cleavage of Proteins	171
4.8.2.1	Cyanogen Bromide	171
4.8.2.2	Heptafluorobutyric Acid	172
4.9	**Peptide Nomenclature**	172
4.10	**Sequence Analysis**	174
4.10.1	Sequence Reagents and Solvents	174
4.10.2	Sample Preparation	175
4.10.2.1	Preparation of Glass Fiber Discs and Application of Protein Carrier	175
4.10.2.2	Sample Application	176
4.10.3	Manual Gas Phase Sequencing	176
4.10.3.1	Coupling	177
4.10.3.2	Cleavage	177
4.10.3.3	Conversion	178
4.10.4	Automated Gas Phase Sequencing	179
4.10.4.1	The Gas Phase Sequencer	179
4.10.4.2	The Sequencer Programme	179
4.10.5	Identification and Quantification of Phenylthiohydantion Amino Acids	184
4.10.5.1	"Off-Line" Identification and Quantification of PTH-Amino Acids by Reverse Phase HPLC	184
4.10.5.2	"On-Line" Identification of PTH-Amino Acids	185

REFERENCES 188

CHAPTER 1

INTRODUCTION

1.1 Histones and the Early Eukaryotes

The manner by which genetic material is organized marks a major difference between the eukaryotes and the prokaryotes. The large amount of DNA found in eukaryotes (often 1 - 2 m long) has to be compacted by a factor 10^4 to be accommodated in a nucleus of about 5 μm diameter (von Holt, 1985). Much of the knowledge of the mechanism by which this compaction is achieved has only been established in the last 15 years. Today, it is well accepted that the first two orders of DNA compaction depends on a family of low molecular weight, basic DNA binding proteins called the histones. The histones are in an equal mass ratio to the DNA (for review see Sperling and Wachtel, 1981). Although discovered in 1884 by A. Kossel, it has taken almost 100 years to partially understand the histones' structure and function - the complexity of which perhaps prompted the rueful comment: "The histones are commonly regarded as unpleasant proteins for vigorous studies" of Luck *et al.* in 1956.

Five major histone types are found in the nucleus of all multicellular eukaryotic organisms. The nucleosomal, or core histones, H2A, H2B, H3 and H4, are present in the nucleus in equimolar amounts. The fifth type of histone, H1, is usually present in half the molar amount of each of the core histones. All these histones share some general characteristics. They are small proteins varying from between 100 - 200 amino acids in length. The histone H1 is the largest and depending on the source, is made up of about 200 amino acids. The histones are all positively charged at physiological pH, due to a richness of lysine and/or arginine residues. The charged residues are generally arranged at the N- and C- terminals of the protein leaving the middle region, or globular domain, hydrophobic. No doubt, the positioning of the charged groups at the protein terminals facilitates binding of the protein to the DNA phosphates. No histone is known to contain tryptophan, while tyrosine and phenylalanine contents are low.

Many amino acid sequences have been determined (eg EMBL Data Bank) either by gene sequencing or by direct protein sequencing. Comparison of such sequences reveal the highly conserved nature of the histones.

The "arginine-rich" histone H4 is the most conserved of all the histones. This may be best illustrated by comparing primary structures of H4 isolated from calf (De Lange *et al.*, 1969a; Ogawa *et al.*, 1969) and that isolated from the morphologically different yeast (Smith and Andresson, 1983). These two sequences differ by only eight amino acids. Sequence comparison of H3 structures also indicate that this histone is conserved, although not to the same degree as that of H4. There are fifteen differences between the H3 of calf (De Lange *et al.*, 1972; Patthy and Smith, 1975; Franklin and Zweidler, 1977) and that of yeast (Brandt and von Holt, 1982). The similarity between the H3 and H4 structures isolated from two organisms that diverged early on in the history of life (Brandt and von Holt, 1982) indicates the evolutionary pressure to maintain the structural integrity of the nucleosome and the importance of the role the histones performs in packing DNA of the eukaryotic cell.

The histones H2A and H2B are not as conserved as the H3 and H4. As observed by von Holt (1985), these two proteins are characterized variations in both composition and length of the N- and/or C- terminal extensions. For H2A type histones, both terminals are found to be variable, while only the N- terminal of the H2B proteins show a low evolutionary stability. The domain structure of both H2A and H2B is generally conserved despite the variability in sequence and length of the terminals.

The lysine rich H1 histone is the most variable of all the histones. Three domains, the N- terminal, C- terminal and the internal domain, may however be distinguished when primary sequences are analysed (Sperling and Wachtel, 1981). The general characteristics of these domains are conserved although the complete primary structure may not be. The N- terminal is rich in basic residues, as well as alanine, proline and serine. The C- terminal is rich in lysine and alanine residues, with a small amount of proline (von Holt *et al.*, 1979). The H1 proteins vary not only from species to species, but also from cell to cell within a species and even within a cell.

An examination of all sequences from the five histone families shows that variability does not occur randomly throughout the proteins. Variation appears to be localized at the N- terminal and in some cases to the C- terminal with the exception of H3. Variability in H3 is found in the internal domain. Changes that occur in the "core" or globular domain of H3, and in some cases H4, are usually conservative point mutations. The terminal regions of H2A and H2B show extensive modification through reiteration, insertions and deletions as well as point mutations. As suggested by von Holt *et al.*, (1979), the conservative evolution of H3, H4 and the conserved domains of H2A and H2B are probably responsible for the structural integrity of the histone complex. The variability in sequence and of length of the H2A and H2B terminals, as well as that of the H1 histone, may reflect differences of DNA compaction in the active chromatin.

From data presented by Wilson *et al.*, (1977), Isenberg (1978) determined percent sequence differences of various proteins as a function of divergence time. As discussed above, the histone H4 is clearly the most conserved protein of all proteins analysed - requiring 400 X 10^6 years for a 1 % difference in amino acid sequence to arise between two lineages. H3 is almost as conserved, requiring 330 x 10^6. On the same scale H2A and H2B require 60 X 10^6 years and H1 8 X 10^6 years for a 1 % difference to occur. When these rates of change are compared with other proteins, the next most stable protein known is glutamate dehydogenase (55 X 10^6 years for a 1 % change), slightly more variable than H2A and H2B. Cytochrome C has an evolutionary period of 15 X 10^6 years for a 1 % change. H1, although the most variable of the histones, is less variable than myoglobin, hemoglobin (α), and hemoglobin (ß) (6,0, 3,7 and 3,3 X 10^6 years respectively).

Their evolutionary stability, and their presence in all eukaryotic cells (with the apparent exception of a few protistans), suggests that a knowledge of histone primary structures may be employed to construct a phylogeny tree. This approach has been successfully used to determine evolutionary relationships by comparing primary structures of cytochrome C, isolated from evolutionary diverse organisms (Dickerson, 1971). Classification of higher organisms reflect their evolutionary relationships. A knowledge of the position on the phylogeny tree, deduced from conserved proteins' primary structures, may assist classification of troublesome organisms such as the algae (see example: Hara *et al.*, 1985).

It is widely accepted that some organelles originated with the endosymbiosis of eubacteria (Gray *et al.*, 1984). Recently it has been shown that *Cryptomonad* algae are chimaeras of two distinct unicellular eukaryotes (Douglas *et al.*,1991). These authors have isolated and sequenced two funtional 18s RNA molecules from the *Cryptommonad*. It was found that the nuclear DNA yielded only one of them. They propose that the nucleomorph yielded the other. By sequence comparison it was found that the nuclear rRNA is most closely related to the fungi *Acanthamoeba* while the nucleomorph rRNA is more related to red algae.

A knowledge of the amino acid sequence of conserved proteins coded for by the nucleus, such as the histones, or coded for by the mitochondrial or chloroplast DNA, might also indicate the nature of the organism prior to the endosymbiotic event. Furthermore, the nature of the symbiont and the period when the endosymbiotic event occurred could possibly be deduced.

To achieve a detailed phylogenetic tree for histone-coding eukaryotes by comparison of histone sequences, requires a wide range of sequences. A scan of known sequences utilizing the GCG program package and the GenEMBL databank, indicates that the majority of solved histone primary structures are those of the higher animals and plants. Within their kingdoms, most of these organisms are closely related in evolution - the "twigs" of the phylogenetical tree. To determine the "branches" of these trees, sequences of early diverging organisms, from fungi, low animals and low plants are required.

Histones have been reported in several species of the algae, fungi, ciliated protozoans and slime molds (Horgen and Silver, 1978; Rizzo, 1985) but very few have been sequenced. This is probably due to the difficulty of extracting the basic proteins and the low yields of protein after purification.

Histones from various species of yeast have been well characterized. Brandt and von Holt (1982) determined the primary structure of H3 isolated from *Saccaromyces cerevisiae* by direct protein sequencing. Wallis *et al.* (1980) demonstrated the presence of two different H2B genes by DNA sequencing. The deduced amino acid sequence showed four differences between them. These authors propose that the origin of the two H2B's was by gene duplication some 190 million years ago. Similarly, Choe *et al.* (1982) has shown that yeast has two H2A genes, coding for

two H2A subtypes, with two differences between them. Smith and Andresson (1983) went on to determine that there are two loci coding for both H3 and H4. The deduced amino acid sequence from the DNA sequence indicates that the H3 and H4 are identical in both loci.

Other fungal histone structures solved include the H2A of *Aspergillus nidulans* by May and Morris (1987). The protein sequence was deduced by sequencing of the gene. Unlike most animal histone genes, there are three introns present. Also different too all known H2A sequences, the first N- terminal residue is threonine as opposed to serine.

Woudt *et al.* (1983) have sequenced the H3 and H4 genes of the fungi *Neurospora crassa*. The organism contained one copy of each gene in its genome. The deduced sequence of the H3 and H4 proteins shows it to be similar to that of yeast.

The Trypanosomatid *Leishmania enrietta*, was found to contain H2B and its structure was solved by isolation of the H2B mRNA (Genske *et al.*, 1990). The structure of H2A of a similar strain, *Leishmania donovani infantum*, was solved by Soto *et al.* (1991). The mRNA for the H2A was found to be plant-like, in that it is polyadenylated and has stem-loops at the 3' terminus.

The histones of the protozoan ciliate, *Tetrahymena*, has also been well studied. The protein sequence of two H2A variants was determined by Fusauchi and Iwai (1983). Bannon *et al.* (1984) sequenced two H4 genes of *Tetrahymena thermophila's* macronucleus. Although the deduced amino acid sequence is identical in both genes, heterogeneity was found in the flanking DNA sequences - one H4 gene being closely linked to the H3 gene. Two H4 variants' primary structures were determined by protein sequencing (Fusauchi and Iwai, 1983), indicating the presence of another H4 gene not detected by Bannon *et al.*. The full sequence of two H3 variants was determined by Hayashi *et al.* (1984) by direct protein sequencing. The partial DNA sequence was determined for two H3 genes by Horowitz and Gorovsky (1985). They observed that the stop codon of higher eukaryotes, TAA, codes for glutamine in the *Tetrahymena*. The authors proposed that this ciliate diverged prior to the branching of fungi, plants and animal. The DNA sequence of two H2B genes (Nomoto *et al.*, 1987) shows that there is a three amino acid heterogeneity between the two variants. Wu *et al.* (1986) solved the structure of

Terahymena H1 gene. The deduced amino acid sequence shows the H1 protein to be small, rich in basic amino acids and, unlike higher eukaryotic H1s, missing the usually conserved globular domain.

The green algae *Volvox carteri* was found to be phylogenetically more closely linked to higher plants than to animals after their respective rRNA's had been compared (Rausch *et al.*, 1989). Müller and Schmitt (1988) sequenced two loci containing the H3-H4 genes of *Volvox carteri*. The coding regions for both were identical. Both loci also contained an intron in the H3 gene, shifted by one base pair relative to each other. This exon-intron organization of the H3 gene is also found in yeast H3, fungal H3 and in vertebrate H3.3 genes. The authors observed that the termination signal of the H3 gene is typical of that found in animal H3.1 genes.

Two non-allelic H2A-H2B gene structures of the alga was solved by Müller *et al.* (1990). The authors found one difference in the encoded protein for H2A and 16 for the H2B. It was also found that, unlike higher plants, the histone mRNA are non-polyadenylated. Furthermore, the mRNAs are terminated at the 3' palindrome by the same mechanism that operates in vertebrates and sea urchin. The authors suggest that *Volvox* histone genes are more closely related to the replication-dependant histone genes of animals. This is contrary to what is expected, if Rausch *et al.'s.* (1989) findings are taken into consideration. Müller and Schmitt (1988) acknowledge that the understanding of *Volvox* histone evolution depends on a greater knowledge of histones of the green algae.

1.2 Scope of this Thesis

Several types of algae have been shown to contain histone or histone-like proteins (Rizzo, 1985). It has been reported that one algal species, *Olisthodiscus luteus*, contains the histones H1, H3 and H4 (Rizzo *et al.*, 1985). No H2A or H2B proteins were detected. The authors claim to have detected a unique histone which they termed the H*0l*. H*0l* was proposed as the ancestral to the H2A and H2B histones, giving rise to these proteins by a gene duplication mechanism. As histone H*0l* would be fulfilling the roles of both H2A and H2B, a knowledge of its primary structure would be extremely useful in understanding the structure/function relationship of the H2A and H2Bs of higher eukaryotes. Furthermore, the primary structures of these algal histones would add to the pool of histone sequences known. This would possibly allow an evolutionary tree to be constructed, thus providing useful insight into the evolution of all the histones. The classification of *O.luteus* is still under debate (to be discussed). A knowledge of the conserved histone structures may indicate the Division and Order in which it truly belongs.

The scope of this thesis is therefore to isolate the histones of the *Olisthodiscus luteus* and to identify them by partial protein sequence analysis. The H*0l* (if present) is to be isolated, purified and the full primary structure determined. A sequence comparison of known histone sequences and those of *O.luteus* determined here would then be undertaken.

1.3 *Olisthodiscus luteus*

Olisthodiscus luteus (*O.luteus*) was first described by Carter (1937) after isolating the organism at Brembridge, Island of Wight, England. She assigned it to the Class Xanthophyceae of the Division Chrysophycophyta.

1.3.1 Morphology

The morphology and ultrastructure was fully documented by Hara *et al.* (1985). The *O.luteus* is a yellow-green marine alga that has no cell wall. It is 15 - 25 μm long, 10 - 16 μm wide and 5 -7 μm thick. By light microscopy the algae can be seen to have a longitudinal furrow on the ventral side and a deep well at a quarter of the cell length from the anterior end. Two flagella arise from the well. One flagella extends anteriorly 1,2 - 1,5 of the cell length. The second flagella, 0,5 - 1,2 of the cell length is housed in the shallow furrow. The cell swims smoothly without rotation. Six to thirteen chloroplasts are clearly visible with the light microscope, as is a teardrop shaped nucleus at the anterior of the cell.

By electron microscopy, the authors made the following observations (figure 1.1):-
* A thin cytoplasmic periplast lies between the plasmolemma and the chloroplast.
* The chloroplasts are covered by two double membranes the outer membrane being the chloroplast endoplasmic reticulum and the inner one being the chloroplast envelope.
* Each chloroplast, 3 - 4 μm long and 2 - 3 μm wide (Cattolico *et al.*, 1976), has pyrenoids protruding anteriorly.
* The pyrenoids are filled with granular material, penetrated by cytoplasmic canals and bounded by a double membrane associated with the chloroplast envelope.
* The chloroplast lamella consists of 2 - 3 thylakoids, none of which invade the pyrenoid matrix.
* Two ring-shaped nucleoids are found at each pole of the chloroplast.

The anterior nucleus is connected to the flagellar bases. Golgi bodies cover the anterior lateral surface of the nucleus, while rough endoplasmic reticulum surrounds the posterior half. Many mitochondria are found between the nucleus and the peripheral chloroplasts.

2.2 Nuclei Isolation

Nuclei, isolated by the method described in section 4.2.3 and visualized by Hoescht fluorescent staining, is shown in figure 2.1a. Intact algal cells, immobilised with gluteraldehyde, were visualized in the same manner and are shown in figure 2.1b. *O.luteus* chloroplasts fluoresce a bright red when viewed under ultraviolet light in the presence of, or without, Hoescht stain. Areas of high DNA concentration, such as the nucleus, fluoresce a brilliant blue in the presence of stain. Figure 2.1b indicates that the chloroplasts appear slightly larger than the blue nucleus. It is evident from figure 2.1a that the nuclei form the major composition of the nuclei isolation preparation, however several red plastids may be seen in the field of focus. The nuclei could be further purified by differential centrifugation through a self-generating Percoll gradient (Nothacker and Hilderbrandt, 1985), resulting in a purer nuclei preparation as judged by phase contrast microscopy. Qunatitation of yield was determined by reading absorbance at 260 nm of nuclei dissolved in .4 M NaCl. The DNA content after differential centrifugation through Percoll was less than half compared to the sample prior to centrifugation. This seems to indicate nuclei lysis. No differences in extracted protein, as judged by SDS - PAGE, were observed between the two samples. It was decided to forfeit nuclei purity in favour of a higher nuclei yield.

Several factors in the nuclei isolation procedure had a significant effect on nuclei yield. A loose dounce homogenizer was found to be far superior to a Virtis 45 homogenizer or a domestic Kenwood liquidizer. Although the latter two effectively ruptured cells, the nuclei tended to be sheared, resulting in a lower yield of isolated nuclei. Hexylene glycol was found to be essential for stabilizing the nuclei (Wray *et al.*, 1977). Although Rizzo and Burghardt (1983) reported hexylene glycol to be more effective at a final concentration of 1 molar, no difference in nuclei yield or stability was observed when the reagent was present in 0,5 molar or 1 molar final concentrations. A noticeably positive effect on nuclei yield was observed when the newly lysed cells were allowed to stir in the nuclei isolation medium for 1 hour at $4^{o}C$ prior to centrifugation. Presumably, this was due to giving the stabilizing hexylene glycol and salts time to percolate through the nuclei.

FIGURE 2.1a: Isolated *O.luteus* nuclei (Section 2.2.) viewed under a Nikon Diaphot-TMD Microscope fitted with a TMD-EF Flourescent attachment. Samples were stained with Hoescht stain prior to photographing. (Magnification X 200)

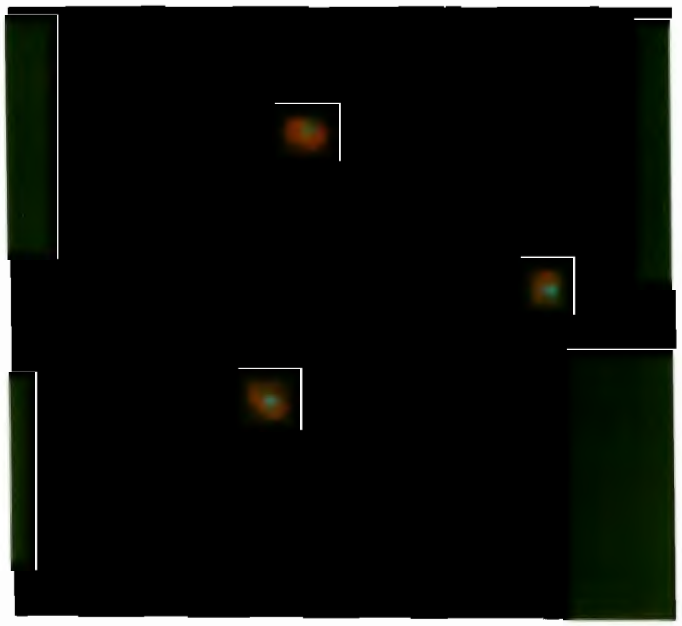

FIGURE 2.1b: Intact *O.luteus* cells, fixed with 0.1% gluteraldehyde and viewed as for Figure 2.1a. (Magnification X 200)

Triton X-100, Surfynol 485 and Nonidet P-40 were all investigated as to their suitability as detergents in this isolation protocol. As found by Rizzo and Burghardt (1983) Nonidet P-40 was found to be the most suitable, in that the cell membrane and most plastid membranes were effectively dissolved, while the nuclear membrane remained intact.

Thiodiglycol was present in the nuclei isolation medium to protect methionines from oxidation by scavenging free radicals (Zweidler, 1978). This problem appears to be prevalent when working with plant cells (J. Rodriques, Personal com.). TDG plays no role in stabilising nuclei during isolation procedures.

It should be noted that a comparison of nuclei yields resulting from different methods and buffer compositions is rather subjective. Factors, other than medium composition and protocol, such as age of culture and stage of cell cycle have an effect on nuclei yield and cannot always be controlled or quantified.

2.3 Analysis of Nuclear Proteins

To investigate the major protein fractions present in the *O.luteus* nucleus, isolated nuclei were boiled in 10 % SDS sample loading buffer (section 4.4) microfuged and analysed by SDS - PAGE (Figure 2.2). It appears that there are four major protein components present in the *Olisthodiscus luteus* nucleus. For ease of discussion, they have been labelled A, B, C and D: protein A being the slow migrating component and protein D the fastest migrating component. This protein complement was compared with calf thymus histones. Protein A has a similar mobility to calf thymus H1. Similarly, the distance migrated by protein B and D compares favorably with calf thymus H3 and H4 respectively. The major *O.luteus* fraction, protein C, did not co-migrate with any of the calf thymus histones, its electrophoretic mobility is between that of calf thymus H2A and H4. An *O.luteus* protein migrating a distance between that of calf thymus H2A and H2B was also present in small amounts. From stain intensities, it would appear that protein C is present in double the amount of protein D. Protein B is present in amounts slightly less than protein D and protein A is present in about half that amount. As compared with calf thymus histones, the molecular weights of the four

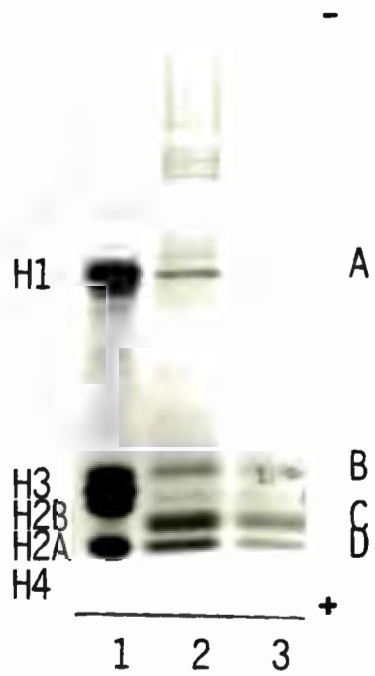

FIGURE 2.2: SDS-PAGE results solubilised *O.luteus* nuclei.
Lane 1: 120 µg calf thymus histone as standard
Lane 2 and 3: Solubilized *O.luteus* nuclei (Section 2.3). Lane 2 contains twice the amount of solubilised nuclei than lane 3.
Migration from top to bottom.

major proteins as judged by SDS - PAGE, appear to be: protein A = 22 500 daltons, protein B = 15 600 daltons, protein C = 13 000 daltons and protein D - 11 600 daltons.

Proteins B, C and D compare favorably with histones isolated from *O.luteus* (Rizzo *et al.*, 1985). However, the authors refer to a protein migrating faster than calf thymus H1 but slower than calf thymus H3, as H1. Figure 2.2 shows no discrete band in this region. The possibility exists that protein A, of similar molecular weight as calf thymus H1, is susceptible to proteolytic trimming, which forms the "H1" protein observed by Rizzo *et al.* (1985).

From the electrophoresis pattern of solubilized *O.luteus* nuclei, it is evident that there are many discrete high molecular weight protein components with slower mobilities than calf thymus H1. These may be contaminants such as ribosomal proteins or HMG proteins, or may be dimers and trimers of proteins A - D (despite being denatured in the presence of mercaptoethanol). These bands could also represent proteolytic products of a few high molecular weight proteins. Further investigation would be required to identify these components.

Based on their abundance in the *O.luteus* nucleus and the similarity in behaviour when analysed by SDS - PAGE, as compared with calf thymus histones and to the histones isolated by Rizzo *et al.*, (1985), it would be tempting to identify proteins A -D as histone proteins at this stage. However, further characterization of these proteins is necessary before such an assumption could be made.

2.4 Histone Extraction

If proteins A, B, C and D were indeed histones, one would assume that they would extract from nuclei by methods proven to be successful when histones from higher eukaryotes have been purified. 2 M NaCl extraction (Ohlenbusch *et al.*, 1967), HCl extraction (Johns, 1964), H_2SO_4 extraction (Fambrough and Bonner, 1966) and protamine displacement (Van der Westhuizen and Von Holt, 1971) histone extraction methods were performed (section 4.3.) and the results analysed and compared by SDS - PAGE (section 4.4).

2.4.1 Sodium Chloride Histone Extraction

The electrophoresis pattern of 2 M NaCl extracted (section 4.3.1) is shown in figure 2.3. Compared with the electrophoresis pattern of solubilized nuclei, some differences are immediately obvious. Proteins B, C, and D are present in both preparations. Two new protein bands not found in the solubilized nuclei are resolved in the 2 M NaCl extract. Protein A is present in smaller amounts in the NaCl extract as compared with the solubilized nuclei. The two new bands migrate slower than protein B, but faster than protein A. The intensities of the stained bands and therefore the quantity of protein present varied between these two bands from one extract to the following. In some 2 M NaCl extracts, the slower migrating protein of this doublet was found to be present in higher amounts than the faster, contrary to what appears in figure 2.3. The origin of these bands, labelled A1 and A2, is discussed in section 2.4.5. The relative quantities of proteins B, C, and D in the salt extract appears to represent the relative amounts of each found in the solubilised nucleus. Several high-molecular weight proteins were also found to be present, but in comparitively lesser amounts than solubilized nuclei. Typically, a total of 2 mg of lypholysed protein would be isolated from 5 L of culture with a cell count of $\pm$ 4 x 10^5 cell / mL, .

2.4.2 Sulphuric Acid Extraction

The results of H_2SO_4 histone extraction, when analysed by SDS - PAGE, are also shown in figure 2.3. Of all the histone extraction procedures (section 4.3), H_2SO_4 extraction proved to be the least successful (section 4.3.3). On average,

only 0,5 mg of total protein was extracted and recovered from 5 L of culture. The only protein bands present were small amounts of protein C and proteins A1 and A2 (discussed in section 2.4.5). The H_2SO_4 extraction procedure was not pursued.

2.4.3 Hydrochloric Acid Histone Extraction

The HCl histone extraction procedure (section 4.3.2.) resulted in an average of 2,5 mg protein from 5 L of culture. The electrophoresis pattern is shown in figure 2.4. Several differences exist between the HCl-extracted protein electrophoresis pattern and that of solubilized nuclei. No protein that migrates in the vicinity of H1 was found. Proteins A1 and A2 were found to be present in almost equal amounts. Very little protein D was present. Protein D is possibly not effectively extracted under these conditions. Protein B is present in slightly lesser amounts than proteins A1 and A2. A band between protein B and protein C, which is present in small amounts when nuclei were solubilized, is shown here to be present in a higher amount. It is possible that this acid soluble protein is a product of proteolytic trimming suffered by a high molecular weight protein. The stain intensity of this band, relative to the other bands, varies with each HCl protein extraction procedure. As compared with solubilized nuclei, only a small amount of higher molecular weight proteins were extracted in the presence of HCl.

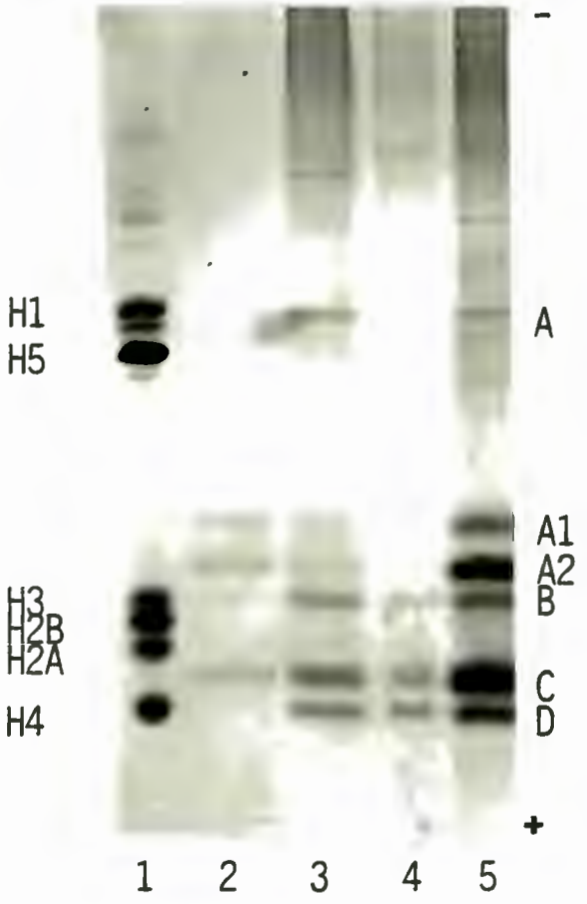

FIGURE 2.3: SDS-PAGE results of exctracted *O.luteus* nuclear proteins.

Lane 1: 120 μg chicken erythrocyte histones as standard

Lane 2: Sulphuric acid extracted *O.luteus* nuclear proteins (Section 2.4.2.)

Lane 3 and 4: Solubilised nuclei. Lane 3 contains twice the amount of solubilised nuclei as compared to lane 4.

Lane 5: 2M sodium chloride extraction of *O.luteus* nuclear proteins (Section 2.4.1.)

For *O.luteus* samples, 1/10 of total protein extracted from nuclei isolated from 5 L culture was analysed by PAGE.

Migration from top to bottom.

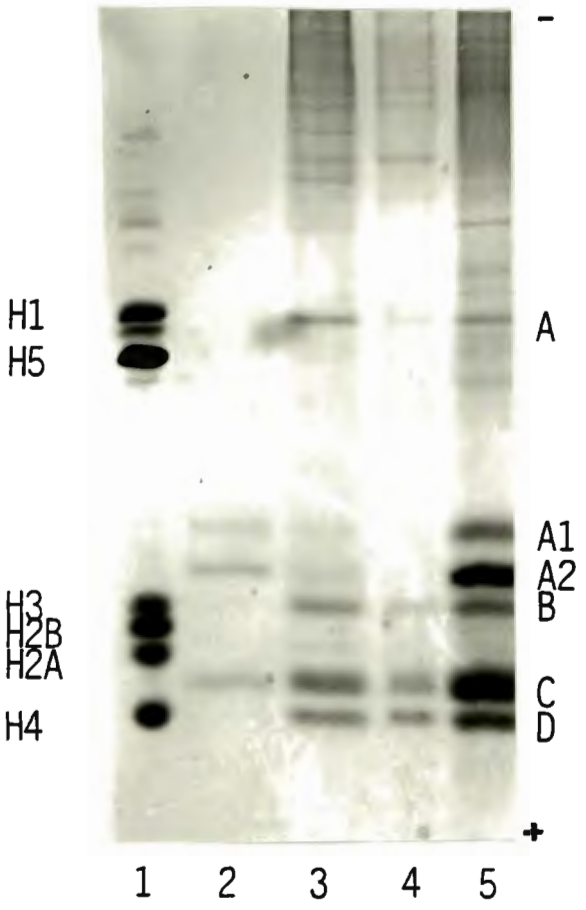

FIGURE 2.3: SDS-PAGE results of exctracted *O.luteus* nuclear proteins.

Lane 1: 120 µg chicken erythrocyte histones as standard

Lane 2: Sulphuric acid extracted *O.luteus* nuclear proteins (Section 2.4.2.)

Lane 3 and 4: Solubilised nuclei. Lane 3 contains twice the amount of solubilised nuclei as compared to lane 4.

Lane 5: 2M sodium chloride extraction of *O.luteus* nuclear proteins (Section 2.4.1.)

For *O.luteus* samples, 1/10 of total protein extracted from nuclei isolated from 5 L culture was analysed by PAGE.

Migration from top to bottom.

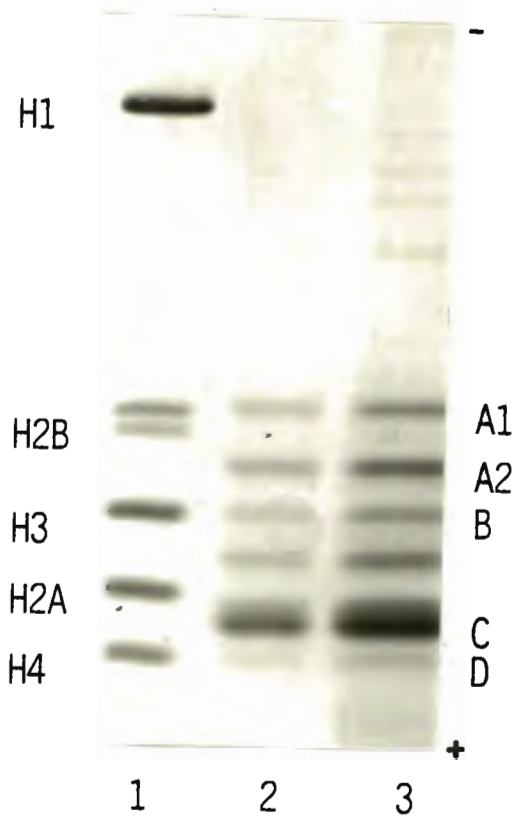

FIGURE 2.4: SDS-PAGE results of hydrochloric acid extracted *O.luteus* histones.

Lane 1: 100 µg sea urchin sperm (*Parechinus angulosus*) histones as standard.

Lane 2 and 3: 1/20 and 1/10 resp. of total hydrochloric acid extracted *O.luteus* nuclear proteins performed on nuclei isolated from 5 L culture as described in section 2.4.3..

Migration from top to bottom.

2.4.4 Protamine Histone Displacement

The protamine displacement of histones procedure proved to be the most successful for extracting *Olisthodiscus luteus* nuclear proteins (section 4.3.4). Figure 2.5a illustrates the Sephadex - G50 gel filtration elution profile, whereby the nuclear proteins were separated from the protamine. Aliquots of the eluted peaks were analysed by SDS - PAGE (figure 2.5b). Proteins A1, A2, B, C, and D are the major proteins present. Compared with solubilized nuclei, a smaller amount of high-molecular-weight proteins are present. As in the previous two extraction protocols, no protein band corresponding to higher eukaryotic H1 is present. Protein D appears to be present in a lesser amount than the other proteins. However, protease activity is strongly suggested. Two fainter bands between proteins B and C were found to vary from one extraction to the next. Similarly, a band, faster migrating than protein D, is clearly visible. It is possible that the latter band represents a trimmed protein D. A summation of the stain intensities of this band and protein D would be comparable to the intensity of the bands of proteins A1, A2, and B.

The amount of histones present in eukaryotic cells is generally equal to the mass of the nuclear DNA. The theoretical yield of histones from 5 L of eukaryotic cells at a cell density of 3×10^5 / mL and a DNA content of $1,66 \times 10^{-12}$ g / cell is 2,49 mg. The protamine displacement protocol yielded 8 mg of protein as judged by absorbance at 230 nm. This is four times greater that the theoretical yield. If histones are present in this extraction, they are contaminated with other proteins or substances that absorb at 230 nm. Some high, as well as low, molecular weight proteins can be seen in figure 2.5b. A summation of the amounts of protein present in these bands and present in amounts to small to be detedcted could well exceed the total amount of proteins found in the major bands.

The reason for the increase in yield of protein, compared with other extraction methods, is probably due to the displacing activity of the protamine. Assuming proteins A - D are indeed histone proteins bound to the DNA, the NaCl, H_2SO_4 and HCl extraction methods extract histones by diminishing the ionic interactions between the DNA and the basic protein. This is achieved by either neutralizing the DNA phosphates by dropping the pH or, by increasing the ionic strength by the addition of salts. No consideration has been given to the possibility of non-specific binding of the proteins to other substances. (Possibly an acidic

substance not found in higher organisms where NaCl, HCl and H_2SO_4 histone extraction methods have proven to be adequate). Protamine was observed by Mirsky and Ris (1951) to actively displace histones from nucleo-protein complexes. It is likely that protamine would also bind to sites that non-specifically bind histone proteins. This extraction procedure not only breaks the histone/DNA bond by raising the ionic strength and thus reduces the binding strength, but also ensures that the proteins do not rebind as a result of the protamine presence on the DNA. Furthermore, the high concentration of protamine effectively swamps possible binding sites where non-specific interactions with the histone proteins may occur.

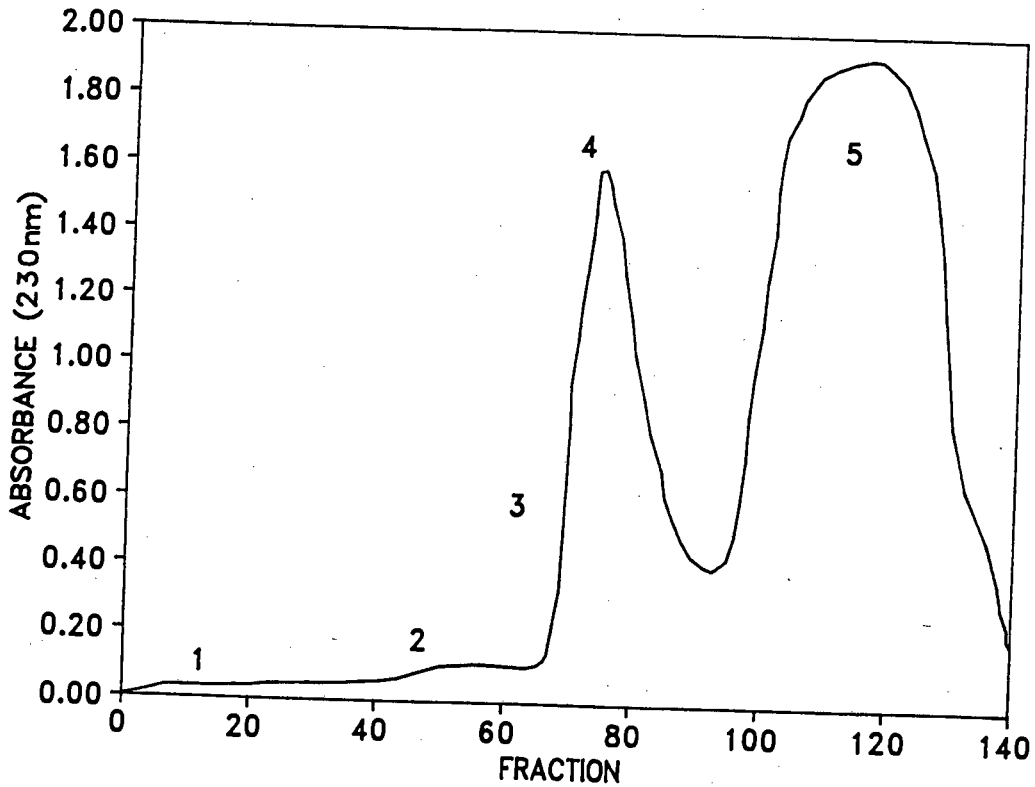

FIGURE 2.5a: Molecular exclusion chromatography of protamine displaced *O.luteus* nuclear proteins on a column (3 X 90 cm) of Sephadex-G50 (Section 2.4.4.).
Eluent: 10 mM Tris-HCl, pH 7,4; 0,15 M NaCl.
Fraction size: 1,5 mL.

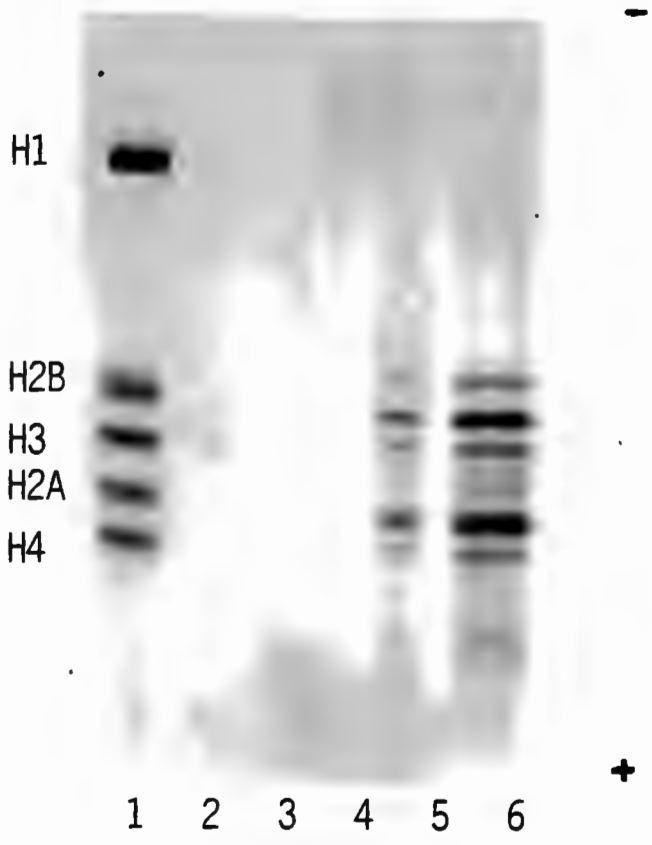

FIGURE 2.5b: SDS-PAGE results of column fractions 1 - 5 as indicated in figure 2.5a..

Lane 1: 120 µg sea urchin sperm (*Parechinus angulosus*) histones.

Lane 2 - 6: Fractions 1 -5 respectively. 1/50 of total volume after dialysis was used for PAGE analysis.

Migration from top to bottom.

2.4.5 Conclusions from Protein Extraction Protocols

Proteins A1, A2, B, C, and D have been shown to have molecular weights in the region expected for histone proteins. Their abundance in the nuclei has been clearly demonstrated. The acid extraction procedure indicates that the proteins are basic. As is shown by the effectiveness of the protamine displacement procedure, it would appear that these proteins do bind to the DNA in the nucleus. The assumption that these proteins, or at least some of them, are histone proteins, may now be safely made. The identity of each fraction represented by bands in the electrophoretic pattern cannot be achieved simply by comparing mobilities during SDS - PAGE. Positive identification can only be made after sequence analysis has been conducted.

The electrophoretic pattern of the extracted basic proteins compare favorably with the histones extracted by Rizzo *et al.* (1985), with the exception of Protein A. Protein A has been observed by Rizzo *et al.* (1985) and has been shown in this study to be present in solubilised nuclei. Protein A was however not found to be present in any of the protein extracts discussed above. Several possibilities could be responsible for this observation. It has been noticed that H1 is the most susceptible to endogenous proteolytic degradation (Wolf Brandt, Personal comm). In the course of this study, it is possible that protein A, which may be histone H1, is completely degraded. If the proteolytic activity resulted in a slight trimming at one or both of the protein terminals, this might result in proteins A1 and protein A2. It is also possible that protein A is not a histone or DNA-binding protein. Assuming that protein A is not a histone-type protein, and is not broken down to form protein A1 or protein A2 or both, the absence of proteins A1 and A2 in figure 2.2 requires explanation. If these proteins were very basic, the addition of 10 % SDS would possibly result in the formation of inverse pseudo-micelles. This would lead to an insoluble hydrophobic complex forming, which is pelleted when the solubilized nuclei sample was centrifuged during the course of the sample preparation (section 4.4). Electrophoresis of the soluble fraction would then not show the presence of protein A1 or A2, although they are present *in vivo*. The true nature of proteins A1 and A2 can only be established by partial sequence analysis.

Protein C compares favorably with Rizzo *et al.* (1985) histone *H01*. *H01*, the authors suggest, is a single histone like protein, functioning both as H2A and

H2B. However, a close inspection of protein C in figure 2.4 strongly suggests the presence of two proteins. This may be a non-histone contamination electrophoresing with the *H01* protein band, or possibly an indication that *H01* is in fact two independent histones that co-migrate when analysed by SDS - PAGE.

Proteins B and D compare well with the H3 and H4 histone proteins identified by Rizzo *et al.* (1985). The ratio of the amounts of protein B and D as to the other extracted proteins was found to vary between different extraction procedures. This has been observed for histones H3 and H4 (W. Brandt, personal com). If proteins B and D are H3 and H4 respectively, they would be arginine rich relative to the other histones. This would result in the proteins B and D binding more strongly to the DNA and consequently making quantitative extraction more difficult.

2.5 Proteolysis Inhibitor Assay

The presence of fast migrating bands found by SDS - PAGE seems to indicate proteolytic activity. Such activity, left uninhibited, can reduce yields and generate a complex mixture of fragments. In an attempt to inhibit this activity, various known protease inhibitors were assayed as described in section 4.6. The electrophoretic pattern is shown in figure 2.6.

Lane 2 sample was sea urchin sperm (*Parechinus angulosus*) histones (a gift from Sylva Schwager), maintained at -60^{o}C until required. No proteolytic activity is evident. To check for the possibility of a contaminating protease in this preparation, an aliquot of the preparation was maintained overnight at 37^{0}C prior to electrophoresis. Very little proteolytic activity, or none at all, was expected for these two samples. In an attempt to detect the smallest amount of proteolytic products, double the amount of sample as compared to the other samples was loaded onto the gel. By the decrease of the amount H1 present, a small amount of proteolytic activity is apparent in the incubated sample (lane 3). All samples were prepared and treated as explained in section 4.6 and in the legend of figure 2.6. Most of the inhibitors assayed displayed some inhibitory action. Lane 4 (sea urchin sperm, cell lysate and no inhibitors) shows complete proteolysis. EDTA, NaBisulphite, as well as low or high pH, hardly offer the histone proteins any protection from degradation. Surprisingly, the serine protease inhibitors PMSF and BZA, both used extensively as protease inhibitors in histone and chromatin studies, offer very little protection to the histone proteins under these circumstances. Leupeptin, Pepstatin, pCMB and TLCK offer some protection, but proteolytic activity is not completely inhibited. TPCK, a serine protease inhibitor appears to protect the core histones from proteolysis - but there is a total absence of H1. As a result in this assay, undegraded H1 would indicate an effective protease inhibitor. The only inhibitor assayed that shows the presence of some undegraded H1 is EGTA.

Assuming *O.luteus* histone proteins would react similarly to sea urchin sperm histone proteins under the same circumstances, it was decided that EGTA should be included in all nuclei isolation buffers. Later on in the course of study, Nothacker and Hildebrandt's (1985) buffer system, described in section 4.1.1, was found to be superior to any other system tried (data not shown). These authors

called for an EGTA concentration of 5 mM. We found this concentration sufficient to inhibit much of the unwanted proteolytic degradation of proteins under study.

The sample analysed in lane 5 contained all the tested inhibitors, including EGTA. It would be expected that the proteolysis in this sample would also be inhibited. It is possible that the EGTA is affected by one of the other inhibitors, such as sodium molybdate, and therefore cannot offer the protein substrate protection.

It was also found that proteolysis decreased significantly if all procedures of protein purification were performed at 4^0C, and that a quick (± 2 hours) nuclei isolation protocol was better than a more thorough, time consuming one (although the latter may result in a cleaner nuclei preparation).

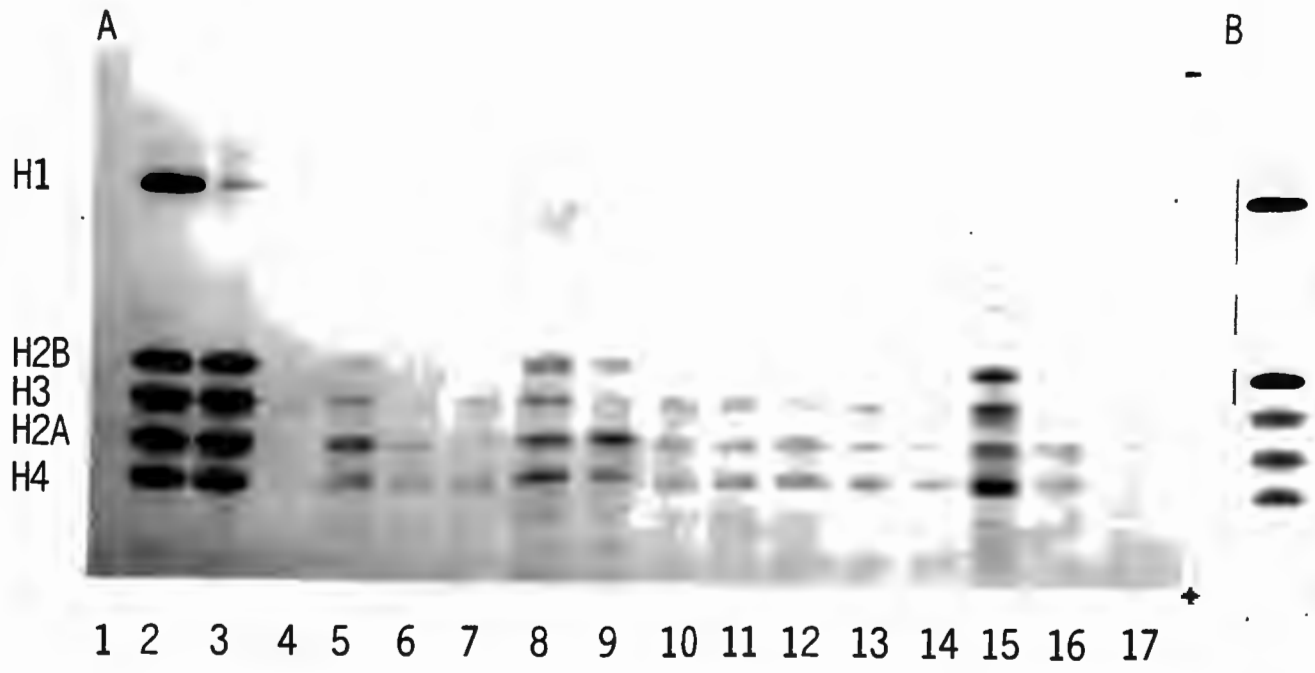

FIGURE 2.6: SDS-PAGE results of the proteolytic inhibitor assay as discussed in section 2.5..

Gel A: Proteolysis Experiment

Lane 1 - 17 corrospond to tubes 1 - 17 of Table 4.1.. The inhibitors in each tube is summarised below.

Lane 1: No substrate[1]
Lane 2: Frozen substrate[1]
Lane 3: No proteolytic factor
Lane 4: No inhibitors
Lane 5: All inhibitors
Lane 6: PMSF
Lane 7: BZA
Lane 8: TPCK
Lane 9: TLCK
Lane 10: pCMB
Lane 11: Leupeptin
Lane 12: Pepstatin
Lane 13: Na_2MoO_4
Lane 14: EDTA
Lane 15: EGTA
Lane 16: Sodium bisulphate
Lane 17: Low pH

1 400 μg sea urchin sperm (Parechinus angulosus) total histones. All other samples contain 200 μg histones as substrate.

Gel B: Control

SDS - PAGE of 200 μg sea urchin sperm (Parechinus angulosus) total histones after extraction and prior to any treatment or storage.

Migration from top to bottom.

2.6 Fractionation of *Olisthodiscus luteus* Basic Proteins

2.6.1 Column chromatography on Bio-Gel P60 Column

In an attempt to fractionate *O.luteus* histone (or histone-like) proteins, gel chromatography over a column of Bio-Gel P60 as described in section 4.7.1 was performed. The elution profile of the algal proteins is shown in figure 2.7. Figure 2.7a is the electrophoretic pattern of aliquots taken from high-absorbing fractions (as measured by absorbance at 230 nm) and analysed by SDS-PAGE.

Five major peaks, labelled from 1 (early eluting) to 5 (late eluting) were resolved. Peak 1 shows the presence of a large shoulder. This peak was therefore divided into two: Peak 1a and Peak 1b. Peaks 2, 3 and 4 have a maximum absorbance (230 nm) considerably lower than that of peaks 1 and 5. Peak 5, the largest, was expected to contain the urea and mercaptoethanol used in the sample preparation.

It is clear from figure 2.7a that none of the eluted peaks contain a pure, homogeneous protein fraction. Peaks 1a and 1b contain no, or very little protein, that corresponds with the sea urchin sperm (*Parechinus angulosus*) histones used here as a standard. Some high-molecular weight proteins do appear to be present. A high percentage contamination of the nucleic basic proteins may be predicted from the theoretical yield of expected protein as discussed in section 2.4.4.

Peak 2 contains both protein A1 and protein A2. Peak 3 contains a single major protein fraction previously labelled protein C (figure 2.2). Peak 4 contains two major protein bands - one corresponds electrophoretically with protein B and the other with protein C. Protein C appears to be a heterogenous mixture of at least two proteins. A comparison of the electrophoretic patterns of peaks 3 and 4 shows that the protein in peak 3 has a mobility marginally slower than that of the faster migrating protein in peak 4. It is not difficult to imagine that if these proteins were to be present in the same sample, their SDS - PAGE pattern would suggest a single protein, present in large amounts. The slower migrating protein will now be referred to as protein C1 and the faster migrating protein in peak 4 as protein C2.

Peak 5 contains protein D which, under these conditions, did not separate from the urea and mercaptoethanol components of the sample.

All fractions isolated (A - D) are still contaminated by a large number of proteins. In some cases, this contamination may represent as much as 50 % of the total proteins present in that fraction. There are strong indications that the major proteins have been subjected to proteolytic activity. This can be seen by the appearance of "smudged" bands slightly below the major bands (figure 2.7a).

Protein B in peak 4 appears to be present in relatively smaller amounts compared to the original protamine extraction (figure 2.5b). This might indicate that protein B is susceptible to non-specific binding to the column packing material.

It is clear from figure 2.7a that the major proteins need further purification prior to any form of sequence analysis. This was achieved by reverse-phase HPLC. Improved peak resolution, by manipulating pH and salt concentrations (Von Holt and Brandt W.F., 1977) and the addition of protease inhibitors to inhibit proteolytic activity, would result in a better performance of the Bio-Gel P60 chromatography procedure. However, the success of protein separation by High Performance Liquid Chromatography precluded the necessity of using Bio-Gel P60 gel chromatography altogether.

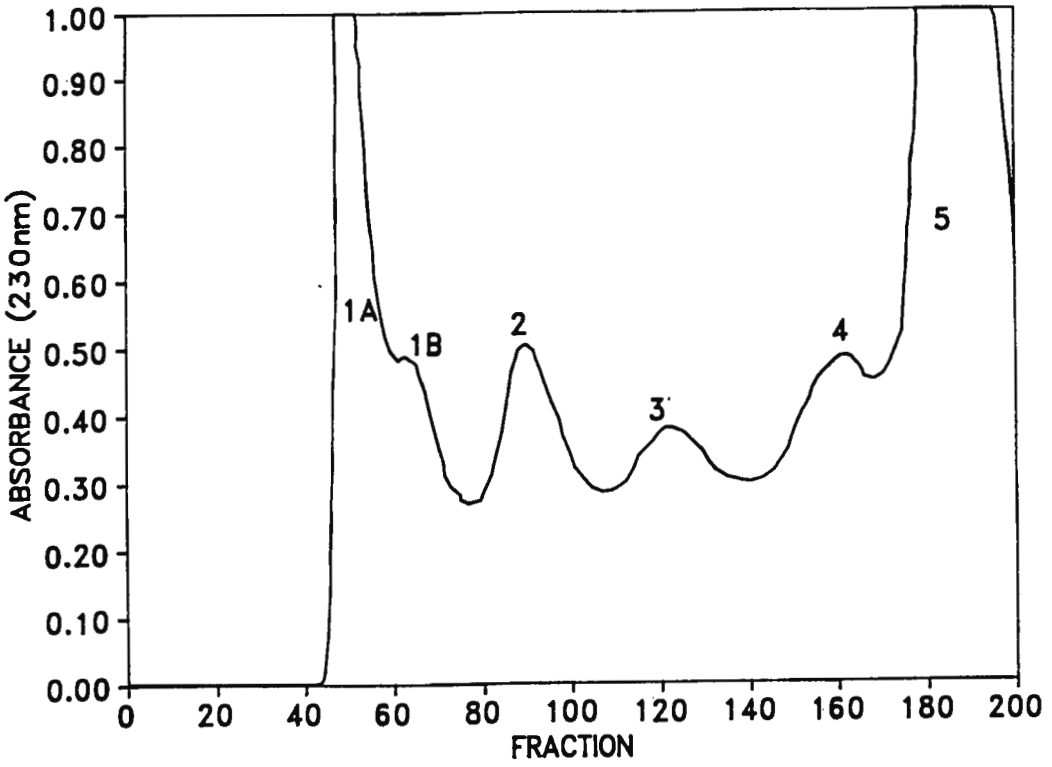

FIGURE 2.7: Molecular exclusion chromatography of *O.luteus* nuclear basic proteins on a column of Bio-Gel P60 (3 X 120 cm).
Eluent: 0,02 N HCl; 50 mM NaCl.
Sample Buffer: 6 M urea; 50 mM NaC; 1% mercaptoethanol. Sample made 0,02 N HCl just prior to applying to column.
Fraction size: 600 µL.

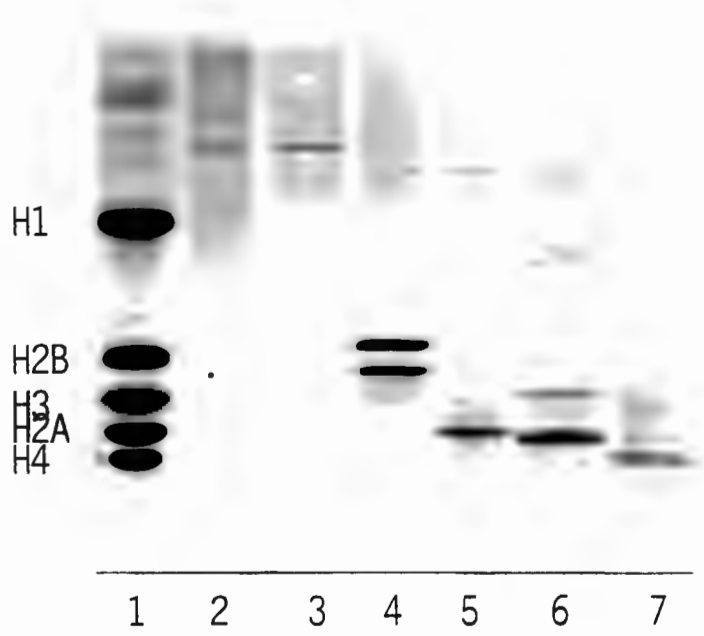

FIGURE 2.7a: SDS-PAGE results of column fractions 1 - 6 as indicated in Figure 2.7. (Section 2.6.1.).
Lane 1: 130 μg sea urchin sperm (*Parechinus angulosus*) histones.
Lane 1 - 7: Fractions 1a, 1b, 2, 3, 4 and 5 respectively.
Migration from top to bottom.

* The reverse phase HPLC method is fast; the entire process can be completed in 90 minutes as opposed to the ±24 hours required for gel chromatography.
* In addition, the shorter time also seems to reduce the non-specific proteolytic activity.
* The buffer system employed here has the advantage over other HPLC buffer systems (examples given by Certa and Von Ehrenstein, 1981) and that of the Bio-Gel P60 eluting buffer, in that the buffer constituents are volatile and can be easily removed by exposure to a stream of N_2 and by lyophilization.
* The most important advantage of this HPLC method is that, except for proteins A1 and A2, all other major proteins are fractionated into separate and homogeneous peaks that can be easily collected and directly utilized for sequence analysis.

A drawback of the HPLC system is the amount of total protein that may be processed in a single HPLC run. Typically, a maximum of 6 mg total protein could be loaded without leading to peak merging and broadening. However, due to the speed and effectiveness of the system, several runs could be achieved in one day. This allowed for processing of enough sample for further analysis, while remaining more time-efficient than normal gel chromatography.

A comparison was done between the HPLC elution profile of *O.luteus* nuclear proteins and chicken erythrocyte total histone proteins chromatographed under the same conditions (figure 2.9). Although appearing smaller by SDS - PAGE, proteins A1 and A2 elute at a time similar to that of chicken erythrocyte H1. Protein C2 (peak 10 figure 2.8) elution time corresponds well with chicken erythrocyte H2B. Peak 11 (figure 2.8) or protein C1 is similar, but slightly earlier, than chicken erythrocyte H2A. Similarly, peak 15 and 18 (proteins D and B respectively) compare favorably with chicken erythrocyte H4 and H3 elution times. This seems to indicate hat the major *O.luteus* proteins, A - D, are indeed histone proteins and are fairly closely related to their higher eukaryotic equivalents. The similarity in elution times between *O.luteus* proteins and those of chicken erythrocyte histones strongly suggests that proteins A1 and A2 correspond to H1, protein B to H3, protein C2 to H2A, protein C1 to H_2B and protein D to H4. Definitive identification can however only be made after partial primary structure determination of these proteins has been conducted.

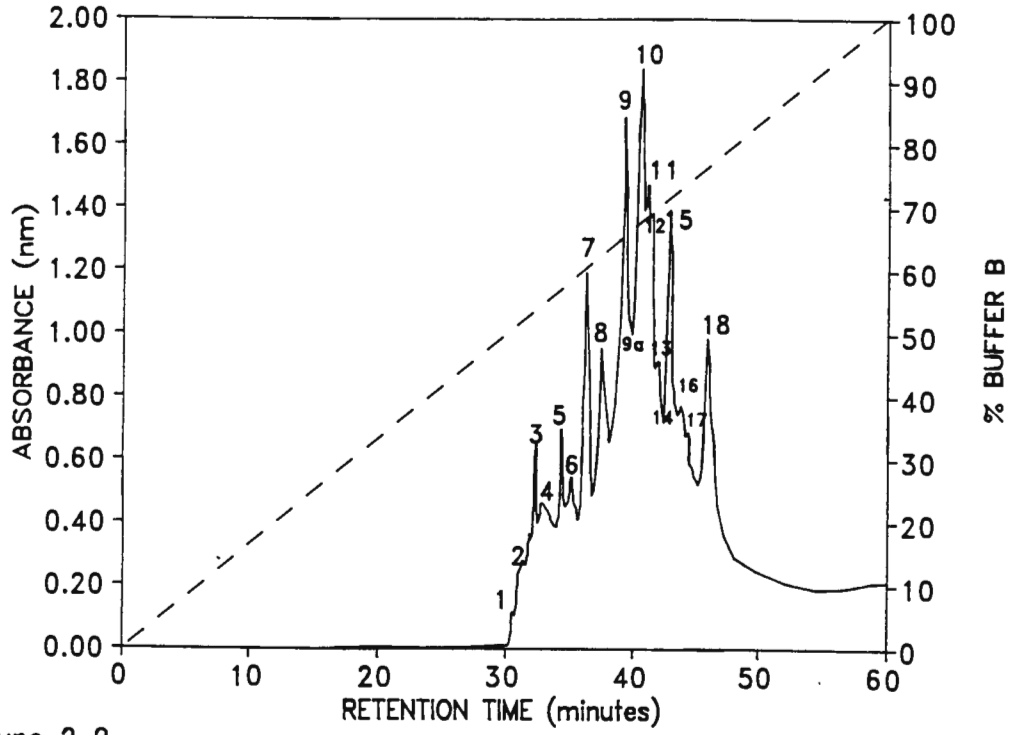

Figure 2.8

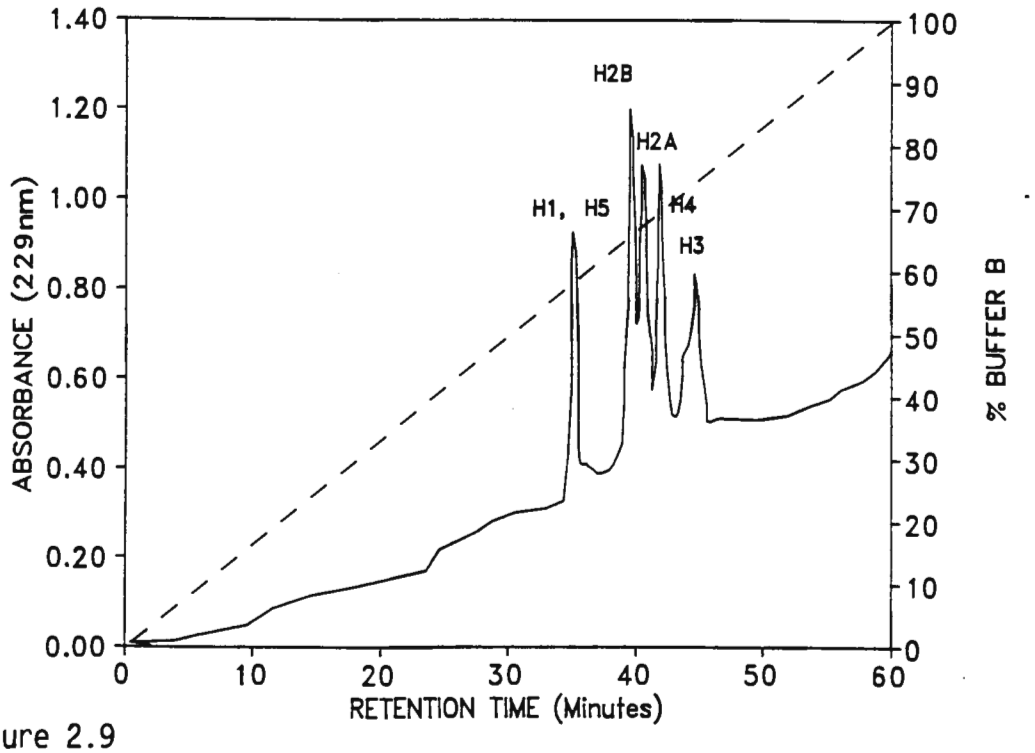

Figure 2.9

FIGURE 2.8 and 2.9: Reverse-phase high performance liquid chromatographic separation of *O.luteus* basic nuclear proteins (figure 2.8) and chicken erythrocyte histones (figure 2.9).

Sample: 6 mg *O.luteus* protein dissolved in 6 M urea, 500 µL Buffer A (figure 2.8).

5 mg chicken erythrocyte histones in 6 M urea, 500 µL Buffer A (figure 2.9).

Mobile Phase: A = 0,1 % (v/v) HFBA in water.

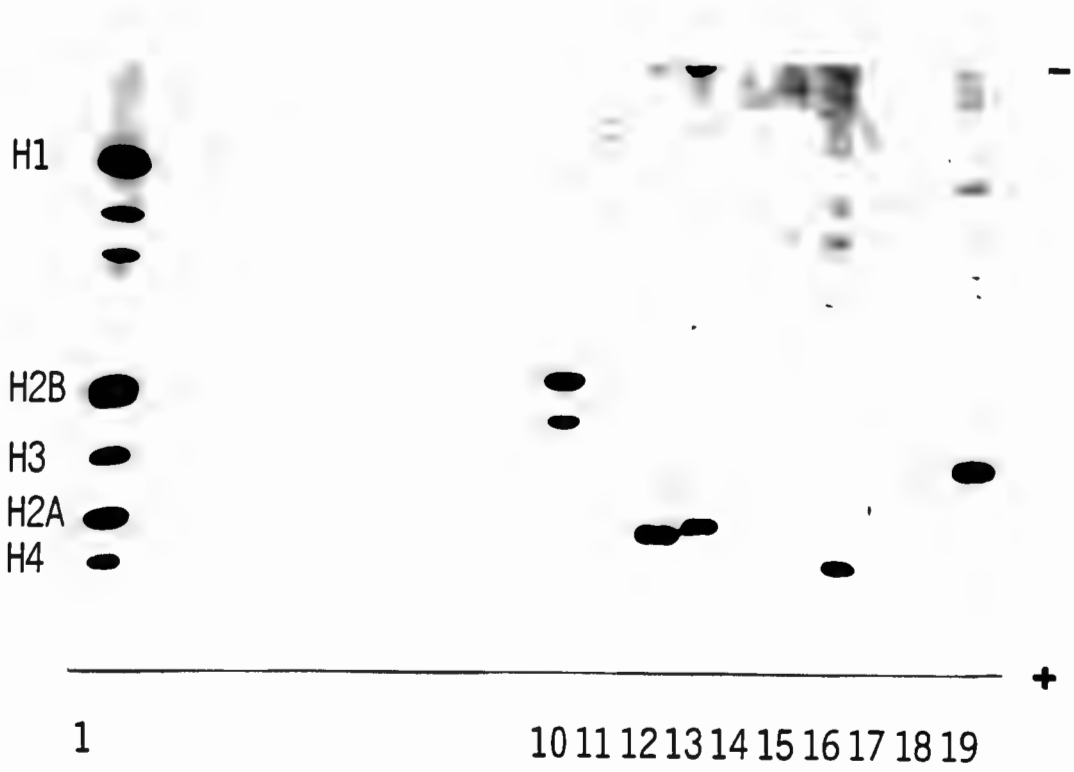

FIGURE 2.8b: SDS-PAGE results of *O.luteus* basic nuclear protein separated by HPLC.
Lane 1: 120 μg sea urchin (*Parechinus angulosus*) histones
Lanes 10, 12, 13, 16 and 19 are peaks 9, 10, 11, 15 and 18 respectively as indicated in Figure 2.8..
Fraction 14 (the trough between peaks 11 and 15 was not analysed by SDS-PAGE.

2.7 Elucidation of the Partial Primary Structure of Proteins A1 and A2

2.7.1 Electroelution of Proteins A1 and A2

Proteins A1 and A2 have defied separation by gel chromatography over Bio-Gel P60 or by HPLC using HFBA / Acetonitrile buffer systems, or by HPLC employing a TFA / Acetonitrile buffer system (data not shown). The fact that proteins A1 and A2 did separate when analysed by SDS - PAGE, led to the development of an electroelution system described in (section 4.5). Enough separated protein A1 and A2 samples were recovered to allow application of the Edman Degradation procedure.

2.7.2 Partial Primary structure determination of Protein A1 and Protein A2

Protein A1 and A2 were not subjected to any form of chemical or enzymatic cleavage procedures. After electroelution as described in section 4.5 and section 4.10, both samples were loaded directly onto the sequencer. Protein A1 was subjected to 48 cycles of Edman degradation and protein A2 was subjected to 46 cycles. The yields of PTH-amino acids for protein A1 and A2 are shown in Tables 2.1 and 2.2 respectively.

The resulting sequences were compared with known bovine histone sequences. Significant homology was found when protein A1 and protein A2 were compared with *Bos primigenius taurus* (bovine) H1 (Liao and Cole, 1981). In the alignment below, the calf thymus H1 sequence is used as the consensus sequence, (.) indicates homology to the consensus and (-) indicates a deletion. Mutations are indicated by identifying the changed residue by single letter amino acid notation. Unidentified amino acids are indicated by X.

Olisthodiscus luteus

```
Protein A1:                                                .TYY
Protein A2:                                                TYY
Bovine H1:   SETAPAAPAA APPAEKTPVK KKAAKKPAGA RRKASGPPVS
             1         10         20         30         40

Protein A1:  D.VKD..V.L .D.N.S.MQ. I..YIE.NQK V..-..HY-L RAA..
Protein A2:  .L.KA..L.L ...N.S.AQ. I..YILENNK IEF--QQTFL RXA..
Bovine H1:   EMITKAIAAS KERSGVSLAA LKKALAAA-G YDVEKNNSRI KLGLKSLVSK
                       50         60         70         80         90
```

Olisthodiscus luteus protein A1 shows 40 % homology when compared with calf thymus H1. Similarly, *Olisthodiscus luteus* protein A2 shows 32 % homology. The homologous sequences align to the conservatively evolved domain of the H1 protein family. Although the homology was found to be less than 50 %, a perusal of known H1 sequence and other histones sequences, will indicate that the histone H1 family is the most rapidly evolving of the histones. This homology, together with the hydrophobicity (as seen by HPLC) of proteins A1 and A2 leave little doubt that they are equivalent to H1.

Initially, it was thought that proteins A1 and A2 were identical proteins, having suffered proteolytic degradation to different extents. When the two sequences are aligned however, there appears distinct differences between the two. This would indicate that the algae have at least two H1 genes. The two H1 proteins sequenced here show a 60 % homology.

Both algal H1 proteins appear to be trimmed N- terminally when compared with calf thymus H1. If this trimming was due to a protease, one would expect a protease "hot spot" just prior to the conservative domain of the algal H1s. This would result in both algal H1 variants to be cleaved, with what appears to be 100 % efficiency, at almost the same residue. The domain structure of the H1s could possibly lead to the specific cleavage, the extended nose and tail domain being

more susceptible to cleavage as a result of conformational constraints of the globular domain. However, protein A1 (H1a) has a one N- terminal residue extension - proline - when compared with protein A2 (H1b). No protease enzyme reported is known to cleave either N- and C- terminally at proline residues. This would indicate that the shortened N- terminal sequences of the *O.luteus* H1s, when compared with higher eukaryotes, is gene coded. However, the possibility remains that both proteins have undergone non-specific or specific C- terminal trimming. Despite precautions to prevent non-specific proteolytic activity, no H1 proteins larger than the ones reported here were isolated.

2.8 Elucidation of the Partial Structure of Protein B

2.8.1 Peptide Production from Protein B

As described in section 4.9.11 protein B (0,5 mg) was subjected to *Staphylococcus aureus* endoproteinase digestion at pH 4,0. The peptides produced were fractionated by HPLC as described in section 4.7.2, and the resulting chromatogram is shown in figure 2.10. The elution profile showed little similarity to sea urchin (*Parechinus angulosus*) histones H1, H2A, H2B, H4 and wheat H3 digested and fractionated under the same conditions. The following peptides were sequenced : B.V8.1: B.V8.2; B.V8.3; B.V8.6; B.V8.7 and B.V8.11.

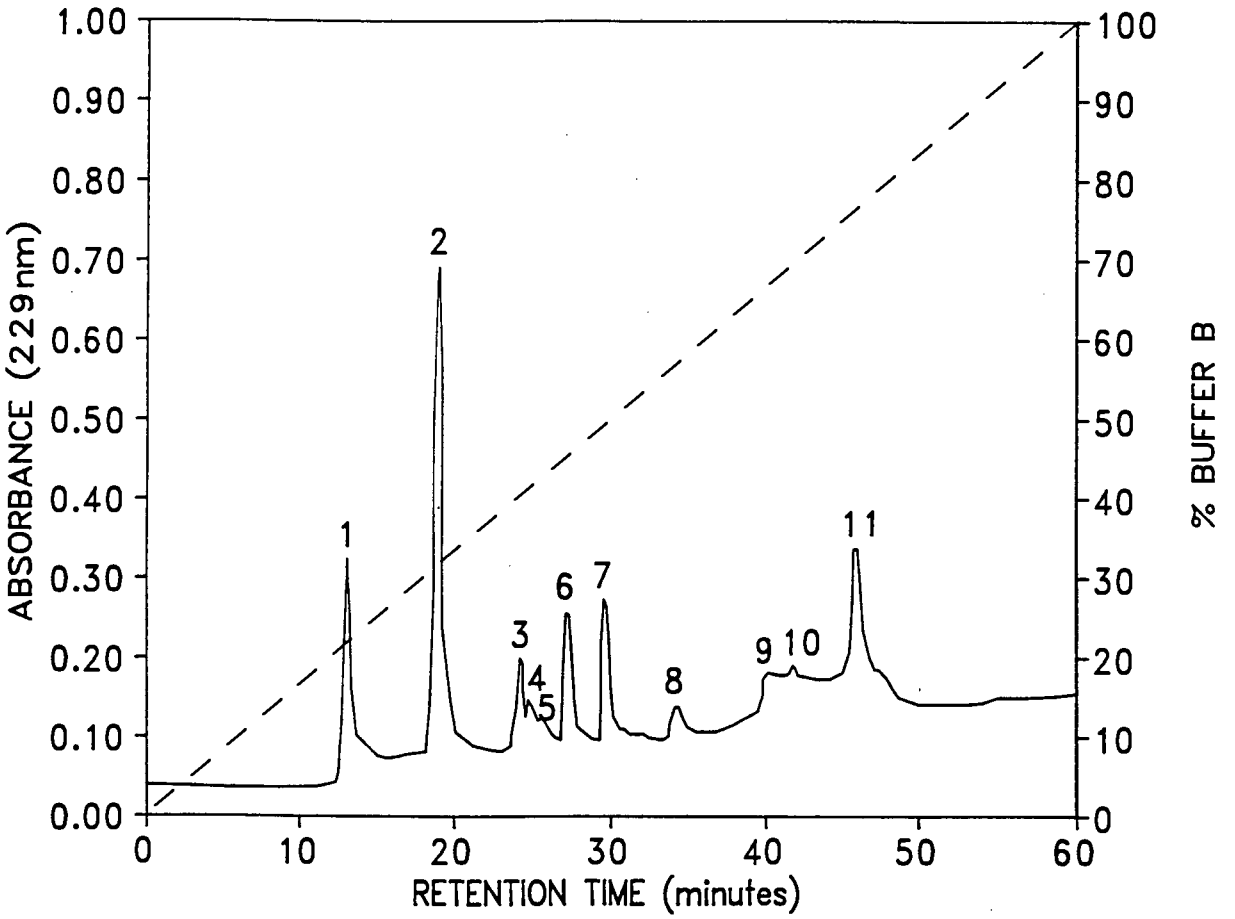

FIGURE 2.10: Reverse-phase high performance liquid chromatographic separation of Protein B digested with *Staphylococcus aureus* V8 protease.
Sample: 0,5 mg Protein B; 25 µg *Staphylococcus aureus* V8 protease; 200 µL 0,05 M ammonium acetate.
Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile
Flow Rate: 0,7 mL / minute

2.8.2 Partial Primary Protein Structure of Protein B

The partial primary structure of protein B was determined by subjecting the uncleaved protein as well as the peptides discussed in section 2.8.1 to Edman degradation by using the gas-phase sequencer (section 4.10). The resulting amino acid yields, after sequencing uncleaved protein, is shown in Table 2.3. The ninth cycle of Edman degradation yielded lysine, of which 82.9 % was mono-methylated. The yield of amino acids as determined from the sequencing of peptides produced by *Staphylococcus aureus* V8 endoproteinase digestion of protein B is reported in Table 2.4 - Table 2.7. In all cases, one-fifth of the isolated peptides were loaded onto the sequence disc. Peptide B.V8.11 did not yield a sequence, even after the total amount of peptide under the peak was applied to the disc. Peptide B.V8.2 results were the same as for uncleaved protein. Data from the sequencing of peptides B.V8.2 and B.V8.11 has therefore not been shown here.

The resulting sequences were compared with bovine histone sequences. 100 % homology was found between the sequenced peptides and those of *Bos primigenius taurus* (bovine) histone H3. The alignment is shown below.

```
                 ┌──── uncleaved protein b ────────
Protein B:       ..........  ..........  ..........
Bovine H3:       ARTKQTARKS  TGGKAPRKQL  ATKAARKSAP
                 1         10          20          30

                                                Peptide
                 ┐                            ┌ B.V8.1┐┌
Protein B:       .                            ..........
Bovine H3:       ATGGVKKPR   YRPGTVALRE  IRRYQKSTEL
                          40          50          60

                 Peptide B.V8.6 ┐
Protein B:       ..........  ...
Bovine H3:       LIRKLPFQRL  VREIAQDFKT  DLRFQSSAVM
                          70          80          90

                           Peptide
                       ┌ B.V8.7 ┐┌ Peptide B.V8.3
Protein B:             ...  ..........  ..........
Bovine H3:       ALQEACEAYL  VGLFEDTNLC  AIHAKRVTIM
                          100         110         120

                 ┐
Protein B:       .
Bovine H3:       PKDIQLARRI  RGERA
                          130
```

The same notation as in section 2.7.2 has been used here except that spaces within the sequence represent unsequenced domains. The peptides constitute 57,8 % of the total bovine H3 sequence. Protein B can therefore be positively identified as H3. This is in keeping with observation of protein B's electrophoretic and chromatographic behaviour.

It is accepted that the sequences of the H3 families, like those of H4, are generally conserved in evolution. It is remarkable however that two organisms as divergent as bovine and marine algae show 100 % homology for the peptides sequenced. Differences may exist in domains not yet sequenced. A full comparison between *O.luteus* H3 and that of other organisms is presented in chapter 3.

2.9 Elucidation of the Partial Primary Structure of Protein C1

2.9.1 Peptide Production from Protein C1

0,5 mg of protein C1 was digested with *Staphylococcus aureus* V8 proteinase at pH 4 (section 4.8.1.1) and resulting peptides separated by HPLC (section 4.7.2). The consequent elution profile (figure 2.11) was compared with elution profiles of sea urchin sperm (*Parechinus angulosus*) H4, H2A and H2B treated in the same manner. No similarities in elution profiles were found. In the case of C1, three major and four minor peaks were resolved. Two of the major peaks (peaks 5 and 8) appear to have "shoulders". This would indicate a heterogeneous peptide population represented by these peaks. This heterogeneity could be due to several possibilities. Two different peptides could possibly be eluting at similar times. Several partial digestion products bracketing the same hydrophobic domain are eluting close together or, finally, the peptide population may contain variants of the same sequence. All resulting peptides, except for the minor peptides C1.V8.6 and C1.V8.7 were sequenced.

2.9.2 Primary Structure Determination of Peptides generated from Protein C1

The uncleaved protein C1 was subjected to Edman degradation in an attempt to determine its N- terminal sequence. Several attempts yielded no detectable amino acid sequence. It was concluded that this protein was blocked to the Edman degradation chemistry. Following this, the protein was cleaved by *Staphylococcus aureus* V8 protease. The fractionated peptides analysed by gas phase-sequencing are described in section 2.9.1. Of the peptides analysed, the following failed to yield a detectable amino acid sequence: C1.V8.1, C1.V8.8 and C1.V8.9. These peptides may represent uncleaved protein C1, or the N- terminal peptide after cleavage, or a partial cleavage product with the N- terminal attached, or possibly a non-protein contaminant. Of the peptides that did yield a sequence, the PTH-amino acid yield are given in tables 2.9 through to table 2.12.

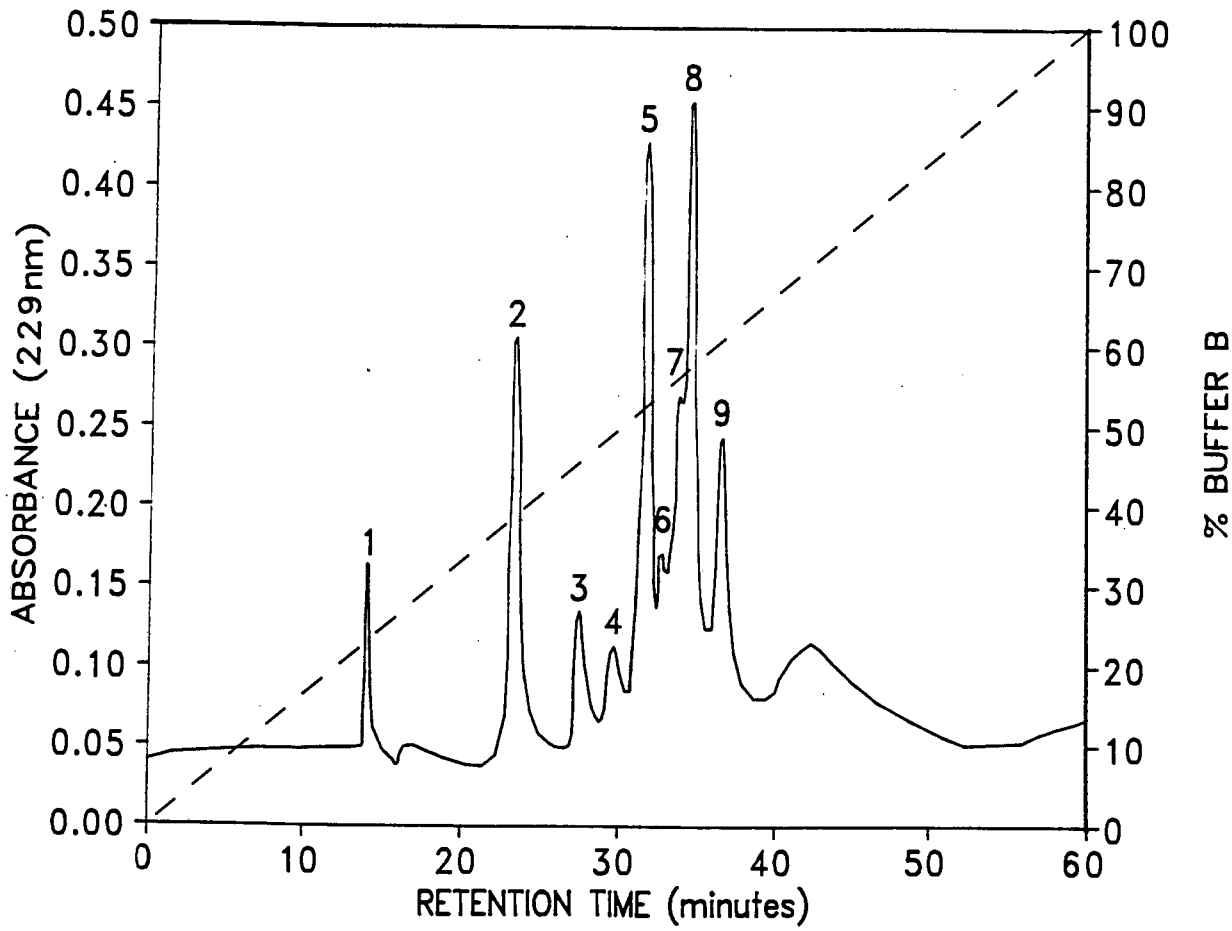

FIGURE 2.11: Reverse-phase high performance liquid chromatographic separation of Protein C1 digested with *Staphylococcus aureus* V8 protease.

Sample: 0,5 mg Protein C1; 25 µg *Staphylococcus aureus* V8 protease; 200 µL 0,05 M ammonium acetate.

Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile

Flow Rate: 0,7 mL / minute

The sequences were then compared with known calf thymus histone sequences. Excellent homology was found between protein C1 and bovine (*Bos primigenius taurus*) H2A (Sautiere *et al.*, 1974). The *O.luteus* protein C1 peptides sequenced represent 41 % of total bovine H2A sequence. Of the 41 %, a homology of 93.1 % was found between the two sequences. This would indicate that the true nature of protein C1 is that of H2A. The aligned sequences are shown below.

```
Protein C1:
Bovine H2A:   <SGRGKQGGKA   RAKAKTRSSR   AGLQFPVGRV
                        10           20           30

                                  Peptide      Peptide
                              ┌ C1.V8.4. ┐   ┌ C1.V8.3.
Protein C1:                   ·····  ···  ····
Bovine H2A:    HRFLRKGKYA   ERVGAGAPVY   LAAVLEYLTA
                        40           50           60

                  ┐ ┌─── Peptide C1.V8.2 ──────┐
Protein C1:       ··· ·····   ·····S····   ··V···V
Bovine H2A:    EILELAGNAA   RDNKKTRIIP   RHLQLAIRND
                        70           80           90

                     ┌── Peptide C1.V8.5 ──┐
Protein C1:          ·····N·   ··········   ·
Bovine H2A:    EELNKLLGKV   TIAQGGVLPN   IQAVLLPKKT  ESHHKAKGK
                       110          120          130          140
```

The notation used in the alignment is as described in section 2.7.2, except that the symbol < represents an acetylated residue.

Bovine H2A is blocked to Edman degradation due to the N- terminal amino acid being acetylated serine. It is tempting to reason that the peptides that did not yield a sequence, as well as the uncleaved *O.luteus* H2A protein, are blocked to sequencing for the same reason, although this has not been confirmed. Heterogeneity between peptides sequenced and bovine H2A was found to be (bovine sequence used here as the consensus sequence) at position 77 where threonine has changed to serine, position 84 where leucine has changed to valine, position 88 indicates a isoleucine change to valine and position 110 where lysine has changed to asparagine. For peptide C1.V8.4 to be generated under the proteolytic

conditions employed here, one would expect a glutamic residue to precede it. However, bovine H2A at position 45 indicates a glycine. It is possible, then, that there has been a mutation from glycine to glutamic acid, although this has not been confirmed. Alternatively, it is also possible that this peptide is a product of a contaminating protease. The sequence of Peptide C1.V8.3 resulted in position 62 being identified as glutamic acid. Theoretically, as this is a point of cleavage under the proteolytic conditions employed, one would not expect any further sequence from that peptide. As can be seen, this is not the case. A possible explanation is that the cleavage rate at the C- terminal side of the glutamic acid residue at position 65 is faster than that at the preceding glutamic residue. This would result in the glutamic residue at position 62 being only two residues from the peptide's C-terminus. *Staphylococcus aureus* V8 protease would not recognise this residue due to the endo-proteolytic nature of the enzyme.

Judging from the alignment of the sequences of the protein C1 peptides and the Bovine H2A, it is clear that the two proteins are closely related. It may be concluded that the C1 protein is a H2A histone.

2.10 Elucidation of the Complete Primary Structure of Protein C2

2.10.1 Production of Peptides from Protein C2

2.10.1.1 Digestion of Protein C2 with *Staphylococcus Aureus* V8 Endoproteinase

The elution profile resulting from peptide fractionation after protein C2 was digested with *Staphylococcus aureus* V8 endoproteinase is shown in figure 2.12. When the protein C2 peptide elution profile was compared with the elution patterns of sea urchin (*Parechinus angulosus*) histones digested under the same conditions, several similarities were observed (figure 2.12b). The two largest peaks in both digests elute at similar times. The first major peak elutes early (at about 35 - 40 % Buffers B) and the other major peak, which appears to be hydrophobic, is one of the last peak to elute (70 - 75 % Buffer B). Many of the elution times for the *O.luteus* protein C2 peptides correspond to peaks of the sea urchin sperm H2B digest. By comparing profiles, it would appear that protein C2 is very closely related to higher eukaryotic H2B. To determine if this was the case, the following peptides were sequenced: C2.V8.3; C2.V8.15 (the two largest peptides produced), C2.V8.6 and C2.V8.2.

2.10.1.2 Digestion of Protein C2 with Endoproteinase Arg-C.

0,5 mg of *O.luteus* protein C2 was digested with endoproteinase Arg-C (section 4.8.1.3). The resulting peptides were separated by HPLC and the resulting elution profile is shown in figure 2.13. The following peptides were selected and sequenced: C2.ER.2; C2.ER.1; C2.ER.13 and C2.ER.11.

2.10.1.3. Digestion of Protein C2 with Endoproteinase Asp-N

The proteinase Asp-N was used to digest 0,5 mg of protein C2 by the method described in (section 4.8.1.2). The resulting peptides were fractionated by HPLC and the elution profile of the peptides shown in (figure 2.14). The following peptides were selected for sequencing: C2.ED.4; C2.ED.5 and C2.ED.11.

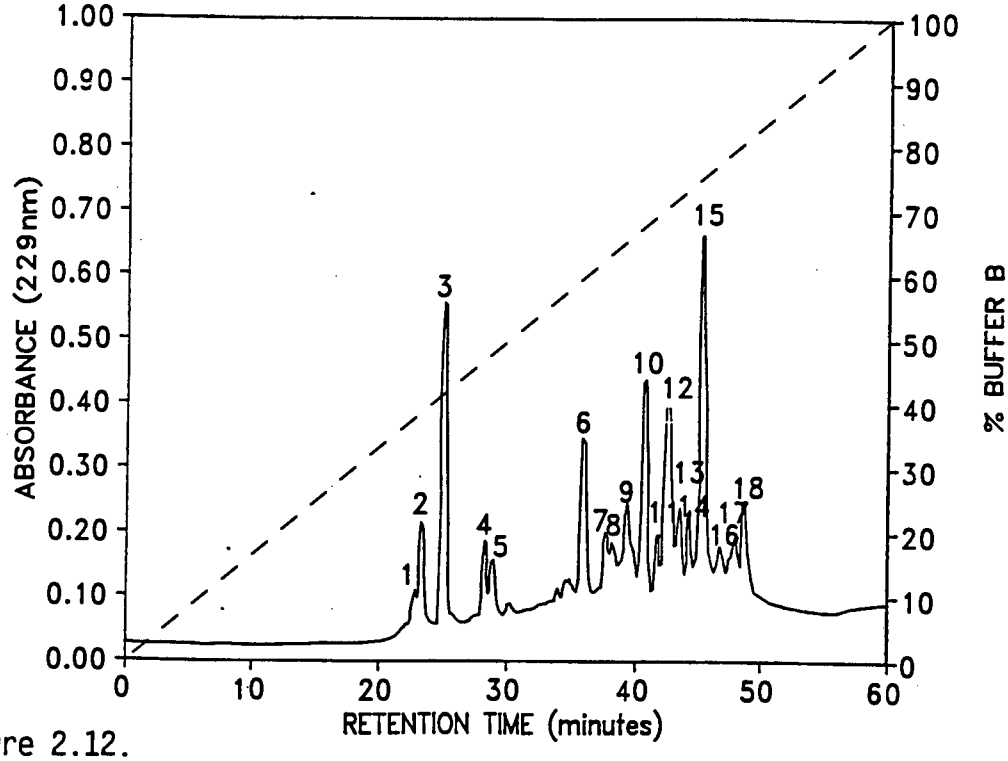

Figure 2.12.

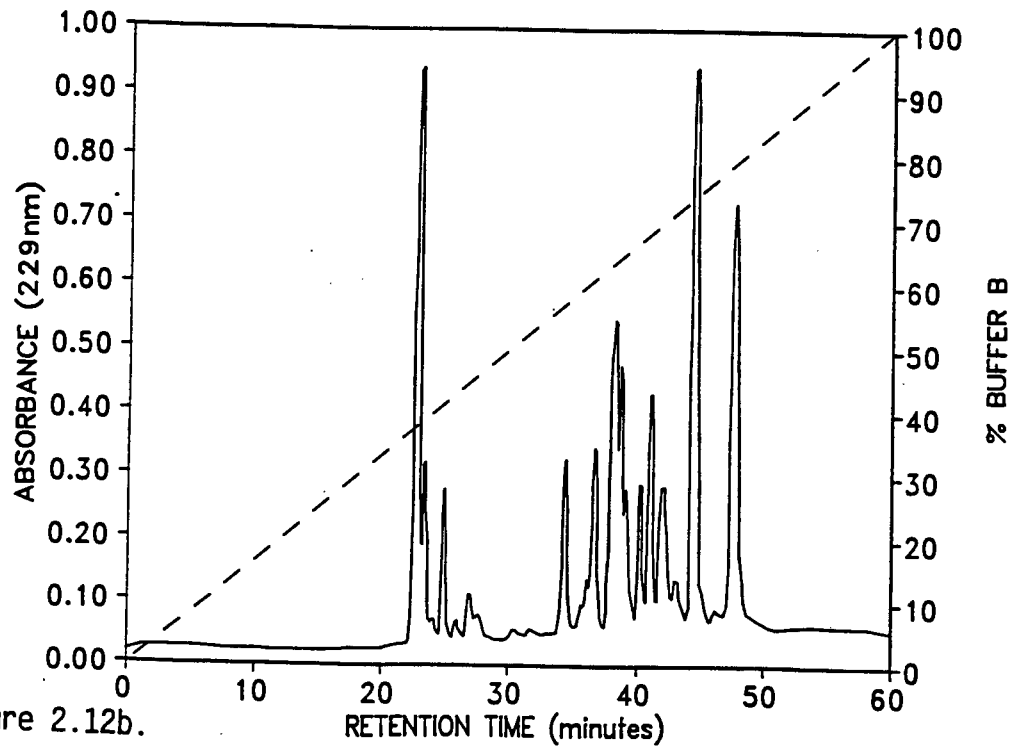

Figure 2.12b.

FIGURE 2.12 and Figure 2.12b: Reverse-phase high performance liquid chromatographic separation of Protein C2 (Figure 2.12) and sea urchin sperm (*Parechinus angulosus*) histone H2B (Figure 2.12b) digested with *Staphylococcus aureus* V8 protease.

Sample: 0,5 mg Protein C2; 25 µg *Staphylococcus aureus* V8 protease; 200 µL 0,05 M ammonium acetate (Figure 2.12).
0,5 mg sea urchin sperm H2b; 25 µg *Staphylococcus aureus* V8 protease; 200 µL 0,05 M ammonium acetate (Figure 2.12b).

Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile

Flow Rate: 0,7 mL / minute

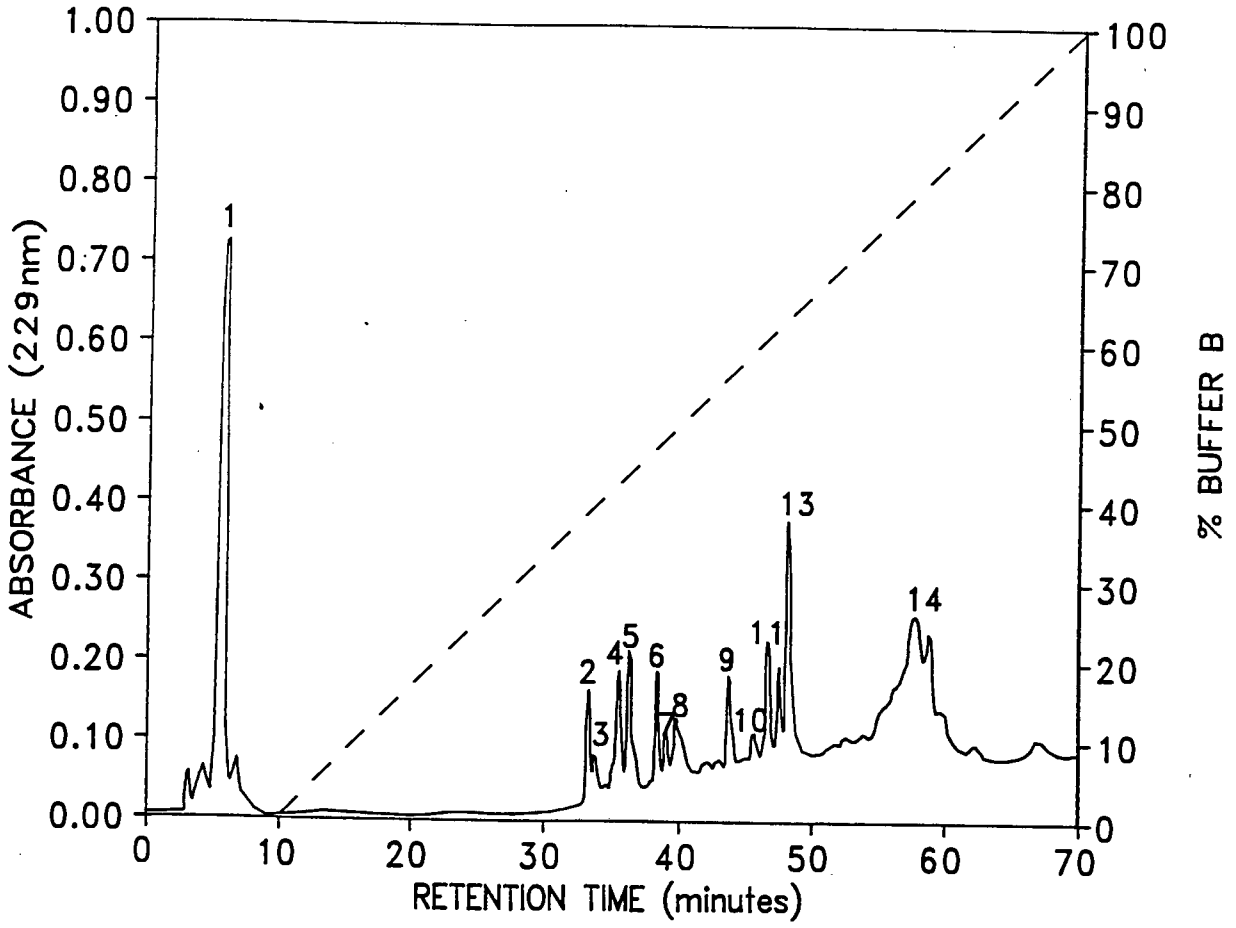

FIGURE 2.13: Reverse-phase high performance liquid chromatographic separation of Protein C2 digested with Endoproteinase Arg-C.

Sample: 0,5 mg Protein C2; 10 μg Endoproteinase Arg-C in 200 μL 10 mM Tris-HCL, pH 8,0.

Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile

Flow Rate: 0,7 mL / minute

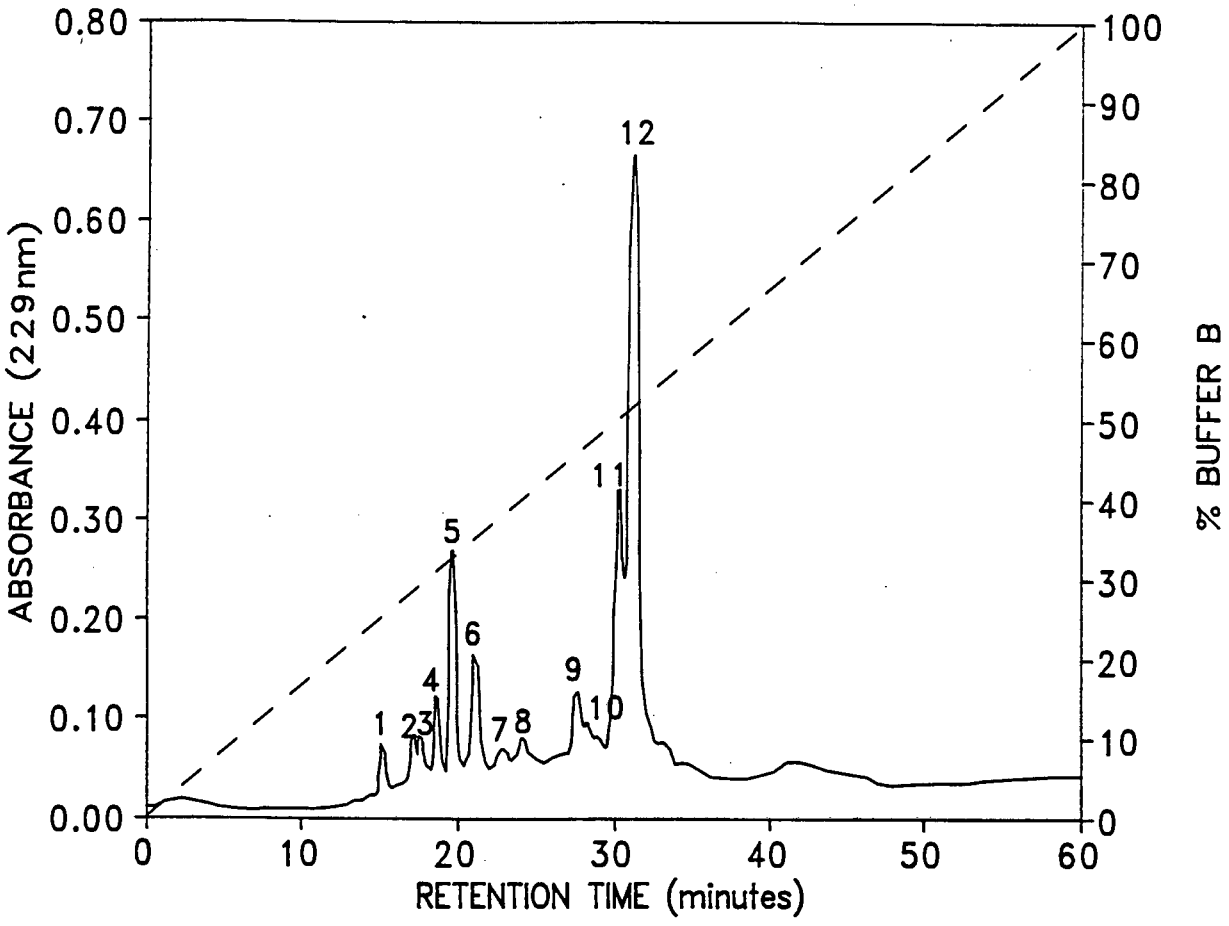

FIGURE 2.14: Reverse-phase high performance liquid chromatographic separation of Protein C2 digested with Endoproteinase Asp-N.

Sample: 0,5 mg Protein C2; 2 μg Endoproteinase Asp-N in 200 μL 10 mM Tris-HCL, pH 7,5.

Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile

Flow Rate: 0,7 mL / minute

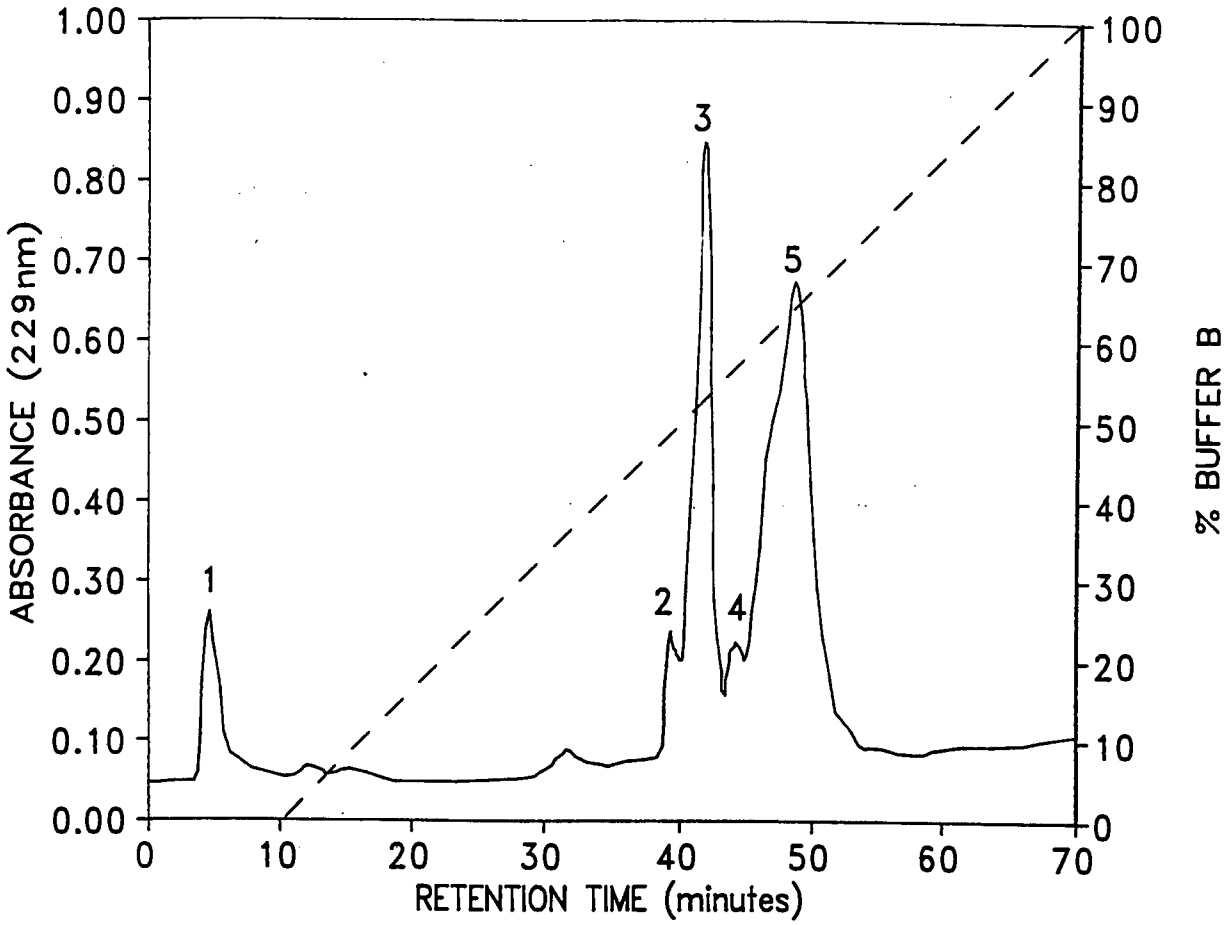

FIGURE 2.15: Reverse-phase high performance liquid chromatographic separation of Protein C2 chemically cleaved with cyanogen bromide.
Sample: 0,5 mg cleave Protein C2 in 200 µL Buffer A.
Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile
Flow Rate: 0,7 mL / minute

2.10.1.4 Chemical Cleavage of Protein C2 with Cyanogen Bromide

Cyanogen bromide was used to cleave 0,5 mg of protein C2, as described in (section 4.8.2.1). As can be seen in the HPLC elution profile (figure 2.15.), two large peptides were separated. The only peptide sequenced was C2.Cn.5.

2.10.2 The Determination of the Complete Primary Structure of Protein C2.

The complete primary structure of protein C2 has been established by sequencing uncleaved protein C2, as well as the peptides generated by *Staphylococcus aureus* V8 protease, Endo-Arginine protease, Endo-Aspartic protease and by the chemical procedure of cyanogen bromide cleavage. Of the peptides analysed by the gas-phase sequencer, only peptides C2.ER.1 (Section 2.10.1.2) and C2.V8.4 (section 2.10.1.1) did not yield an amino acid sequence. The amino acid yields, after each cycle of the Edman degradation procedure imposed on the protein and peptides, are reported in tables 2.13 through to 2.24. The peptide alignment used to determine the complete protein structure is shown in figure 2.16.

Protein C2 was aligned to bovine (*Bos primigenius taurus*) histone H2B sequenced by Iwai *et al.*, (1972), and is shown below.

```
Protein C2:            A   ·TP····A··   ·P··A·S··N
Bovine H2B:   PEPAKSAPAP   KKGSKKAVTK   AQKKDGKKRK
                      10           20           30

Protein C2:   K··V·T··S·   I·········   ·····KRG·S
Bovine H2B:   RSRKESYSVY   VYKVLKQVHP   DTGISSKAMG
                      40           50           60

Protein C2:   ·····I····   ··L·······   ·R·····LS
Bovine H2B:   IMNSFVNDIF   ERIAGEASRL   AHYNKRSTIT
                      70           80           90

Protein C2:   ··········   M·········   ··········  F··N
Bovine H2B:   SREIQTAVRL   LLPGELAKHA   VSEGTKAVTK  YTSSK
                     100          110          120     125
```

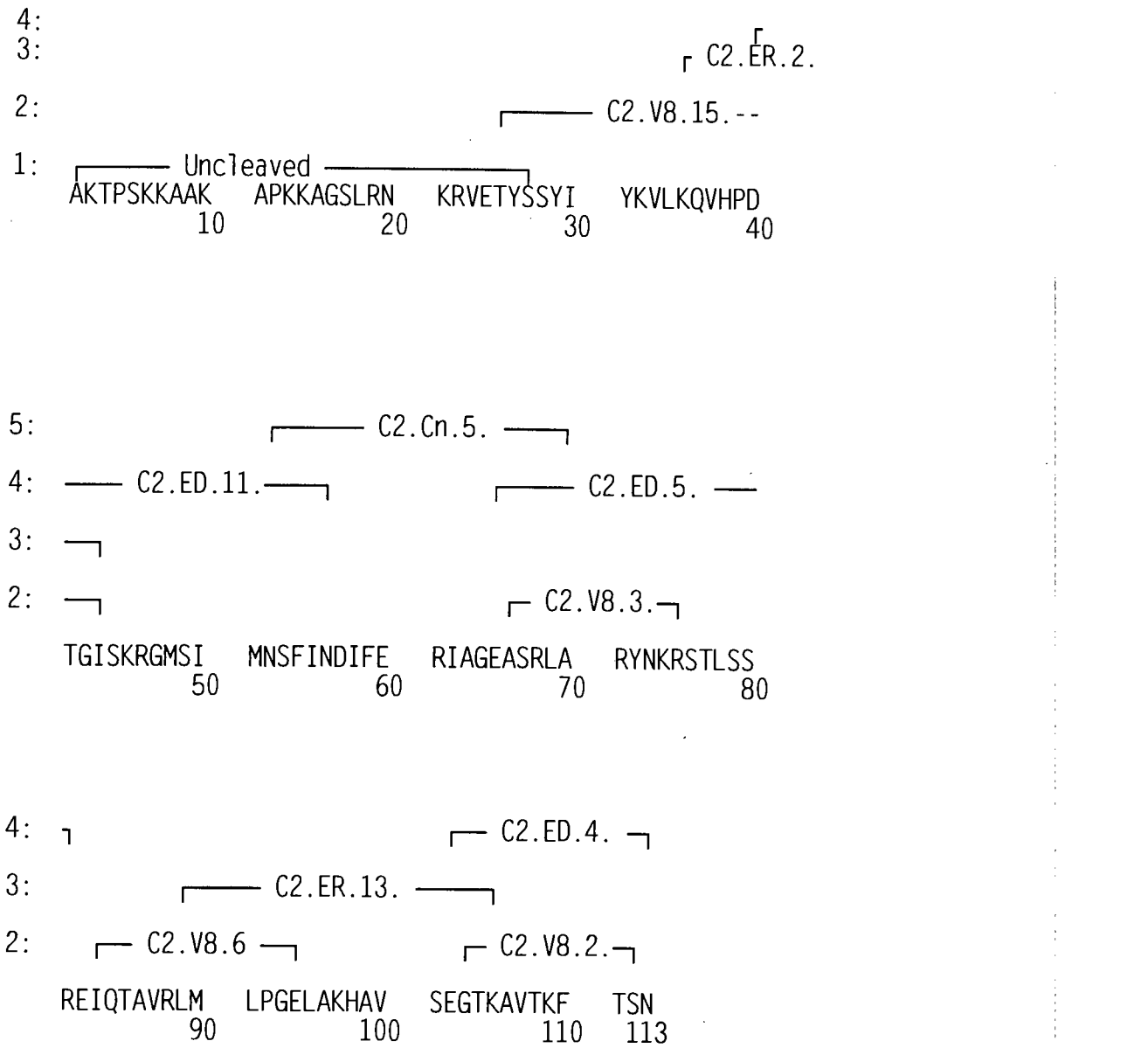

FIGURE 2.16: The complete structure of protein C2 with the peptide alignment used to solve the structure. Line 1 shows the N- terminal sequence of uncleaved protein C2; Line 2, the sequences of peptides generated by *Staphylococcus aureus* V8 digestion; line 3, peptides resulting from Endo-Arginine digestion; line 4, peptides resulting from Endo-Aspartic digestion and line 5, the peptide resulting from cleavage of protein C2 with CnBr. The first and last amino acid positions of the peptide are represented by ⌐ and ¬ respectively.

The full length of the sequenced protein C2 (H2B) is 113 amino acids, 90,4 % the length of bovine H2B. The N- terminal of *O.luteus* H2B is 11 residues shorter than its bovine counterpart. This would explain the observation that the protein migrates at a similar rate as the *O.luteus* H2A, when analysed by SDS-PAGE. The protein sequence alone cannot determine whether the trimming is gene coded or whether it is due to a specific post-transcriptional event. Due to it being highly consistent, not only as observed in this laboratory, but also in others (Rizzo, 1985), it seems unlikely that the trimming was a non-specific occurrence.
As compared with the bovine H2B, *O.luteus* H2B appears to have two deletions. One of these is at position 19, where threonine has been deleted, and the other at position 32, where serine has been deleted. As can be seen from the above alignment, there are 24 mutations. The majority of these changes (13) occur in the N- terminal region of the molecule. In the C- terminal domain, 3 mutations can be found and in the core region, 8 mutations. The homology between the two proteins (not including the trimmed N- terminal) is therefore 78 %.

The glutamic acid residue at position 93 has not been observed experimentally. The identity of this residue was deduced by the fact that
i) for Peptide C2.V8.6 to be cleaved by *Staphylococcus aureus* V8 protease, there would have to be a glutamic acid residue preceding it; and
ii) for Peptide C2.ED.5, lacking the full C- terminal to be formed by Endo-Aspartic protease, there would have to be either an apartic acid or a glutamic acid residue.
The fact that the *Staphylococcus aureus* V8 did cleave at this site, vouches for glutamic acid. This residue would not be attached to either peptide, as the Endo-Aspartic protease cleaves N- terminally of the residue and the *Staphylococcus aureus* V8 C- terminally of the residue. No overlapping peptide was found. However, when compared with other known H2B sequences, it can be seen that the sequence x-Ser-Arg-Glu-Ilu-Gln-x is conserved in all known H2B sequences. The chance that the residue in question is not glutamic acid, or that there is an insertion in the *O.luteus* sequence appears remote.

Of the endoproteases employed in this study, only *Staphylococcus aureus* V8 protease cleaved with the specificity claimed by the suppliers (Boerhinger Mannheim). Endo-Arginine protease, used under the suppliers' recommended

conditions (section 4.8.1.3), cleaved with what appears to be equal efficiency at the N- terminal of lysine residues. Similarly, Endo-Aspartic protease (section 4.8.1.2) cleaved at the N- terminal of glutamic residues. It is possible that the dual specificity of these enzymes has gone undetected by suppliers due to the use of short peptides as test substrates. If a test peptide is either too short or has a recognition site too close to one of the terminals, the enzyme will not cleave, due its endoproteolytic nature. In this way, cleavage recognition sites could be disguised and not be recognised by workers.

It may be concluded that the *O.luteus* algal nucleus not only contains the histone H2A, which appears closely related to higher eukaryotic H2As, but also the histone H2B, similar to those found in higher organisms. This study has found no evidence for the presence of the protein *HO1*, as proposed by Rizzo *et al.* (1985). The presence of the *HO1* protein, claimed to perform the function of both the H2A and the H2B histones, was based on SDS-PAGE and peptide mapping studies. It seems likely that the protein band, reported to be present in double the amounts when compared to the other major nuclear proteins when analysed SDS-PAGE, is in fact a heterogeneous band containing both the H2A and H2B histones, as reported here.

2.11 Elucidation of the Complete Structure of Protein D

2.11.1 Production of Peptides from Protein D

2.11.1.1 Digestion of Protein D with *Staphylococcus Aureus* V8 Endoproteinase

0,5 mg *O.luteus* H4 was digested with *Staphylococcus aureus* V8 endoproteinase. The resulting peptides were fractionated by HPLC. The elution profile of the protein D digest can be seen in figure 2.17. The following peptides were sequenced: D.V8.2; D.V8.3; D.V8.4 and D.V8.5.

2.11.1.2. Separation of Peptide D.V8.4

It was found that the peak labelled D.V8.4 was not homogeneous and contained at least two peptides. Two peptides were successfully separated by HPLC, utilising the HFBA buffer system (section 4.7.2) normally used to separate total nuclear proteins (figure 2.18). Separation of the two peptides under these HPLC conditions is probably due to the different arginine content in each peptide. The peptide D.V8.4b was sequenced.

2.11.1.3 Fragmentation of Peptide D.V8.3

Peptide D.V8.3 appears to be a large peptide. However, it yielded no sequence when placed on the gas-phase automatic sequencer. It was assumed that this peptide may be the N- terminal peptide, and that the N- terminal amino acid may be blocked to sequencing. Further fragmentation was attempted by incubating the peptide D.V8.3 with *Staphylococcus aureus* V8 endoproteinase, but at a higher pH

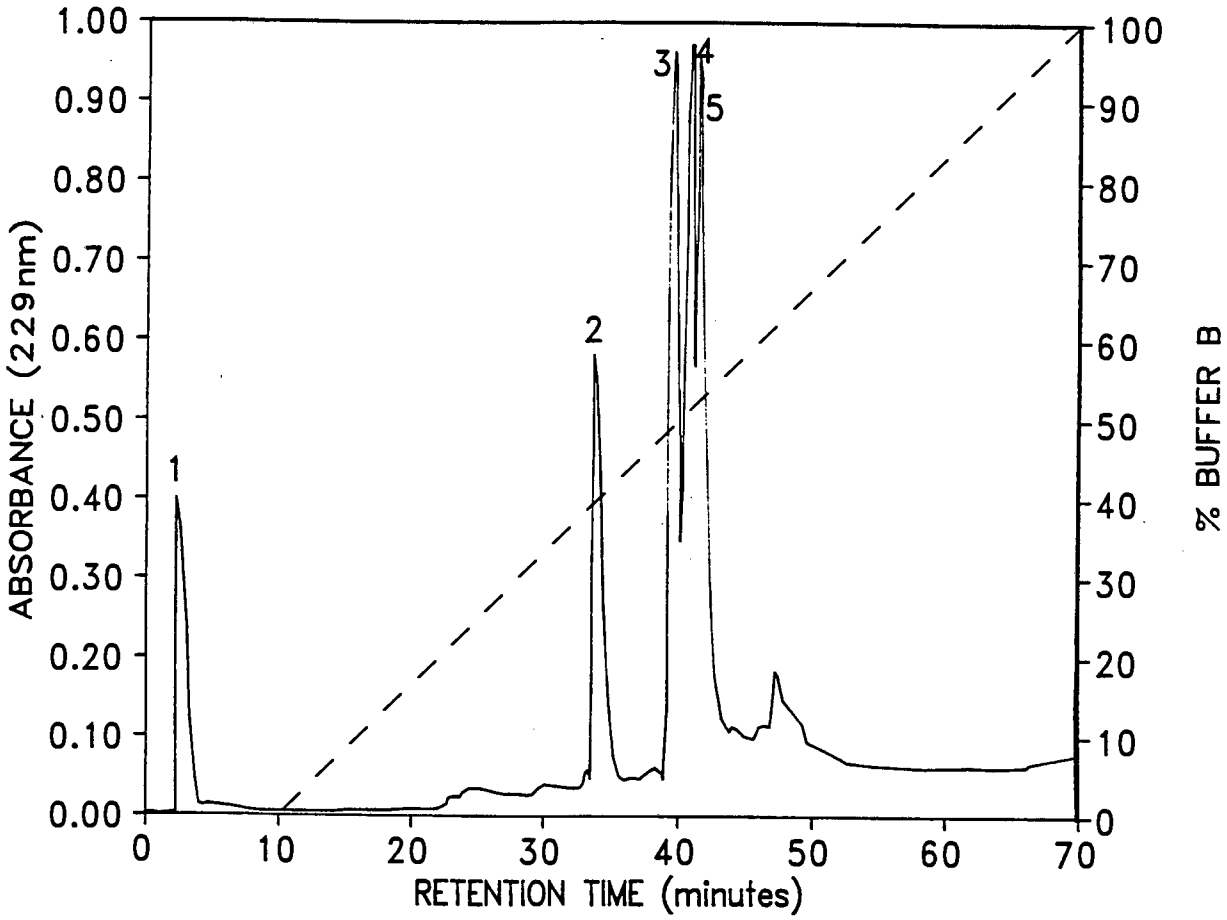

FIGURE 2.17: Reverse-phase high performance liquid chromatographic separation of Protein D digested with *Staphylococcus aureus* V8 protease.
Sample: 0,5 mg Protein D; 25 µg *Staphylococcus aureus* V8 protease; 200 µL 0,05 M ammonium acetate.
Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile
Flow Rate: 0,7 mL / minute

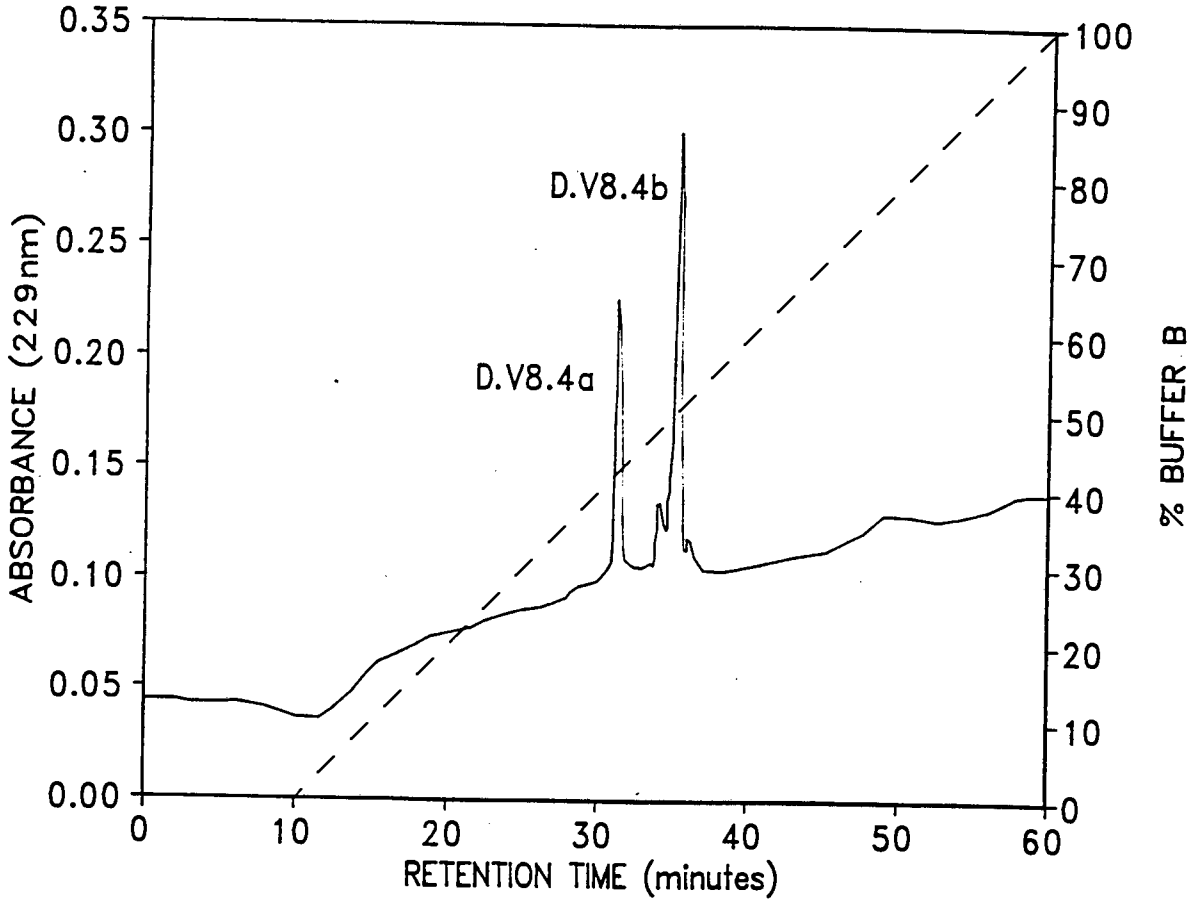

FIGURE 2.18: Reverse-phase high performance liquid chromatographic separation of Peptide D.V8.4..

Sample: Total remaining Peptide D.V8.4. in 200 μL Buffer B.

Mobile Phase: A = 0,1 % (v/v) HFBA in water.

B = 0,1 % (v/v) HFBA in 70 % (v/v) acetonitrile

Flow Rate: 0,7 mL / minute

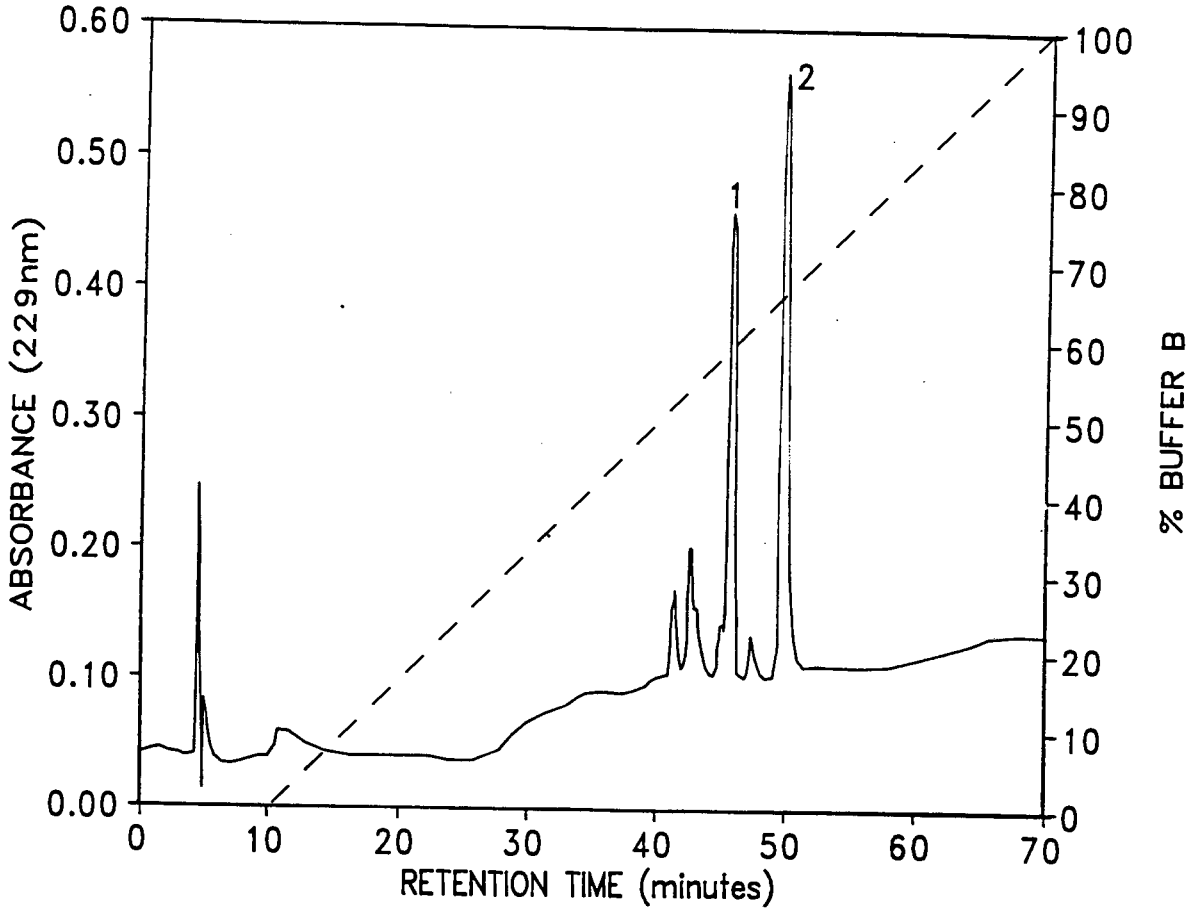

FIGURE 2.19: Reverse-phase high performance liquid chromatographic separation of Peptide D.V8.3. digested with *Staphylococcus aureus* V8 protease at high pH.

Sample: Total remaining Peptide D.V8.3.; 20 µg *Staphylococcus aureus* V8 protease dissolved in 200 µL KH_2HPO_4, pH7,0.

Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile

Flow Rate: 0,7 mL / minute

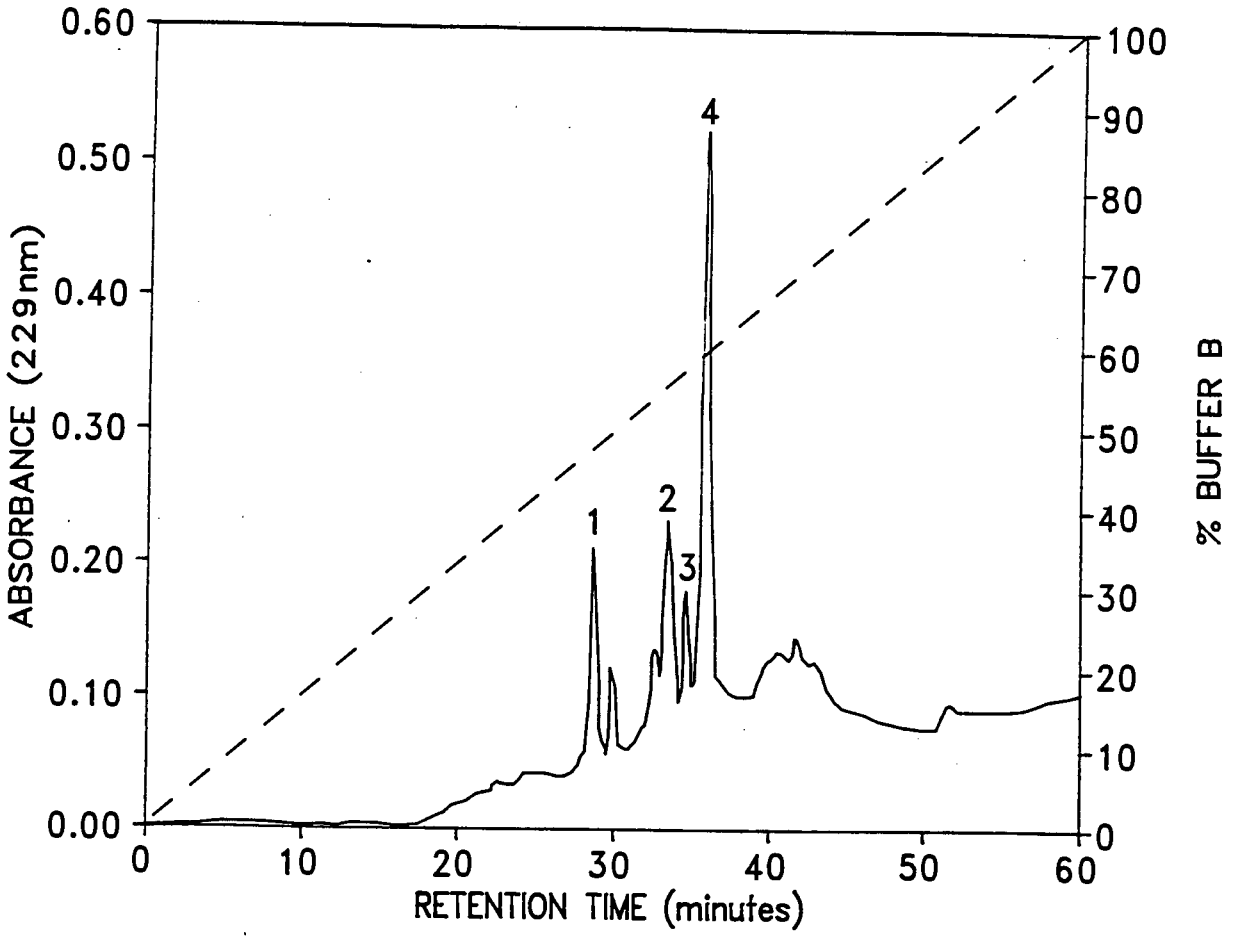

FIGURE 2.20: Reverse-phase high performance liquid chromatographic separation of Protein D chemically cleaved with heptafluorobutyric acid.
Sample: Cleaved Protein D dissolved in 200 µL Buffer B.
Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile
Flow Rate: 0,7 mL / minute

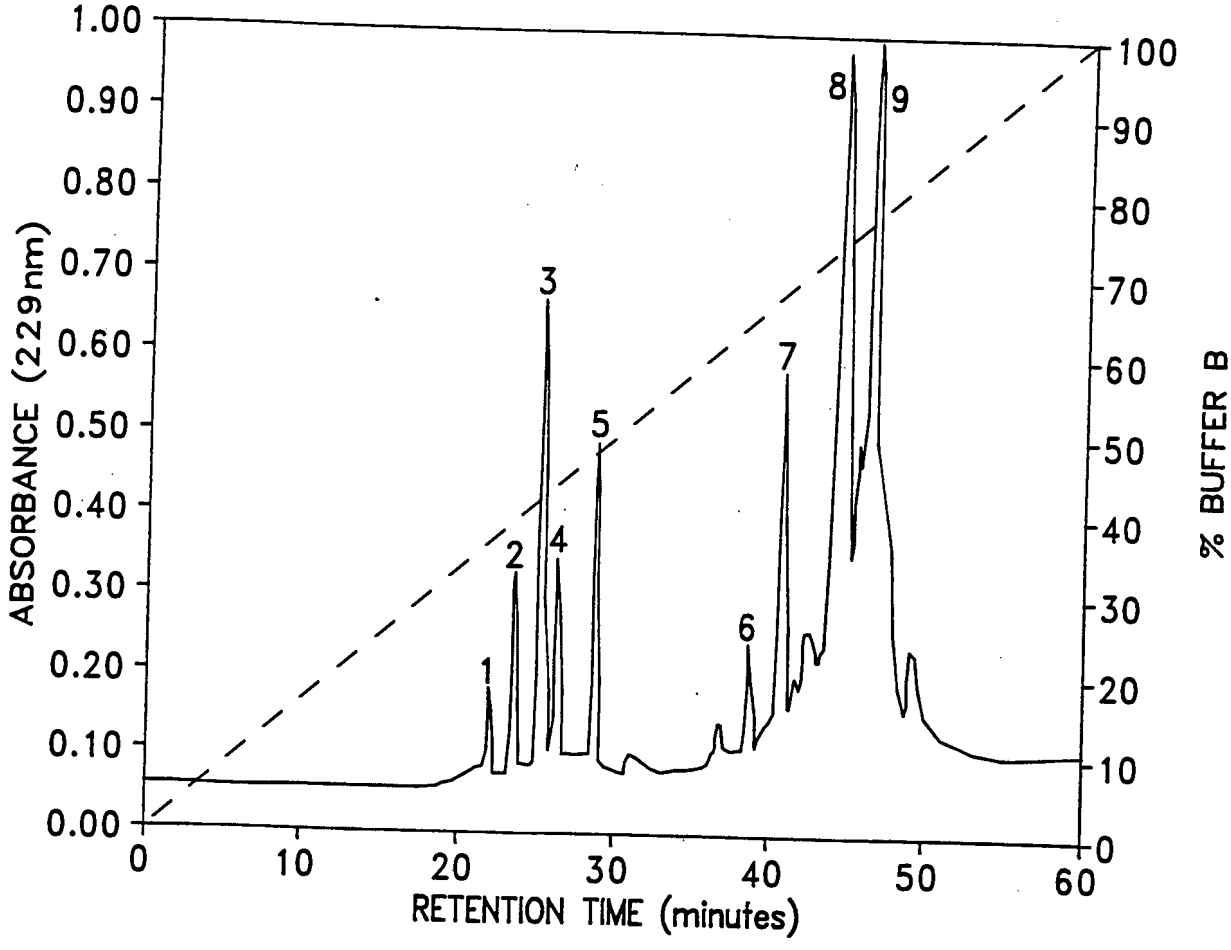

FIGURE 2.21: Reverse-phase high performance liquid chromatographic separation of Protein D digested with Endo-lysine proteinase.
Sample: 0,5 mg Protein D; 5 μg protease dissolved in 300 μL 0,1 M NH_4HCO_3, pH8,3.
Mobile Phase: A = 0,1 % (v/v) TFA in water.
B = 0,1 % (v/v) TFA in 70 % (v/v) acetonitrile
Flow Rate: 0,7 mL / minute

(section 4.8.1.1), which would result in cleavage at aspartic acid residues. The resulting elution profile, after HPLC separation, is shown in figure 2.19. Two major peptides [(D.V8.3).V8H.1 and (D.V8.3).V8H.2] were resolved and sequenced.

2.11.1.4 Chemical Cleavage of Protein D with Heptafluorobutyric Acid

0,25 mg protein D was incubated with heptafluorobutyric acid as described in (section 4.8.2.2). The HPLC elution profile after peptide separation (figure 2.20) shows that four major peptides were resolved. Only peptide D.HF.2 was sequenced.

2.11.1.5 Digestion of Protein D with Endo-Lysine Proteinase

0,5 mg protein D was digested with lysine endoproteinase as described in section 4.8.1.4. The peptide elution profile is shown in figure 2.21. Nine major peaks were resolved. The peptides that were sequenced are: D.EK.2; D.EK.3; D.EK.4; D.EK.5; D.EK.7; D.EK.8 and D.EK.9.

2.11.2 The Complete Primary Structure of Protein D

The amino acid sequence of protein D was determined by subjecting the peptides discussed in section 2.12.1 to the Edman degradation procedure (section 4.10). Generally, one-fifth of the peptides was spotted onto the sequencing glass fiber disc, unless otherwise stated. The uncleaved protein yielded no amino acid sequence when analysed by the gas-phase sequencer. This would suggest that protein D, as was the case for *O.luteus* H2A (protein C1), has a modified N-terminal residue, which precludes the possibility of utilising the Edman degradation chemistry. Of the peptides sequenced, peptides D.V8.3 and (D.V8.3).V8H.2 yielded no amino acid sequence. As the peptide (D.V8.3).V8H.1 did yield a sequence and it is a product of the D.V8.3 peptide, it is safe to assume that the D.V8.3 and (D.V8.3).V8H.2 peptides are not non-protein contaminants of

the peptide fractionation process, but rather the N- terminal peptides of protein D.

When peptide D.V8.4. was sequenced (table 2.27), the following amino acids were identified in the first five cycles:

Cycle	Amino Acid
1	Glu, Thr, His
2	Thr, Arg, Ala
3	Arg, Gly, Lys
4	Gly, Val, Arg
5	Val, Leu, Lys

It is evident from this result that at least three peptides co-eluted under the same peak when the peptides generated by *Staphylococcus aureus* V8 protease were fractionated by HPLC (figure 2.17). This peak was further fractionated as described in section 2.11.1.2. Two discrete peaks were resolved (figure 2.18). The peak D.V8.4b was sequenced and the results are shown in table 2.28. Unfortunately, due to an accident in the laboratory, the peptide D.V8.4a was lost. Lack of material precluded the possibility of repeating the experiment. However, when the sequence for the peptide D.V8.4b was subtracted from the sequencing results of the peptide peak D.V8.4 (table 2.27) the following amino acids were found per cycle:

Cycle	Amino acid
1	Glu, Thr
2	Thr, Arg
3	Arg, Gly
4	Gly, Val
5	Val, Leu
6	Leu, Lys
7	Lys, Val
8	Val, Phe
9	Phe, Leu
10	Leu, Glu
11	Glu

Although two peptides are present, it is easily deduced that the peptides are the same, except for one peptide having a Glu residue at its N- terminal. The sequences of the two peptides are therefore:

Peptide 1: Glu-Thr-Arg-Gly-Val-Leu-Lys-Val-Phe-Leu-Glu

Peptide 2: Thr-Arg-Gly-Val-Leu-Lys-Val-Phe-Leu-Glu

These peptides could only result from a protein cleaved with a C- terminal endo-glutamic acid protease, if the sequence being cleaved was -x-x-Glu-Glu-x-x-. The protease has the option of cleaving C- terminally of either glutamic acid residue. If cleavage occurs in between the two glutamic acids, one peptide would have a glutamic acid as its N- terminal. This residue would not be recognised by an endo-protease and is therefore not cleaved.

Two modified amino acids were found during the course of sequencing protein D. The fifth cycle of Edman degradation of Peptide D.V8.4b (table 2.28) resulted in 100 % di-methyl-lysine. The lysine found in the 16th cycle of peptide D.HF.2. (table 2.30) was also modified - 91,5 % of the lysine was found to be acetylated.

The amino acid yields of all the peptides sequenced are shown in tables 2.25 through to 2.36.

The complete structure of protein D was determined by aligning peptide sequences as shown in figure 2.22.

The Protein D peptide sequence was then compared with bovine histone sequences. Excellent homology was found between bovine (*Bos primigenius taurus*) histone H4 (Ogawa *et al.*, 1969; Quagliarotti *et al.*, 1969). The alignment of the *O.luteus* H4 (protein D) to bovine histone H4 is shown below. Homology to the Bovine sequence is indicated by (·). < indicates an assumed acetylated N- terminal residue. The acetylated lysine is indicated by * and the methylated lysine by +.

```
Protein D:  <·········  ·····*···R  ·········
Bovine H4:  <SGRGKGGKGL  GKGGAKRHRK  VLRDNIQGIT
                    10           20          30

Protein D:  ·········  ·········  ·········
Bovine H4:  KPAIRRLARR  GGVKRISGLI  YEETRGVLKV
                    40           50          60

Protein D:  ·········  ·····R·+·  ·········
Bovine H4:  FLENVIRDAV  TYTEHAKRKT  VTAMDVVYAL
                    70           80          90

Protein D:  ·········  ··
Bovine H4:  KRQGRTLYGF  GG
                   100
```

Protein D (H4) is 102 residues in length - exactly the same as bovine H4. Only two mutations were found: at position 20, where the bovine methyllysine has changed to arginine, and, similarly, at position 77, where lysine has also been substituted for an arginine. Both sequences have acetyllysine at position 16. However, the algal lysine appears to be almost 100 % acetylated whereas the bovine lysine is only 50 % acetylated (Ogawa *et al.*, 1969). The modification of the *O.luteus* lysine at position 77 to di-methyllysine is not shared by the corresponding lysine in the bovine H4 sequence. The acetylated serine at position 1 is homologous to both. The conservation of these amino acid modifications and molecular length would indicate that they are particularly important for maintaining the function of the histone H4 in the role it plays in chromatin formation. However, it is notable that only two differences can be found in the sequences between two evolutionary divergent creatures such as algae and bovine.

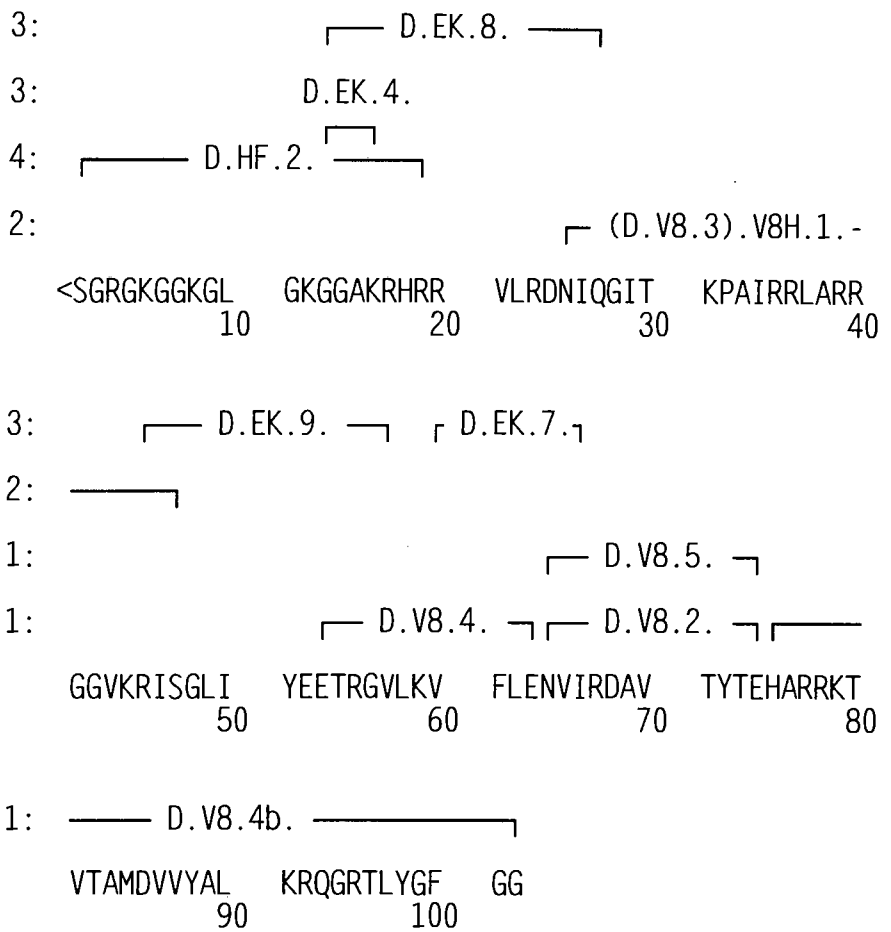

FIGURE 2.22 The complete amino acid sequence for protein D. The peptides used for the elucidation of the structure are shown to bealigned to the complete sequence. Line 1 shows peptides generated by *Staphylococcus aureus* V8 protease; Line 2, peptides generated by cleaving at aspartic residues with *Staphylococcus aureus* V8 protease at high pH, Line 3 shows peptides produced from Endo-lysine protease digestion and line 4 shows peptide incubated with HFBA. < indicates an acetylated residue. Other modified amino acids are not indicated.

2.12 Tables of Sequencing Results

		PROTEIN A1 (Uncleaved)		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Pro	-	932	243
2	Thr	28	418	83
3	Tyr	227	1449	1547
4	Tyr	1449	1547	416
5	Asp	74	960	274
6	Met	41	1267	349
7	Val	0	747	198
8	Lys	246	968	445
9	Asp	73	799	289
10	Ala	286	1260	480
11	Ile	66	509	126
12	Val	29	459	171
13	Ala	388	1029	579
14	Leu	79	872	0
15	Lys	360	1037	525
16	Asp	33	577	251
17	Arg	29	516	195
18	Asn	82	727	324
19	Gly	86	471	208
20	Ser	62	113	147
21	Ser	113	147	70
22	Met	51	422	216
23	Gln	58	325	204
24	Ala	392	672	
25	Ile	38	173	130
26	Lys	358	565	648
27	Lys	565	648	471
28	Tyr	77	255	200
29	Ile	47	128	96
30	Glu	0	172	188

TABLE 2.1: Results of Automatic Gas-Phase Edman Degradation of Protein A1 (Uncleaved). The increase above background of an amino acid (in picomoles) is assigned to position R, the background concentration of that amino acid at position R-1 and the carry over of that amino acid at position R+1.

PROTEIN A1 (Uncleaved)				
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
31	Ala	361	567	464
32	Asn	59	212	193
33	Gln	40	181	167
34	Lys	53	517	376
35	Val	17	323	85
36	Glu	169	231	215
37	Phe	11	76	84
38	Lys	38	455	432
39	Gln	127	171	163
40	His	0	44	50
41	Tyr	49	126	136
42	Leu	45	109	120
43	Arg	0	21	7
44	Ala	388	455	468
45	Ala	455	468	458
46	Leu	79	110	121
47	Lys	335	377	367

TABLE 2.1 Continued

PROTEIN A2 (Uncleaved)				
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Thr	-	383	85
2	Tyr	104	1548	1658
3	Tyr	1548	1658	311
4	Glu	46	1205	232
5	Leu	0	1438	438
6	Ile	0	484	115
7	Lys	327	1130	438
8	Ala	118	1317	1557
9	Ala	1317	1557	369
10	Ile	0	1134	201
11	Leu	0	1403	277
12	Ala	184	1337	394
13	Leu	227	1578	290
14	Lys	554	1181	584
15	Glu	158	523	203
16	Arg	0	480	146
17	Asn	39	861	182
18	Gly	63	416	157
19	Ser	0	132	151
20	Ser	132	151	85
21	Ala	235	582	262
22	Gln	65	510	148
23	Ala	245	651	308
24	Ile	66	475	145
25	Lys	350	548	765
26	Lys	548	765	530
27	Tyr	70	347	181
28	Ile	0	202	98
29	Leu	0	837	209
30	Glu	103	173	186
31	Asn	46	243	367
32	Asn	243	367	159
33	Lys	398	455	447
34	Ile	0	151	147
35	Glu	103	170	157

TABLE 2.2: Results of Automatic Gas-Phase Edman Degradation of Protein A2 (Uncleaved)

PROTEIN A2 (Uncleaved)

Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
36	Phe	0	250	165
37	Gln	54	184	211
38	Gln	184	211	105
39	Thr	0	164	76
40	Phe	51	166	184
41	Leu	0	271	200
42	Arg	0	100	172
43	X			
44	Ala	212	256	242
45	Leu	22	175	154
46	Lys	365	502	449

TABLE 2.2 Continued

		PROTEIN B (Uncleaved)		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Ala	-	1344	158
2	Arg	4	417	44
3	Thr	43	449	49
4	Lys	100	1657	143
5	Gln	24	837	48
6	Thr	78	332	125
7	Ala	196	1006	342
8	Arg	54	717	167
9	Lys	254	1024[1]	319
10	Ser	262	527	236
11	Thr	0	247	61
12	Gly	177	661	700
13	Gly	661	700	273
14	Lys	181	744	227
15	Ala	400	682	259
16	Pro	109	325	170
17	Arg	125	423	204
18	Lys	195	496	277
19	Gln	21	327	24
20	Leu	220	501	269
21	Ala	277	554	363
22	Thr	53	184	69
23	Lys	184	330	217
24	Ala	320	492	504
25	Ala	492	504	314
26	Arg	162	234	308
27	Lys	144	190	176
28	Ser	0	141	84
29	Ala	300	363	269
30	Pro	69	146	113
31	Ala	269	284	225

1 175 pico moles lysine and 849 picomoles mono-methlyated lysine

TABLE 2.4 Results of Automatic Gas-Phase Edman Degradation of Peptide B (Uncleaved)

		PEPTIDE B.V8.1.		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Ile	-	760	0
2	Arg	16	187	267
3	Arg	187	267	85
4	Tyr	0	1011	107
5	Gln	19	948	90
6	Lys	41	780	135
7	Ser	0	396	86
8	Thr	0	381	59
9	Glu	0	104	-

TABLE 2.5 Results of Automatic Gas-Phase Edman Degradation of Peptide B.V8.1

		PEPTIDE B.V8.3.		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Asp	-	139	37
2	Thr	0	81	30
3	Asn	0	201	69
4	Leu	22	214	100
5	Cys	33	233[1]	83
6	Ala	51	225	156
7	Ile	0	81	52
8	His	0	104	0
9	Ala	55	174	160
10	Lys	46	131	42
11	Arg	83	235	98
12	Val	85	125	0
13	Thr	0	79	10
14	Ile	0	75	94
15	Met	0	155	106
16	Pro	0	53	0

1 As judged at 313 nm, and compared relative to internal standard, norleucine. No cystines is detected if DTE is present in the sequencing buffers.

TABLE 2.6 Results of Automatic Gas-Phase Edman Degradation of Peptide B.V8.3

		PEPTIDE B.V8.6.		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Leu	29	1273	1303
2	Leu	1273	1303	88
3	Ile	0	583	73
4	Arg	53	709	116
5	Lys	0	897	141
6	Leu	0	949	139
7	Pro	0	430	221
8	Phe	0	493	279
9	Gln	0	395	194
10	Arg	78	539	339
11	Leu	26	542	282
12	Val	0	228	142
13	Arg	208	427	363
14	Glu	30	11	53

TABLE 2.7. Results of Automatic Gas-Phase Edman Degradation of Peptide B.V8.6.

		PEPTIDE B.V8.7.		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Ala	-	804	147
2	Tyr	0	558	42
3	Leu	118	706	91
4	Val	6	392	53
5	Gly	36	482	181
6	Leu	59	501	159
7	Phe	37	386	139
8	Glu	48	59	54

TABLE 2.8 Results of Automatic Gas-Phase Edman Degradation of Peptide B.V8.7

		PEPTIDE C1.V8.2.		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Leu	-	270	57
2	Ala	13	439	51
3	Gly	18	186	48
4	Asn	3	131	57
5	Ala	12	178	234
6	Ala	178	234	75
7	Arg	3	146	19
8	Asp	0	163	2
9	Asn	0	86	34
10	Lys	0	45	54
11	Lys	45	54	15
12	Ser	2	34	17
13	Arg	1	85	15
14	Ile	0	76	91
15	Ile	76	91	45
16	Pro	0	37	19
17	Arg	0	43	3
18	His	7	92	8
19	Val	0	39	16
20	Gln	0	12	6
21	Leu	0	12	5
22	Ala	0	19	1
23	Val	0	28	-

TABLE 2.9 Results of Automatic Gas-Phase Edman Degradation of Peptide C1.V8.2

		PEPTIDE C1.V8.3		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Tyr	-	300	32
2	Leu	21	203	17
3	Thr	3	134	32
4	Ala	14	345	48
5	Glu	15	133	9
6	Ile	2	69	8
7	Leu	0	20	-

TABLE 2.10 Results of Automatic Gas-Phase Edman Degradation of Peptide C1.V8.3

		PEPTIDE C1.V8.4		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Ala	-	133	75
2	Gly	72	196	55
3	Ala	75	203	89
4	Pro	16	135	56
5	Val	51	173	74
6	Tyr	78	156	77
7	Leu	42	115	52
8	Ala	98	129	144
9	Ala	129	144	-

TABLE 2.11 Results of Automatic Gas-Phase Edman Degradation of Peptide C.V8.4

		PEPTIDE C1.V8.5		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Leu	-	189	17
2	Asn	0	171	36
3	Lys	1	170	33
4	Leu	1	230	243
5	Leu	230	243	35
6	Gly	14	213	67
7	Asn	1	136	35
8	Val	1	228	51
9	Thr	1	125	34
10	Ile	0	138	9
11	Ala	12	149	64
12	Gln	1	18	2
13	Gly	11	96	85
14	Gly	96	85	36
15	Val	2	103	32
16	Leu	2	60	24
17	Pro	1	59	11
18	Asn	0	40	17
19	Ile	2	64	-

TABLE 2.12 Results of Automatic Gas-Phase Edman Degradation of Peptide C1.V8.5

PROTEIN C2 (UNCLEAVED)				
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Ala	-	679	89
2	Lys	22	287	102
3	Thr	25	181	21
4	Pro	21	290	69
5	Ser	21	96	28
6	Lys	81	414	499
7	Lys	414	499	175
8	Ala	95	193	459
9	Ala	193	459	170
10	Lys	119	350	144
11	Ala	170	399	202
12	Pro	30	143	86
13	Lys	118	254	354
14	Lys	254	354	214
15	Ala	147	258	179
16	Gly	91	151	122
17	Ser	17	41	31
18	Lys	123	220	153
19	Arg	16	33	29
20	Asn	50	105	83
21	Lys	117	143	114
22	Arg	22	39	35
23	Val	20	38	37
24	Glu	33	45	42
25	Thr	11	17	15
26	Tyr	24	56	17
27	Ser	0	18	21
28	Ser	18	21	-

TABLE 2.13 Results of Automatic Gas-Phase Edman Degradation of Protein C2 (Uncleaved)

		PEPTIDE C2.V8.2		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Gly	-	579	94
2	Thr	17	226	75
3	Lys	0	160	56
4	Ala	0	412	124
5	Val	0	354	129
6	Thr	8	145	58
7	Lys	4	120	64
8	Phe	0	259	91
9	Thr	21	108	60
10	Ser	0	71	0

TABLE 2.14 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.V8.2

		PEPTIDE C2.V8.3		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Ala	-	195	45
2	Ser	31	119	59
3	Arg	0	79	30
4	Leu	0	66	23
5	Ala	0	105	75
6	Arg	0	82	51
7	Tyr	0	87	48
8	Asn	0	42	36
9	Lys	0	32	16
10	Arg	14	54	29

TABLE 2.15 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.V8.3

		PEPTIDE C2.V8.6		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Ile	-	295	22
2	Gln	9	215	43
3	Thr	0	92	23
4	Ala	7	226	50
5	Val	0	178	64
6	Arg	0	74	26
7	Leu	0	163	62
8	Met	0	255	17
9	Leu	62	183	70
10	Pro	0	162	45
11	Gly	7	53	41
12	Glu	5	58	-

TABLE 2.16 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.V8.6

		PEPTIDE C2.V8.15		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Thr	-	80	11
2	Tyr	15	146	41
3	Ser	5	30	31
4	Ser	30	31	5
5	Tyr	10	88	30
6	Ile	0	19	12
7	Tyr	30	99	27
8	Lys	0	5	3
9	Val	0	76	26
10	Leu	0	7	5
11	Lys	0	9	4
12	Gln	0	43	10
13	Val	0	41	18
14	His	0	18	0
15	Pro	0	35	13
16	Asp	44	49	38
17	Thr	0	8	6
18	Gly	8	27	16
19	Ile	0	12	0

TABLE 2.17 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.V8.15

		PEPTIDE C2.ER.2		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Lys	-	38	17
2	Gln	0	125	17
3	Val	0	143	23
4	His	0	86	44
5	Pro	18	94	43
6	Asp	3	103	1
7	Thr	27	35	18
8	Gly	5	51	23
9	Ile	0	44	-

TABLE 2.18 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.ER.2

		PEPTIDE C2.ER.11		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Leu	-	8	1
2	Met	0	17	1
3	Leu	1	9	2

TABLE 2.19 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.ER.11

		PEPTIDE C2.ER.13		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Leu	-	269	5
2	Met	0	411	20
3	Leu	5	271	13
4	Pro	0	276	16
5	Gly	4	205	52
6	Glu	0	195	49
7	Leu	0	188	40
8	Ala	7	192	53
9	Lys	1	139	51
10	His	0	259	86
11	Ala	22	177	55
12	Val	0	129	46
13	Ser	0	78	24
14	Glu	5	59	30
15	Gly	4	56	27
16	Thr	0	49	16
17	Lys	1	16	-

TABLE 2.20 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.ER.13

		PEPTIDE C2.ED.5		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Glu	-	1371	164
2	Ala	43	1305	148
3	Ser	0	45	0
4	Arg	3	419	84
5	Leu	35	749	107
6	Ala	156	942	149
7	Arg	29	412	89
8	Tyr	14	590	138
9	Asn	4	306	7
10	Lys	3	71	11
11	Arg	19	160	49
12	Ser	8	57	18
13	Thr	8	104	33
14	Leu	4	110	33
15	Ser	18	28	32
16	Ser	28	32	15
17	Arg	3	24	13

TABLE 2.21 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.ED.5

		PEPTIDE C2.ED.4		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Glu	-	771	109
2	Gly	9	507	86
3	Thr	15	254	41
4	Lys	7	226	29
5	Ala	21	611	118
6	Val	6	230	41
7	Thr	6	168	40
8	Lys	0	124	20
9	Phe	2	354	68
10	Thr	17	157	41
11	Ser	0	83	21
12	Asn	19	124	67

TABLE 2.22 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.ED.4

		PEPTIDE C2.ED.11		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Asp	-	842	65
2	Thr	0	284	42
3	Gly	2	551	60
4	Ile	17	290	50
5	Ser	45	176	27
6	Lys	4	273	31
7	Arg	27	486	80
8	Gly	17	465	89
9	Met	5	460	61
10	Ser	16	143	24
11	Ile	4	113	19
12	Met	24	313	54
13	Asn	10	300	62
14	Ser	6	65	13
15	Phe	3	144	34
16	Ile	3	52	10
17	Asn	18	64	41

TABLE 2.23 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.ED.11

		PEPTIDE C2.Cn.5		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Asn	-	1972	439
2	Ser	98	458	98
3	Phe	44	1259	224
4	Ile	9	406	103
5	Asn	210	1019	292
6	Asp	84	1053	302
7	Ile	30	336	112
8	Phe	40	643	257
9	Glu	68	488	63
10	Arg	8	271	56
11	Leu	144	535	255
12	Ala	361	631	372
13	Gly	93	372	268
14	Glu	76	229	176
15	Ala	268	573	396
16	Ser	0	33	31
17	Arg	17	55	52
18	Leu	61	133	128
19	Ala	188	252	-

TABLE 2.24 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.Cn.5

		PEPTIDE D.V8.2		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Asn	-	504	40
2	Val	0	1153	115
3	Ile	4	457	25
4	Arg	0	37	6
5	Asp	2	498	243
6	Ala	0	675	199
7	Val	4	372	169
8	Thr	0	202	191
9	Tyr	0	310	227
10	Thr	191	345	212
11	Glu	22	47	-

TABLE 2.25 Results of Automatic Gas-Phase Edman Degradation of Peptide D.V8.2

		PEPTIDE D.V8.5		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Asn	-	88	28
2	Val	0	249	59
3	Ile	0	105	5
4	Arg	0	188	5
5	Asp	47	182	100
6	Ala	20	208	86
7	Val	0	304	131
8	Thr	14	91	50
9	Tyr	0	160	81
10	Thr	50	76	34
11	Glu	10	132	-

TABLE 2.26 Results of Automatic Gas-Phase Edman Degradation of Peptide D.V8.5

		PEPTIDE D.V8.4		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Glu	-	251	44
	Thr	-	67	103
	His	-	57	41
2	Thr	67	103	20
	Arg	30	73	44
	Ala	20	145	17
3	Arg	73	44	12
	Gly	1	62	126
	Lys	1	5	1
4	Gly	62	126	64
	Val	16	142	211
	Arg	44	12	5
5	Val	142	211	78
	Leu	1	57	147
	Lys	5	1	19
6	Leu	57	147	45
	Lys	1	19	57
	Thr	2	58	10
7	Lys	1	57	24
	Val	78	139	133
8	Val	139	133	47
	Phe	0	28	83
	Thr	10	29	14
9	Phe	28	83	28
	Leu	0	34	53
	Ala	1	61	24
10	Leu	34	53	28
	Glu	0	9	18
	Met	0	22	14
11	Glu	9	18	19
	Leu	53	28	13

TABLE 2.27 Results of Automatic Gas-Phase Edman Degradation of Peptide D.V8.4

		PEPTIDE D.V8.4b		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	His	-	206	25
2	Ala	23	182	66
3	Arg	6	124	404
4	Arg	124	404	159
5	Lys	0	48[1]	13
6	Thr	2	96	32
7	Val	0	194	126
8	Thr	32	76	37
9	Ala	17	123	50
10	Met	5	165	92
11	Asp	38	109	102
12	Val	8	91	173
13	Val	91	173	111
14	Tyr	0	116	72
15	Ala	1	110	98
16	Leu	1	90[2]	80
17	Lys	0	21	32
18	Arg	0	55	92
19	Gln	0	1	34
20	Gly	8	32	55
21	Arg	5	35	67
22	Thr	14	34	28
23	Leu	15	18	46
24	Tyr	11	35	54
25	Gly	74	95	82
26	Phe	0	15	34
27	Gly	82	96	83
28	Gly	96	83	77

1 100 % di-methyl lysine
2 Carry-over increases and becomes greater than the "in-step" cleavage.

TABLE 2.28 Results of Automatic Gas-Phase Edman Degradation of Peptide D.V8.4b

PEPTIDE (D.V8.3).V8H.1				
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Asn	-	334	73
2	Ile	11	433	76
3	Gln	0	362	73
4	Gly	13	288	86
5	Ile	10	180	94
6	Thr	0	92	50
7	Lys	0	147	71
8	Pro	0	136	82
9	Ala	1	170	109
10	Ile	0	53	83
11	Arg	0	124	262
12	Arg	124	262	102
13	Leu	0	83	50
14	Ala	16	89	81
15	Arg	43	65	141
16	Arg	65	141	112
17	Gly	13	55	92
18	Gly	55	92	19
19	Val	0	32	4
20	Lys	0	5	1
21	Arg	5	63	29
22	Ile	4	15	14
23	Ser	0	25	2

TABLE 2.29 Results of Automatic Gas-Phase Edman Degradation of Peptide (D.V8.3).V8H.1

Residue Number	Amino Acid	PEPTIDE D.HF.2 Yield in picomoles		
		R - 1	R	R + 1
1	Ser	-	2748	204
2	Gly	112	1879	257
3	Arg	3	746	213
4	Gly	257	2034	337
5	Lys	20	1672	239
6	Gly	337	2194	2187
7	Gly	2194	2187	324
8	Lys	110	1424	275
9	Gly	324	1923	316
10	Leu	35	1376	203
11	Gly	316	1514	321
12	Lys	55	1259	268
13	Gly	321	1287	1657
14	Gly	1287	1657	542
15	Ala	59	1427	455
16	Lys	42	1036[1]	408
17	Arg	9	1618	661
18	His	6	1009	697
19	Arg	661	1633	891

1 91,5 % acetylated lysine

TABLE 2.30 Results of Automatic Gas-Phase Edman Degradation of Peptide D.HF.2

Residue Number	Amino Acid	PEPTIDE D.EK.2. Yield in picomoles		
		R - 1	R	R + 1
1	Gly	-	432	1105
2	Gly	432	1105	48
3	Lys	15	2055[1]	114

1 90 % acetylated lysine.

TABLE 2.31 Results of Automatic Gas-Phase Edman Degradation of Peptide D.EK.2

		PEPTIDE D.EK.3		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Arg	-	578	156
2	Gln	1	531	101
3	Gly	14	406	204
4	Arg	25	437	51
5	Thr	0	123	65
6	Leu	0	191	127
7	Tyr	2	196	169
8	Gly	21	166	141
9	Phe	0	97	99
10	Gly	141	172	166
11	Gly	172	166	-

TABLE 2.32 Results of Automatic Gas-Phase Edman Degradation of Peptide D.EK.3

		PEPTIDE D.EK.4		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Gly	-	60	57
2	Gly	60	57	41
3	Ala	16	65	21
4	Lys	0	36	13

TABLE 2.33 Results of Automatic Gas-Phase Edman Degradation of Peptide D.EK.4

		PEPTIDE D.EK.7.		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Val	-	28	6
2	Phe	0	19	2
3	Leu	0	20	2
4	Glu	0	17	6
5	Asn	0	14	9
6	Val	5	25	14
7	Ile	0	4	-

TABLE 2.34 Results of Automatic Gas-Phase Edman Degradation of Peptide D.EK.7

		PEPTIDE D.EK.8		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Gly	-	1396	1363
2	Gly	1396	1363	275
3	Ala	94	895	125
4	Lys	0	581	34
5	Arg	63	566	102
6	His	0	529	129
7	Arg	102	915	483
8	Arg	915	483	157
9	Val	84	610	95
10	Leu	42	424	98
11	Arg	171	1019	602
12	Asp	74	539	88
13	Asn	53	288	127
14	Ile	54	494	76
15	Gln	68	288	-

TABLE 2.35 Results of Automatic Gas-Phase Edman Degradation of Peptide C2.ED.11

		PEPTIDE D.EK.9		
Residue Number	Amino Acid	Yield in picomoles		
		R - 1	R	R + 1
1	Arg	-	60	25
2	Ile	0	231	12
3	Ser	0	288	137
4	Gly	13	181	67
5	Leu	13	178	53
6	Ile	3	148	47
7	Tyr	5	218	51
8	Glu	5	123	144
9	Glu	123	144	62
10	Thr	23	85	55
11	Arg	3	56	26
12	Gly	29	89	51
13	Val	19	101	-

TABLE 2.36 Results of Automatic Gas-Phase Edman Degradation of Peptide D.EK.9

CHAPTER 3

Conclusions and Discussions

The results discussed in Chapter 2 clearly demonstrate that the nucleus of the alga *Olisthodiscus luteus* contains the five families of histone proteins found in all higher eukaryotes. By partial sequence analysis the core histones H2A and H3 have been positively identified. Similarly, two variants of H1 have been partially sequenced. The full primary structure of the H2B and H4 histone proteins has been elucidated. No evidence in support of the existence of the histone H*O1*, as proposed by Rizzo *et al.* (1985), was found.

In this chapter, the partial or complete structures of the *Olishthodiscus luteus* histones will be discussed and compared with other known sequences. To achieve this, the Genetics Computer Group sequence analysis software package (GCG program version 7), purchased from Genetics Computer Group Inc., was used in conjunction with the "Swissprot" protein data bank (version 21.0).

3.1 Genetic Computer Group Multiple Sequence Analysis

3.1.1 Data Banks

Mini data banks were constructed by searching the "Swissprot" protein data bank (version 21.0) for each type of histone (e.g. H4) and deleting all incomplete structures. To each histone family data bank, the corresponding partial or complete *Olisthodiscus luteus* histone sequences were entered.

3.1.2 Multiple sequence Alignment

In performing multiple sequence alignment analysis of the histone family data banks, the assumption that all histone sequences within a family data bank share a common ancestor has been taken. The GCG program multiple sequence alignment procedure ("PileUp" in GCG format) follows the method of Feng and Doolittle

(1987). These authors follow the notion that more trust can be placed in the comparison of more recently diverging sequences, than those that had diverged earlier. Furthermore, they propose that if a gap had to occur between divergent protein sequences due to an insertion or deletion, that gap will always be present in the progressive divergence of these two sequences. In the words of Feng and Doolittle: "Once a gap, always a gap."

The multiple sequence alignment procedure progressively aligns two sequences and determines a comparative score for each possible alignment, the best alignment being that with the highest score. The score is determined by arithmetic addition of values derived at each position along the alignment. These values are determined by comparison of the two aligned sequence symbols to the GCG default symbol comparison table. This symbol comparison table, the same as used by Gribskov and Burgess (1986), has been derived from the Mutational Difference Matrix (MDM) of Schwartz and Dayhoff, (1979). Gribskov and Burgess (1986) have rescaled the MDM in such away that identical residues score a value of 1,5. The score for non-identical comparisons have been adjusted to a mean of -0,17, with a standard deviation of 0,364. For example, the resulting value in comparing a Phe ↔ Tyr mutation is 1,491 while an Ala ↔ Trp mutation scores -0,677.
Having determined the best alignment score between each sequence, the program forms pair clusters of similar sequences. Each cluster is then aligned to the next most related cluster as determined by the scoring process (the comparative score between any two positions in the clusters is the average of the scores for all possible symbol comparisons at that position). By this pairwise comparison, the formation of a dendogram is achieved. Final sequence alignment of sequence is achieved by pairwise alignments of increasingly dissimilar sequences, as dictated by the dendogram.

When required, the GCG program inserts gaps into a sequence to maximise the alignment score. When this occurs, the same sized gap is inserted in the same position of all the sequences in that cluster. This is in keeping with Feng and Doolittle's (1987) concept: "Once a gap, always a gap". Following Feng and Doolittle's notion that early divergent sequence, comparisons carry less integrity than those of later divergent sequences and that the requirement to insert gaps would occur most likely during the alignment of early divergent sequences, the score of such an alignment is penalised. The GCG program default for a gap

insertion penalty holds a value of 3, with a further 0,1 each time the gap is lengthened by one residue to maximise alignment.

In the case of incomplete structures of *Olisthodiscus luteus* H1, H2A and H3, multiple-sequence alignment was performed twice. For the first alignment, the *Olisthodiscus luteus* partial sequences were entered into the data banks with random gap lengths inserted between peptide sequences. The GCG program, during the course of the first alignment, adjusts the gap lengths such that the peptides would be maximally aligned to the other sequences. The gaps represent unsequenced protein and not insertions or deletions in the case of peptide alignments. As a result of the gap penalty, clustering of the alignments is nonsensical.

To overcome the gap penalty, a second multiple-alignment procedure was undertaken, using the first complete alignment with modifications as the sequence data bank. The modification simply consisted of editing out from all the complete structures, the sequences that corresponded to the interpeptide gaps of the *Olisthodiscus luteus* histones. The data bank therefore consisted of fictional protein structures created by the tandem linking of short sequences that corresponded to the *Olisthodiscus luteus* peptides. Any gaps that were introduced within the peptide sequence were left unedited. The second multiple-sequence alignment results in a dendogram constructed from the clustering of known peptide sequences without a penalty for unknown sequences data.

As it is impractical to fully reference every entry represented in the sequence alignments, sequences are referenced by quoting the "Swissprot" data bank accession number for that sequence.

To facilitate phylogenetic discussion, a similarity matrix was constructed utilizing the GCG Distances program. The matrix consists of values resulting from a comparison of each sequence to the other, where the number of matches between two sequences (including gaps) is divided by the average length of the protein. A match is defined when the value in the symbol comparison table of Gribskov and Burgess (1986) exceeds a threshold value of 0,7. This value is proportional to the frequency of Arg ↔ Lys and Val ↔ Leu mutations. Only matrix values pertaining to *Olisthodiscus luteus* will be documented.

3.2 *Olisthodiscus luteus* Histone H1

The role of the histone H1 is not fully understood. However, it is known that H1 is loosely bound to the nucleosome (Thoma *et al.*, 1983), and that it is associated with spacer region DNA (the linker DNA between nucleosomes) (Noll, 1977). The binding of H1 results in the closing of two full turns of DNA around the nucleosome (Allan *et al.*, 1980). This results in the 180-220 base pair protection of DNA observed when chromatin is subjected to nuclease activity (Hewish and Burgoyne, 1973; Noll, 1974).

Two models have been proposed for the chromatin tertiary structure. Condensation of chromatin to form a solenoid, or 30nm fiber, has been proposed by Finch and Klug (1976). Zentgraf *et al.* (1980) proposed the structure to be in the form of "superbeads" - large granules, 36nm-47nm in diameter, containing 20-26 nucleosomes. Weintraub (1984) proposed that these superstructures are H1 dependent and that they may provide a means by which gene activation could be controlled.

The mechanism by which histone H1 performs these functions is unclear. Thoma *et al.* (1983) determined that the basic C- terminal of H1 binds to DNA more effectively than the N- terminal and as effectively as the whole H1 protein. The correct placing of H1 molecule with respect to the nucleosome appears to be controlled by the N- and C- terminals (Allan *et al.*, 1986). It was also found (Allan *et al.*, 1986) that the globular domain and the C- terminal of H1 was required to induce folding of the chromatin into a higher ordered structure. The N- terminal was found to be uninvolved in this function.

Böhm *et al.* (1988) established that the accessibility of the highly conserved glutamine residue (position 67 in alignment below) of calf thymus H1 to *Staphlococcus aureus* V8 protease is highly dependent on the preceding eight residues - the last residues of the N- terminal. The authors conclude that the N-terminal might dictate the tertiary folding of the globular domain. A change in this sequence would change its binding to the nucleosome and the ability to form higher order structures. This, in turn, could provide a mechanism for transcription control. However, this does not appear to be relevant in the case of the *Olisthodiscus luteus* Histone H1.

Figure 3.1 shows the alignment of *Olisthodiscus luteus* H1 protein sequences to several other H1 structures. The N- terminal domain includes residues 1 - 59 and the globular domains start at residue 60. The most striking observation of the alignment is that the H1s of *Olisthodiscus luteus* lack the N- terminal domain. It could be argued that this is due to proteolytic activity but, for reasons discussed in section 2.7, this appears unlikely.

The nucleosomal structure found in *Olisthodiscus luteus* appears typical of that found in higher eukaryotes. Although the *Olisthodiscus luteus* DNA repeat length after brief digestion with micrococcal nuclease is 220 bp (Shupe *et al.*, 1980) and is larger than the eukaryotic average, it is within the observed size range (Lewin, 1980). Shupe *et al.* (1980) found that the length of DNA associated with the nucleosomal core was 140 bp. This value is typical for nucleosomal structures.

No-higher order chromatin structures have been observed in the *Olisthodiscus luteus* nucleus. It is tempting to propose a relationship between the lack of evidence for 30 nm fibers or superbeads to the lack of the N- terminal domain in the *Olisthodiscus luteus* H1 histones. It is possible that the tertiary structure of the globular domain, which is sensitive to the N- terminal domain primary structure (Böhm *et al.*, 1988), is no longer able to form a higher order structure. However, without supportive evidence this remains speculation. It is possible that the N-terminal domain (or lack of it) may influence other aspects of chromatin, such as the phasing of nucleosomes along the DNA. Although little concrete can be said regarding the lack of this domain, it does provide a useful system for further study on the structure/function relationship of H1.

On inspection of the alignments shown in figure 3.1 it is apparent that the primary structures are highly conserved. Many of the mutations in this domain are conservative changes such as valine changing to leucine (or visa versa) and glutamic acid changes to or from aspartic acid. Position 81 (glycine) and position 83 (serine) are conserved in all H1s aligned (figure 3.1), except for those of *Tetrahymena*.

It is clear that the two *Olisthodiscus luteus* H1 molecules are most closely related to each other. By inspection, there appears to be little relation between

plants (maize, pea) or other unicellular organisms such as the *Tetrahymena*. From the GCG distances program (Table 3.1) the closest related H1 to *Olisthodiscus luteus* H1a (protein a1; section 2.7.2) is that of the painted sea urchin (*Lytechinus pictus*) (Knowles and Childs, 1986). Similarly, the painted sea urchin H1 is also the most related to *Olisthodiscus luteus* H1b (protein a2; section 2.7.2). The percentage difference between the sea urchin and *Olisthodiscus luteus* H1a (including gaps) is 43 %. The difference between *Olisthodiscus luteus* H1b and the sea urchin H1 is 42 %. Between the two *Olisthodiscus luteus* H1s, the percentage difference is 34 %. Assuming the validity of the molecular evolutionary clock (to be discussed in section 3.7), and that for a 1 % difference to occur between two H1 sequences the unit evolutionary period (UEP) takes 8×10^6 years (Wilson *et al.*, 1977), it would seem that the H1 gene of the common ancestor of *Olisthodiscus luteus* and painted sea urchin diverged 344×10^6 years (for *Olisthodiscus luteus* H1a) and 336×10^6 years for *Olisthodiscus luteus* H1b. Using the same principles, the two *Olisthodiscus luteus* genes diverged some 272 million years ago.

It is appreciated that the analysis here has been performed using a small domain of the full H1 sequence. Although a more complete calculation has to be done on the fully elucidated *Olisthodiscus luteus* H1 sequences, it is possible that using a smaller domain, particularly a conserved domain, might provide an indication of the true result, had the full sequence been used. Furthermore, it must be remembered that the UEP has been calculated from studies of the intact protein and not of a small domain as performed here. This could well compromise the dating of the divergence. It would appear that *Olisthodiscus luteus* H1 gene diverged from that of painted sea urchin $330 - 340 \times 10^6$ years ago. The two *Olisthodiscus luteus* genes diverged 270×10^6 years ago, possibly by a gene duplication mechanism. At this stage it is not possible to determine when the N- terminal domains of the algal H1s were lost. Assuming that this loss is not due to a post-transcription modification, but rather gene coded, two possibilities exist: Either the modification occurred soon after sea urchin and *O.luteus* diverged, prior to the algal H1 gene duplication (the most likely), or the trimmed gene was inherited from the sea urchin ancestor. The echinodermata line has either lost this gene, or does not express it, if the latter is the case.

Although no clear evolutionary relationships can be determined, it is surprising that sequence comparison of part of the conserved H1 domain indicates a

relationship between two such morphologically divergent organisms such as algae and the sea urchin.

FIGURE 3.1 Comparison of histone H1 sequences aligned for maximum homology by the GCG program. H1 sequences from the following sources were compared (common name and H1 type; proper name, accession number for "Swissprot" GCG data bank): Maize H1 *Zea mays* (P23444); Pea H1 *Pisum sativum* (P08283); Duck H1 *Anas platyrhnchos* (P09426); Chicken H1 *Gallus gallus* (P09987); Bovine H1.1 *Bos taurus* (P02253); Rat H1D *Rattus norvegicus* (P15865); Rabbit H1.3 *Oryctolagus cuniculus* (P02251); Trout H1 *Salmo gairdneri* (P06350); Frog H1C1 *Xenopus laevis* (P15866); Pig H1T *Sus scrofa* (P06348); Mouse H1.0 *Mus musculus* (P10922); *Caenorhabditis elegans* H1.1 (P10771); Purple Sea Urchin H1D *Strongylocentrotus purpuratus* (P15870); Sea Urchin H1 *Echinolampas crassa* (P02257); Angulate Sea Urchin H1 *Parechinus angulosus* (P02256); Purple Sea Urchin H1G *Strongylocentrotus purpuratus* (P07796); Painted Sea Urchin H1 *Lytechinus pictus* (P06144); *Olisthodiscus luteus* H1b; *Olishodiscus luteus* H1a; Dumeril's Clam Worm H1A *Platynereis dumerilii* (P06894); Midge H1 *Chironomus thummi thummi* (P21895); Fruit Fly H1 *Drosophila hydei* (P17268) and *Tetrahymena pyriformis* H1 (P12305). Only amino acids that differ to the concensus sequence is shown. Homology is indicated by (.) and a deletion by (-).

Key to superscripts:
1: *Caenorhabditis elegans*
2: Purple Sea Urchin H1D
3: Common Sea Urchin
4: Angulate Sea Urchin
5: Purple Sea Urchin H1G
6: Painted Sea Urchin

FIGURE OVERLEAF

```
                    1                                                  50
          Maize           atd. tet.a...-. a.p.apad.p a.p.ad.na. ....a-----
            Pea     mateepiva. etv.e.i.t. .tti.e.evp ekee.k..ve .t..a.gs.p
           Duck                          .....v.. ..vsa.ga.. ag..---p..
        Chicken                          .....v.. ..vsa.ga.. a...---p..
         Bovine                          .....-.. .........t pv-.......
            Rat                          .....-.. ...pa....t pi-....r.a
         Rabbit                           se..-.e t..pa....s p.........
          Trout                             .ev.p ap..aapa.. pk..------
           Frog                           a...stet tp......pk qk...qqp..
            Pig                           a......p adsv.asvek ppa..rgk.p
          Mouse                                          te nstsap...p
  Caen.elegans1                             sds.v v...vepkvp ...aa....p
    Sea Urchin2                              m.dtd. ap...apstp .---.....-
    Sea Urchin3                         a..pqkr..s .rks.kkspr .sp...spr.
    Sea Urchin4                         .g.pqkr..s .rks.rkspk .spr.as.sp
    Sea Urchin5                                                   ms...p
    Sea Urchin6                                                   ms...p
  O. luteus H1b
  O. luteus H1a
       Clam Worm                                                      .rrr
           Midge                              msd..ie v.pv.v.sp. ....e.kp.s
       Fruit Fly                  .sds vvavs.spvt .qt.saek.v a...pas.sa
    Tetrahymena                              gkqs tsksvtr..k dv..tv.p..
       CONSENSUS        .........V ...P.P.V.. PASETAPAAA PAAAPPAEKA KAKKKKAAKK

                    51                                                 100
          Maize     -t.p.krasp t.l.ya...s e..ts....t ...sy..a.. .edkh.aklp
            Pea     ...s.prnp. s..tye...k d..vs...kn ...qy..a.. .eekq.-qlp
           Duck     aaggska..p .g.s.t...t k....s.... .l..a.l..a l..g---...
        Chicken     aaggakp..p .g.s.t...t k..s.s.... .l..a.l..a l..g---...
         Bovine     p--ag.r... sg.......t k....s...s .v..a.l..a l..a---...
            Rat     a--gg..... sg.......t k....s...s .v..a.l..a l..a---...
         Rabbit     pg.ga..... .g.......t k....s...n .l..a.l..a l..g---...
          Trout     ---aa..p.k .g.s.g.... k..s.s...s .v..a.l..s l..g---...
           Frog     aaggakak.p sg.sa..... ks.s.s...g .v..a.l..a l..g---...
            Pig     vgltgts... psas..k..t e.lsvsq..a .m..a.l..a l..a---...
          Mouse     .r..as.kst d..ky..... ...q.e.n.a ...rqs.q.. .ksh....en
  Caen.elegans1     t.va...apv ....yint.k e..kq..... .a.kq..l.. .sq...l.dn
    Sea Urchin2     ...s.p.tp. s..ky....a s.les...k. ...rq..l.. .k...t..dn
    Sea Urchin3     r....ka-s- -----i---t -----q---r ---vak-qs- ---k-rcdin
    Sea Urchin4     rrka.ra.as t....l...q ...t..... ...aak..s. m....r.dmn
    Sea Urchin5     .v....rvap ....s.q... ...t.....g ...tq..... .....t.dmt
    Sea Urchin6     .t....raap ....t.q... ...t.....g ...nq..... .......din
  O. luteus H1b                tyy.l.k ...l.....n ...a..... .le.n.i--e
  O. luteus H1a                .tyyd.vk d..v...d.n ...m..... .e..q..--.
       Clam Worm  .t.-----a. .....at... t..lg..... ....v..... .....r.dva
           Midge  d.p..p.apr t......... n..kt....g ...vq..... lv.q...dt.
       Fruit Fly  s....ttapp t...tqq..d .s.qn....g ....l..... .s.t..cdaq
    Tetrahymena   aik.vt.kst tpvkt.kaap .sttpi..tt pvkad.k.-- ----------
       CONSENSUS  KKAKKAKRKA AHPPVSEMIV AAIAALKERK GSSL.AIKKY IAANYKVGYD
```

```
                  101                                                  150
Maize             pn-.rkllnv q..k..ag.k .tkv.----- ----n....s sat.pnp.pk
Pea               an-.kklllq n..kn.as.k ..kv.----- ---.....s aa..---.pa
Duck              ...n..... g.....g... .....---- .......... ..pg.t.e.a
Chicken           ...n..... g.....s... .....---- .......... ..pg.t.a.a
Bovine            ...n..... g.....s... .....---- .......
Rat               ...n..... g.....s... .....---- ......... ...asgea..
Rabbit            ...n..... g.....s... ..e..---- ........d ...asgea..
Trout             ...n.....i .v....t... .....---- .......... ...v.a.-..
Frog              ..rn...l.. ...a..t... .t.v..---- s......... ..ql.t.v.a
Pig               ...n..... g.....g..i ....r.---- .........s ...ap.pr.g
Mouse             a---..q... si.r..tt.v .k.....--- v......r.- a-kg..p.rs
Caen.elegans[1]   .iqi.ahhrq ...rg.tska ...ar.---- s..n.r.rv- pekaaaa...
Sea Urchin[2]     a---.vh..q ...rg.ts.q .rhv..---- s......l.- aektktp..-
Sea Urchin[3]     a--..ph.rr ...nq.ks.a .k.vs.---- v..t.r.rvg av--------
Sea Urchin[4]     .--.aph.rr ..rng.as.a .k.vt.---- .....r.rvg av..------
Sea Urchin[5]     k--qgpf.rr ..vkg.as.a .....---- k........- -gk.k.g.sd
Sea Urchin[6]     k--qatf..r ...ag.an.. ...v..---- k........- -gkvkag.te
O. luteus H1b     f--qqtflrx ...
O. luteus H1a     .--k.hylra ...
Clam Worm         r--.apf.rk firka.k--- --.....--- .......rv- -nk-------
Midge             k--.spf..k y...a.e..q .l.....--- k........p aa..k..vvk
Fruit Fly         k--.apf..k y..ns.an.k .....---- k........s ass.k.p.p-
Tetrahymena       -----ihrtk t..es.sdak ----.tvhks a.dkklsrsq .p..r.aa.k
CONSENSUS         VEKLNSRIKL ALKSLV.KGT LVQTKGSA.A TGASGSFKLN KKAKEEKKKP

                  151                                                  200
Maize             ..p..pktga ..pkaaa.pk a.tpa..kp. t...p...p. avvkp.tpak
Pea               v..p..ktaa .a.sv.a.p. a.--p..kav v...v.s.a. ava..pkka.
Duck              tk..p...p. .pa.k.pas. a..p...a-- -----.v..s ........a.
Chicken           tk..p...p. .pa.k.pa.. a..p...a-- -----.v..s ........a.
Bovine
Rat               k...ag.--- -a..k.pag. a..p...tgt at..kst..t ........ag
Rabbit            kp..ag.--- -a.pk.pag. t--p..pk.. aga.k.v..t ....p..ka.
Trout             .k.aa.p.a. .va.k.pa.. k.p..v.a.k av---....s ........tp
Frog              v...---.lv ap..k.pvt. k..p.spk.p ..vs...a.s .......v..
Pig               kv..p.---- ---------. a.t..lvlsr dskspksa.a n.r...srtt
Mouse             v.f..tk.ev ..v.tp.... kp..aa-s.. psk.pk.t-- .v.k-akk.p
Caen.elegans[1]   ..a..p.aa. .pa.a..atg e..a..paa. .pk...tgd. kv.ka.sp.k
Sea Urchin[2]     ..a...tp.. .pa.------ -..t..paak .at.-----. .a.kpaak.k
Sea Urchin[3]     -------.rs aas.n.l..t re.arara.. ..a....rr. aaa..rk.a.
Sea Urchin[4]     ------p..a ..tsaaa..k ka.aaa.k.. r.a.....r. aal...k.a.
Sea Urchin[5]     .q.ari.a.. a.l.a...eq re..alkt.. r.e.v....a a...t.kt.k
Sea Urchin[6]     .q.ar..a.. a.l.a...eq ke..aakt.. r.e.l....a a...---..k
O. luteus H1b
O. luteus H1a
Clam Worm         -----t.vp. ....a..pk. k-------.v ....s..... tnr.------
Midge             kvt..vte.. p...vs.pkt ge..v.kti. ....v.sat. i..pv--..t
Fruit Fly         --.vssve.. s..vtssa.. a...tisa-t ....gv.d.. ls..vvtk.s
Tetrahymena       ivhpakkaaa .p.ta..evk kdt.pvkkd. ..dtkpv..d a..dt....k
CONSENSUS         AAKKKAAKKK KKKA.KKKAA .KKKKKA.KA KKPKAAAKKK PKKAKKPAKA
```

```
                    201                                                  250
          Maize     p.a.p.a..- ...tag.... pl..---.ag ra.aaktsa. dtpg.k.p..
            Pea     ..p.tva-.- .t.pta.... avv.pks.v. pa.vaktsv. ttpg.kva.v
           Duck     .t.....sp. .aa.-agr.. ........a. .v........ a..p......
        Chicken     .t.....sp. .at.-agr.. .t......a. .v......s. a..p......
         Bovine
            Rat     ...---a.sp .ka.-at.a. ..p.....ar .v........ ts.p....p.
         Rabbit     ..-------p .va.-pks.a .v....k.a. .v........ ap.p......
          Trout     --.....sp. .v..pa.aa. ......k.a- ---t...... a..p......
           Frog     .-----.sp. ...--av.s. .vt.....-. .t......a. i..a....g.
            Pig     .aq...rsgr .t.--e..v. qqr.....ar .a....gn.. ltqq.tnpr.
          Mouse     .atp--.... ...vv.v..v ..s.pkka.t v.pkakssa. rgs..k
Caen.elegans[1]     va.p....va .spa...a.. .i..pa...a ..paak.
  Sea Urchin[2]     va.p.....a ..va...t.. ---.kvv..a ..gkgkk.
  Sea Urchin[3]     ..rr.....r .a.a.p
  Sea Urchin[4]     ..r...a... .a..p.k.aa .k..k....s p..a.kpak. sp...k..rs
  Sea Urchin[5]     v..p..... ..aa..pa-- -.k.pa...- ....akkva. ....a...p-
  Sea Urchin[6]     v..p..-... ..a--.ka-- -.k.pa...a ....a.k--. ....a.k.pa
O. luteus H1b
O. luteus H1a
     Clam Worm      -r.pkt..nr n
         Midge      t..p..a.pt .----.va-- -.kpka.p.p kaa..pkva. .-..a..pka
     Fruit Fly      vd..k.e... akda..vgti ..kptt..a. ssaa.pkt.. .ktts.kpk.
  Tetrahymena       dt...t.gs. .n
    CONSENSUS       AKKKAAKKAK KPKKKKAKPK KAAKSPAKKK AKKPKAAKPK PAKKKAAKAK

                    251                                                  296
          Maize     .a.ps...at pvr.apsr.. k.
            Pea     .kv.a..vpv ksv.a.sv.s pv..vsvk.g gr.
           Duck     .a.p...
        Chicken     .a.t...
         Bovine
            Rat     .t.....
         Rabbit     .t.....k
          Trout     .a.p...
           Frog     .a....
            Pig     atnr.
          Mouse
Caen.elegans[1]
  Sea Urchin[2]
  Sea Urchin[3]
  Sea Urchin[4]     pkk...a.g. rkpaa.karr spr.agkr.s pk.ar.
  Sea Urchin[5]     akk.a.p... ...--p.... a.pak.
  Sea Urchin[6]     akk.a.p... .....a.... a..
O. luteus H1b
O. luteus H1a
     Clam Worm
         Midge      .kp.ae.kp. .....s....
     Fruit Fly      vv..aspk.a ........t. sat.
  Tetrahymena
    CONSENSUS       K.AAKKKAKK AAKKPKAKKA .KKK.K..R. ..K..K.... ......
```

	O.luteus H1b	H1a
Maize	0.3934	0.4098
Pea	0.4590	0.4918
Duck	0.3607	0.4426
Chicken	0.3607	0.4426
Bovine	0.3607	0.4426
Rat	0.3607	0.4426
Rabbit	0.3770	0.4590
Trout	0.3607	0.4426
Frog	0.3443	0.4262
Pig	0.3115	0.3934
Mouse	0.4098	0.4262
Caenorhabditis elegans	0.4098	0.4918
Purple Sea Urchin	0.4262	0.4918
Common Sea Urchin	0.3934	0.4262
Angulate Sea Urchin	0.4426	0.4426
Purple Sea Urchin	0.4918	0.4918
Painted Sea Urchin	0.5574	0.5410
O.luteus H1b	1.0000	0.5902
O.luteus H1a	0.5902	1.0000
Clam Worm	0.4590	0.5246
Midge	0.4590	0.4918
Fruit Fly	0.4098	0.4426
Tetrahymena	0.1475	0.1148

TABLE 3.1 Similarity between *O.luteus* H1a and H1b peptides compared to sequences aligned in Figure 3.1 and as discussed in section 3.1.2. The number of matches have been divided by the sequence length. 100 % homology is therefore equal to 1. Only domains which aligned to the *O.luteus* peptides were analysed. Proper names and "Swissprot" accession numbers are as given in the legend of Figure 3.1.

3.3 *Olishtodiscus luteus* Histone H2A

It has been twenty years since the first histone H2A (calf thymus) was sequenced and the full primary structure determined (Yeoman *et al.*, 1972). Since then, many H2A structures have been solved and variants identified from a wide range of sources e.g. - rat (Laine *et al.*, 1976), wheat (Rodrigues *et al.*, 1979), *Strongylocentratus purpuratus* (Sures *et al.*, 1978) and *Volvox carteri* (Muller & Schmitt, 1988). The latter two examples were solved via DNA sequencing. Despite the large number of variants, the overall structure remains highly conserved - a variable N- and C- terminal (residues 1 - 21 and 128 - 154 respectively) and a highly conserved central globular domain. Residue numbering refers to the alignment of H2A structure in figure 3.2 and the domains defined by Böhm *et al.*, (1980). Charge clusters, as observed by Yeoman *et al.* (1972) studying calf H2A, appear to be conserved in all H2As studied. A single acidic amino acid cluster is found between residues 97 - 101. Basic amino acids are found between residues 38 -49; 81 - 87; 91 - 92; 128 - 129 and from 132 to the end of the molecule. A large hydrophobic domain may be found between residues 54 and 79.

D'Anna and Isenberg (1974) showed the H2A and H2B of calf thymus aggregate in a 1:1 ratio. The same was demonstrated for pea H2A and H2B (Spiker & Isenberg, 1977). Moss *et al.* (1976) and DeLange *et al.* (1979) demonstrated that this interaction was dependent on the pH sensitive secondary and tertiary folding of both histones. By UV radiation crosslinking and proton magnetic resonance, it was shown that residue 39 - 105 were important in this interaction. In particular, tyrosine at positions 49, 59 and 67, histidine at positions 40 and 92, phenelalanine at position 34 and proline at 35 were shown to be involved in the H2A - H2B dimer formation.

The alignment of the sequenced *Olisthodiscus luteus* H2A peptides against known H2A sequences, performed by the GCG program, is shown in figure 3.2. The peptides of *Olisthodiscus luteus* H2A align to the conserved central domain of other H2As. It is interesting to note that the mutations that have occurred in the *Olisthodiscus luteus* H2A peptides as compared with the consensus sequence, are at sites that show variation throughout evolution. Threonine (69) is homologous in mammals, aves, sea urchins and volvox. In plants, insects and some animal variants, this residue is alanine. The serine at position 86 is homologous to sea urchins and tomato. In pea and volvox, this residue is asparagine. In all other known

sequences the amino acid was found to be threonine. *Olisthodiscus luteus* shows a conservative mutation at residue 93. Except for tomato H2A and *Olisthodiscus luteus* H2A which have valine, the majority of eukaryotes have leucine at this position. Although all the studied eukaryotes contain glutamine at position 94, the plants and *Tetrahymena* contain leucine at this position. It may be significant in the study of *Olisthodiscus luteus* evolution to note that the alga is homologous to all animals at this position. The residues 107 - 109 are variable throughout the aligned sequences. *Olisthodiscus luteus* contains asparagine at position 109. This is fairly conservative change from glycine and is found in both plants and some animals. Some of 'higher' animals have a non-conservative mutation, resulting in a basic amino acid at this point.
Of the amino acids that were identified to be involved in the H2A - H2B interaction, those that occur in peptides homologous to the *Olisthodiscus luteus* peptides, were found to be conserved viz. Tyr 59, Tyr 67 and His 92.

By inspection, the *Olisthodiscus luteus* peptide sequences appear to be more homologous to higher animals than to the yeasts, fungi or plants. This is particularly evident when comparing the peptide aligned between residues 102 and 121. The GCG distance program, comparing fragments homologous to the *Olisthodiscus luteus* peptides, finds the most homologous sequences to be that of H2A and H2A.2 of sand sea urchin (*Psammechinus miliaris*), H2A of common sea urchin (*Echinolampas crassa*) and the H2A of purple sea urchin (*Strongylocentrotus purpuratus*). The sea urchin H2A sequences are 100 % homologous for the domains investigated (Table 3.2).

```
                    1                                                  50
Tomato              ...-...t.g ag.rk-.gp. k.svtk.ik. .........g ...k-..r..
Pea                 .........g ag.rk..gp. k.svt..v.. .........g ...k-..r..
Wheat               ..g.-.a..v a-a.kf.gp. k.svtk.ik. .........g ...k-..r..
Yeast[1]                    sg .ksg.kaava ks.q...ak. ..a..... ....-.....
Aspergillus                 tg .ksg.kasgs kn.q....k. ..a..... ....-.....
Yeast[2]                    sg .k.gka.s.a kasq...ak. ..t..... ....-r....
Sea Urchin[3]                . ....s-.... t......... .......... ....-.....
Sea Urchin[4]                . ....s-.... t...t..... .......... ....-.....
Sea Urchin[5]                . .....-kakg t.s.t..... .......... ...k-....g
O. luteus
Duck                         . ....q..... .......... .......... ....-.....
Chicken                      . ....q..... .......... .......... ....-.....
Human                        . ....q..... ....t..... .......... ....-.....
Rat                          . ....q..... .......... .......... ....-.....
Mouse                        . ....q.c... ....t..... .......... ....-....s
Frog                         . ....q...t. ....t..... .......... ....-.....
Trout                        . ....t..... ....t..... .......... ....-.....
Midge                        . .....-..vk .......... .......... ....-....g
Fruit Fly                    . .....-..vk g......n.. .......... ....-.....
Clam Worm                    . .....-...k g.s.t..... .......... ....-.....
Marine Worm                  . .....-...k g.s....... .......... ....-.....
Cuttlefish                   . .....-..vk g.s.t..... .......... ....-.....
Starfish                     . .....-.... .......... .......... ....-.....
Volvox                    ma .....ktagk -..v....k. .........a ...k-..k..
Tetrahymena            st.g.g .ka..kt-.s s.qv...a.. .........s ...k-h.r.s
CONSENSUS              MDASTKTKKS GRGKGGGKAR AKAKSRSSRA GLQFPVGRVH RLLR.KGNYA

                    51                                                 100
Tomato              q...s..... .......... .......... .....s.... ..vl......
Pea                 q...t..... .......... .......... .....n..s. ...l......
Wheat               q...s..... .......... .......... k......... ...l......
Yeast[1]            q......... .......... .......... .......... ..........
Aspergillus         q......... .......... .......... .......... ..........
Yeast[2]            q...s..... .t........ .......... .......... ..........
Sea Urchin[3]       k...g..... ........t. .......... .....s.... ..........
Sea Urchin[4]       k...g..... ........t. .......... .....s.... ..........
Sea Urchin[5]       s......... ........t. .......... .....s.... ..........
O. luteus           ----...... ...xxx..t. ...x...... .....s.... ..v....xxx
Duck                .......... ........t. .......... .......... ..........
Chicken             .......... ........t. .......... .......... ..........
Human               .......... ........t. .......... .......... ..........
Rat                 .......... ........t. .......... .......... ..........
Mouse               .......... ........t. .......... .......... ..........
Frog                .......... ........t. .......... .......... ..........
Trout               .......... ........t. .......... .......... ..........
Midge               .......... .......... .......... .......... ..........
Fruit Fly           .......... .......... .......... .......... ..........
Clam Worm           .......... .......... .......... .......... ..........
Marine Worm         .......... .......... .......... .......... ..........
Cuttlefish          q......... .......... .......... .....s.... ..........
Starfish            .......... .......... .......... ........n. ..........
Volvox              .......... ........t. .......... .....n.... ..i.......
Tetrahymena         ....t..... .......... .......... k......... ..il......
CONSENSUS           ERVGAGAPVY LAAVLEYLAA EILELAGNAA RDNKKTRIIP RHLQLAIRND
```

```
                101                                                   150
      Tomato    ...g...a.. ...s..... .np......s av.eek---s ..--a..g.. ....
         Pea    ...g...a.. ...y..... .np.....rk .n..as---t ....s..k.. ....
       Wheat    q..gr..s.. ...h...i.. .np......a aek.ekagaa ....k.tt.. ....
      Yeast1    ........h. .......v.. .n.h....ts ggtg.psqe.
Aspergillus     ........h. .......... .hqn...... p..g.gsqe.
      Yeast2    ........n. .......... .hqn.....s a..t..sqe.
Sea Urchin3     .......... .......... .......... g.ss
Sea Urchin4     .......... .......... .......... a.ss
Sea Urchin5     .......... .......... .......... a..s.
  O. luteus     xx......n. .......... .
        Duck    ........k. .......... .......... .s-h...s.
     Chicken    ........k. .......... .......... .s-h...a.
       Human    ........k. .......... .......... .shh...g.
         Rat    ........k. .......... .......... .shh...g.
       Mouse    ........r. .......... .......... .shh...g.
        Frog    .......... .......... ..s....... .s.ks..s.
       Trout    .......... .......... .......... ...v...
       Midge    .......s.. .......... .......... .skka
   Fruit Fly    .......s.. .......... .......... ..k.
   Clam Worm    .......s.. .......... .......... t..k
Marine Worm     .......s.. .......... .......... q.sk
  Cuttlefish    .......s.. .......... .......... q....
    Starfish    .......s.. .......... .......... a....
      Volvox    ...g....d. ...s..... .h.......s kggkgeeaa
Tetrahymena     .......ant ...d..... .npm...s.s k.tesrgqas qdi
   CONSENSUS    EELNKLLGGV TIAQGGVLPN IQAVLLPKKT EKAAKAK.KL PKSP.KA.KS PKKA
```

Key to superscripts:
1: Fission Yeast
2: Baker's Yeast
3: Common Sea Urchin
4: Purple Sea Urchin
5: Sand Sea Urchin

FIGURE 3.2 Comparison of histone H2A sequences aligned for maximum homology by the GCG program. H2A sequences from the following sources were compared (common name and H2A type, proper name, accession number for "Swissprot" GCG data bank): Tomato H2A *Lypcopersicon Esculentum* (P25469); Garden Pea H2A *Pisum sativum* (P25470); Wheat H2A.2 *Triticum Aestivum* (P02276); Fission Yeast H2A.1 *Schizosaccharomyces pombe* (P04909); *Aspegillus nidulans* H2A (P08844); Baker's Yeast H2A.1 *Saccharomyces cerevisiae* (P04911); Common Sea Urchin H2A *Paracentrotus lividus* (P13630); Purple Sea Urchin H2A *Strongylocentotus purpuratus* (P02271); Sand Sea Urchin H2A.1 *Psammechinus miliaris* (P04735); *Olisthodiscus luteus* H2A; Muscovy Duck H2A *Cairina moschata* (P13912); Chicken H2A *Gallus gallus* (P02263); Human H2A1 *Homo sapien* (P02261); Rat H2A *Rattus norvegicus* (P02262); Mouse H2A.1 *Mus musculus* (P22752); African Clawed Frog H2A1 *Xenopus laevis* (P06897); Rainbow Trout H2A *Salmo gairdneri* (P02264); Midge H2A *Chironomus thummi thummi* (P21896); Fruit Fly H2A *Drosophila melanogaster* (P02267); Dumeril's Clam Worm H2A *Platynereis dumerilii* (P19178); Marine Worm H2A *Sipunculus nudus* (P02270); Common Cuttlefish H2A *Sepia officinalis* (P02268); Common European Starfish H2A *Asterias rubens* (P02269); *Volvox carteri* H2A4 (P16866) and *Tetrahymena pyriformis* H2A.1 (P02273). Only amino acids that differ to the concensus sequence are shown. Homology is indicated by (.) and a deletion by (-).

O.luteus	H2A
Tomato	0.8793
Pea	0.8448
Wheat	0.8621
Fission Yeast	0.9310
Aspergillus	0.9310
Baker's Yeast	0.9138
Common Sea Urchin	0.9655
Purple Sea Urchin	0.9655
Sand Sea Urchin	0.9655
Duck	0.9483
Chicken	0.9483
Human	0.9483
Rat	0.9483
Mouse	0.9483
Frog	0.9483
Trout	0.9483
Midge	0.9138
Fruit Fly	0.9138
Clam Worm	0.9138
Marine Worm	0.9138
Cuttlefish	0.9483
Starfish	0.9138
Volvox	0.9483
Tetrahymena	0.9138

TABLE 3.2 Similarity between *O.luteus* H2A peptides compared to sequences aligned in Figure 3.2 and as discussed in section 3.1.2. 100 % homology has a value of 1. Only domains which aligned to the *O.luteus* peptides were analysed. Proper names and "Swissprot" accession numbers are as given in the legend of Figure 3.2.

3.4 *Olisthodiscus luteus* histone H2B

The histone H2B primary structures are the most evolutionary diverse of the core histones (Von Holt *et al.*, 1979). From H2B sequence alignments generated by the GCG program and shown in figure 3.3., it is evident that most of this diversity occurs in the basic N- terminal. Heterogeneity is due not only to amino acid mutations or deletions, but also to the length of the terminal extensions. The largely hydrophobic C- terminal appears well conserved.

Bradbury and Rattle (1972), on the basis of nuclear magnetic resonance and optical spectroscopic data, suggested that residues 66 - 136 (figure 3.3.) formed the structural part of the molecule. Moss *et al.* (1976) later proposed that the residues 71 - 148 were involved in forming a stable tertiary structure. D'Anna and Isenberg (1974) have shown that H2B forms a dimer with H2A in a 1:1 ratio. The two methionines, residues 93 and 96, were found to be essential for this interaction (Spiker and Isenberg, 1977). The middle third of the H2B molecule, residues 71 - 76, has been found to interact with H2A (Martinson *et al.*, 1979; DeLange *et al.*, 1979). One or more of the tyrosines in this region possibly interacting with a phenylalanine of H2A. The C- terminal third of the conserved domain is believed to interact with H4 in the nuleosomal structure (Martinson *et al.*, 1979). The observation that the molecule is divided into a basic N- terminal and a hydrophobic C- terminal, the latter which interacts with other histone molecules, led to the assumption that the molecule is made up of two distinct domains of very different structures and of different evolutionary stability, performing two different functions. The origin of the histone could possibly have been due to the fusion of two separate genes, fulfilling two different functions. (Elgin *et al.*, 1979; Strickland *et al.*, 1980).

The N- terminal is characterized by being variable in sequence and in length, rich in lysine, serine and proline, and separated from the C- terminal by 7 or 8 basic residues (residues 58 - 68 in figure 3.3.) (Von Helden *et al.*, 1979; Strickland *et al.*, 1980). Von Holt *et al.*, (1979) observed that the pentapeptide Pro- (1-3) basic amino acid, is also found in other DNA binding proteins such as the HMGs. This sequence may provide a structure for reversible association between DNA and nuclear proteins. Lindsey and Thompson (1992) determined through micrococcal nuclease protection studies that the extended N- terminal of wheat and sea urchin sperm H2B interacts with the linker DNA. In part, this may confirm the proposal

of Brandt *et al.* (1988) who suggest that the N- terminal extension may stabilize higher chromatin structures and contribute to the protection against nuclease activity during dormancy or arrested growth, such as in the wheat embryo or the sea urchin sperm cells. Both these cell types exhibit long N- terminal extensions. In contrast, sea urchin blastula cells, which have a high level of chromatin activity, exhibit a short N- terminal. This, too, may explain the particularly short N- terminal of the *Olisthodiscus luteus* H2B aligned in figure 3.3.

Olisthodiscus luteus, maintained under a 12 hour light : 12 hour dark regime, replicates approximately once every 24 hours. This would indicate a highly active chromatin structure. The fact that the H2B has a shortened N- terminal possibly results in the inability of *O.luteus* chromatin to form a stable higher chromatin structure. Furthermore, we have found that ± 50 % of *Olisthodiscus luteus* is acetylated at the second residue, lysine. Acetylation at this point reduces the net positive charge of the N- terminal and therefore its DNA binding potential. This type of modification is believed to promote transcription of the "relaxed" chromatin (Allfrey *et al.*, 1964).

The *Olisthodiscus luteus* H2B contains one Pro-Lys-Lys motif (Von Holt *et al.*, 1979) proposed to be a DNA binding motif, between residues 52 - 53. The basic amino acid domain (residues 58 - 68) that separates the N- terminal from the conservative C- terminal domain is much shorter and less basic than that found in other organisms. This might be due to the organism's need for a "relaxed" and easily accessible chromatin structure.

The domain proposed to bind to H2A was found to be well conserved (residues 71 - 76). All three tyrosines are present throughout the evolutionary alignment. Residue 72 appears to be evolutionary unstable. *Olisthodiscus luteus*, yeasts and some higher vertebrates have serine at this position, while the consensus residue is glycine. The following amino acid, residue 73, is isoleucine in the majority of H2Bs aligned. However, in the case of *Olisthodiscus luteus*, yeast and one sea urchin, this residue has undergone a non-conservative change to serine. Of the mutations found in comparing *Olisthodiscus luteus* H2B C- terminal with other sequences, most were found to be conservative changes. The non-conservative mutations in this domain of *O.luteus* H2B is in no case unique to the algae.

Muller and Schmitt (1988) proposed that certain residues are unique to plant. They are lysine at residue 72; lysine (106); arginine (116); proline (21) and valine (135). Of these, the *Olisthodiscus luteus* H2B is homologous only to the arginine at residue 116. *Olisthodiscus luteus* molecule does have a conservative but apparently unique mutation at position 135 (figure 3.3). In plants, valine is found and in animals, isoleucine or leucine, while in *Olisthodiscus luteus* methionine is found at that position. The GCG distance program finds that the most related sequence to that of *Olisthodiscus luteus* H2B is that of the purple sea urchin (*Strongylocentrotus purpuratus*).

As the N- terminals of the H2Bs are highly variable, and as these variants appear to be specific to the H2B function in the local dynamic chromatin structure, they probably do not reflect the organism's evolutionary history. The distance program was rerun after the N- teminal (residue 1 - 68 in figure 3.3.) had been edited out of each sequence (Table 3.3.). Once again, the purple sea urchin was found to be closely related.

```
                    1                                                  50
Tetrahymena                                                   a pk.apaaa.e
Volvox              ....-..... ..qaaea.ev kae..pkav. apkkke.k.p .k.vakepsa
Yeast1                                                sakae..p.s k.paek.p.a
Yeast2                                                 .aae..p.s k.pa.k--.p
Sea Urchin3                 pr.... s..r.g..r. g..-r.g.p. rgg..a.r.g
Sea Urchin4               psqk..t. r..t.r..t. r..qkg-g-- kgg..a.rgg
Sea Urchin5                   ..q r..t.r..t. r..qkgag-- kgg....rgg
Sea Urchin6                                        aptgq v.........
Mouse                                             ...a...p .p.......
Human                                             ...a...p .p.......
Duck                                              ...a...p .p.......
Chicken                                           ...a...p .p.......
Frog                                              ...a...p .p.......
Trout                                             ...a..-- .p.......
Midge                                            ap..t.g. ....ag-..q
Fruit Fly                                         p..t.g. ....ag-..q
Starfish7                                         p..p.g. gq..ag-..-
Starfish8                                          .p..g. gq..ag-..-
Limpet                                            p..v.s. g...ag-..-
Caen.elegans                                      ap..p... g...ak-.t-
O. luteus                                                 ..tp....a
Wheat                 . bkkpaaenkv eka.ekt.ag kk.kae.rlp .g.ta..e.g
CONSENSUS           MAPK.KEEKP AS....SPSK .SPAK.SP.K .SPEPKKSAK AAKKGSKKAV

                    51                                                 100
Tetrahymena         ..v..aptte k.nkk..s.t .a..... .....v...k ...n.....i
Volvox              ggedgdkks. k.akvaks.t .kl..... ......... .........
Yeast1              ...stst... -..ska...t .ss..... .t.......q .s.......
Yeast2              rd.m.sa.-. -..gkn...t .ss...r... .........n q..p.....
Sea Urchin3         .ggrrrnv-- v....r.r.. .gs..... ......... rg........
Sea Urchin4         .ag.rrrgvq v....r.r.. .g..... ......... r.........
Sea Urchin5         .a-rrr.gaa vr...r.r.. .g..... ......... rg........
Sea Urchin6         .ppras-g.. -..h..... .g..... ......... r..t......
Mouse               t.aq..-... -..k.s.... .s..... ......... ...g......
Human               t.aq..-... -..k.s.... .s..... ......... ...g......
Duck                t..q..-gd. -..kks.... .s..... ......... ...g......
Chicken             t..q..-gd. -..kks.... .s..... ......... ...g......
Frog                t..q..-... -...ks.... .a..... ......... ..........
Trout               t..ag.-g.. -..k.s.... .a..... ......... ...g......
Midge               .ais.d-.k. -...h..... .a..... ......... ..........
Fruit Fly           .nit.t-.k. -.k.-..... .a..... ......... ..........
Starfish7           .gaprt-.k. -r..-..... .g..... ......... r.........
Starfish8           .gaprs-.k. -r..-..... .g..... ......... r.........
Limpet              .aarsg-.k. -rk.-r.... .s..... ......... ..........
Caen.elegans        -v..p.-... -..kka.... .s....r... ......... ..........
O. luteus           .ap..a.s-- --k.n..v.t .ss..... .........k rg.......i
Wheat               geg.tr.rk. gskak.gv.t .k..... .....i.... .........i
CONSENSUS           KKTKKKGDGK .KRRRKRKES Y.IYIYKVLK QVHPDTGISS KAMSIMNSFV
```

```
                    101                                                 150
Tetrahymena         ..s.....l. s.k.vr.... r.l..... ..k..... .r.........
Volvox              .....k..t. ..k.sr...k p..t..... ....v..... ..........
Yeast[1]            ........t. ..k..a...k ....a..... ....i..... .........r
Yeast[2]            ........t. ..k..a...k .......... ....i..... .....t....
Sea Urchin[3]       .......... ....cqa.r. r......... .......... ..........
Sea Urchin[4]       ........a. .g..tt..r. .......... .......... ..........
Sea Urchin[5]       ........a. ....tk..r. .......... .......... ..........
Sea Urchin[6]       .......... ....tq...k .......... .......... ..........
Mouse               ........s. .......... ...t...... .......... ..........
Human               .......... .......... ...t...... .......... ..........
Duck                .......... .......... ...t...... ...s...... ..........
Chicken             .......... .......... ...t...... .......... ..........
Frog                .......... .......... ...t...... .......... ..........
Trout               .......... s......... ...t...... .......... ..........
Midge               ........a. .......... ...t...... .......... ..........
Fruit Fly           ........a. .......... ...t...... .......... ..........
Starfish[7]         ........a. .........k ...t...... .......... ..........
Starfish[8]         ........a. .........k ...t...... .......... ..........
Limpet              ........a. .......... ...t...... .......... ..........
Caen.elegans        ........a. .......... .......... ....i..... ..........
O. luteus           ......l... .....r.... ..l....... ....m..... ..........
Wheat               .....kl... sak..r...k p..t...... s...v..... ..........
CONSENSUS           NDIFERIAGE ASRLAHYNKR STISSREIQT AVRLLLPGEL AKHAVSEGTK

                    151      161
Tetrahymena         ....fs..tn
Volvox              .......a
Yeast[1]            .....s..t. .
Yeast[2]            s....s..a.
Sea Urchin[3]       ......t.r
Sea Urchin[4]       ......t.r
Sea Urchin[5]       ......t.r
Sea Urchin[6]       ......ta.
Mouse               .........
Human               .........
Duck                .........
Chicken             .........
Frog                .......a.
Trout               .........
Midge               .........
Fruit Fly           .........
Starfish[7]         ......t..
Starfish[8]         ......t..
Limpet              .........
Cael.elegans        .........
O. luteus           ....f..n
Wheat               ....f...
CONSENSUS           AVTKYTSSKQ A
```

Legend overleaf

Key to superscripts:

1: Baker's Yeast
2: Fission Yeast
3: Painted Sea Urchin
4: Angulate Sea Urchin
5: Purple Sea Urchin
6: Sand Sea Urchin
7: Common European Starfish
8: Spiny Starfish

FIGURE 3.3 Comparison of histone H2B sequences aligned for maximum homology by the GCG program. H2B sequences from the following sources were compared (common name and H2B type, proper name, accession number for "Swissprot" GCG data bank):*Tetrahymena thermophila* H2B.1 (P08993); *Volvox carteri* H2B3 (P16867); Baker's Yeast H2B.1 *Saccharomyces cerevisiae* (P02293); Fission Yeast H2B.1 *Schizosaccharomyces pombe* (P04913); Painted Sea Urchin H2B.2 *Lytechinus pictus* (P06146); Angulate Sea Urchin H2B.1 *Parachinus angulosus* (P02290); Purple Sea Urchin H2B.1 *Strongylocentrotus purpuratus* (P06145); Sand Sea Urchin H2B.1 *Psammechinus miliaris* (P02287); Mouse H2B.1 *Mus musculus* (P10853); Human H2B (100% 1homologous to bovine) *Homo sapien* (P02278); Muscovy Duck H2B *Cairina moschata* (P14001); Chicken H2B *Gallus gallus* (P02279); African Clawed Frog H2B.1 *Xenopus laevis* (P02281); Brown Trout H2B *Salmo trutta* (P02282); Midge H2B *Chironomus thummi thummi* (P21897); Fruit Fly H2B *Drosophila melanogaster* (P02283); Common European Starfish H2B *Asterias rubens* (P02286); Spiny Starfish H2B *Marthasterias glacialis* (P02285); Sandpaper Limpet H2B *Patella granatina* (P02284); *Caenorhabditis elegans* H2B (P04255); *Olisthodiscus luteus* H2B and Wheat H2B.2 *Triticum aestivum* (P05621).

O.luteus	H2B
Baker's Yeast	0.5404
Fission Yeast	0.5404
Painted Sea Urchin	0.5776
Angulate Sea Urchin	0.5776
Purple Sea Urchin	0.6025
Sand Sea Urchin	0.5715
Mouse	0.5590
Human	0.5652
Duck	0.5590
Chicken	0.5652
Frog	0.5652
Trout	0.5466
Midge	0.5776
Fruit Fly	0.5714
European Starfish	0.5652
Spiny Starfish	0.5652
Limpet	0.5901
Caenorhabditis elegans	0.5590
Wheat	0.5590

TABLE 3.3 Similarity between the complete *O.luteus* H2B protein compared to sequences aligned in Figure 3.3. and as discussed in section 3.1.2. 100 % homology is equal to 1. Proper names and "Swissprot" accession numbers are as given in the legend of Figure 3.3.

3.5 *Olisthodiscus luteus* H3

The histones H3 and H4 are the most conserved of all histone proteins. D'Anna and Isenberg (1974) have shown that these two proteins form a stable tetramer complex consisting of two H3 and two H4 proteins. This strong association is most likely due to hydrophobic bonding (Eickbush and Mohdrianakis, 1978). The H3 - H4 tetramer plays a central role in the self-assembly of the nucleosome. It binds to the DNA prior to the two H2A - H2B dimers and forms an intermediate nucleosome-like structure (Camerini-Otero and Felsenfeld, 1977; Daban and Cantor, 1982)

The alignment of several H3 sequences is shown in figure 3.4. The conserved nature of the proteins can be clearly seen. Of all the mutations, the majority occur in the C- terminal half of the histone. With the exception of *Volvox*, which appears to have a serine deletion at residue 28, H3s are 135 amino acids long. The first 53 amino acids, the N- terminal, is rich in basic amino acids, while the C- terminal appears to be slightly richer in acidic residues - especially the domain between residues 90 - 110.

The region between residues 94 and 106 is rich in amino acids with high potential for a helix formation. Provided the helix is naturally occurring, it has been proposed (Palau and Padros, 1972) that one side of the helix would exclusively consist of, hydrophobic residues and therefore be capable of hydrophobic interaction.
The histone H3 is not entirely a globular structure. The N- terminal (residues 1 - 53) appears to be unstructured (Moss *et al.*, 1976b). Böhm *et al.* (1977) determined that the C- terminal (residues 42 - 120) plays an essential role in the tetramer complex formation. The last 15 residues of the C- terminal were not found to affect nucleosomal assembly or reconstruction.

With the exception of yeast, the cysteine at position 110 is highly conserved. It has been shown that any modification of this cysteine, such as by carboxymethylation, reduces the interaction between H3 and H4 (Lewis, 1976; Lewis and Chiu, 1980). Camerini-Otero and Felsenfeld (1977) showed that a H3 dimer formed by a covalent Cys 110 - Cys 110 linkage, had little effect on the ability of the H3 protein to reconstitute into a nuclosome. This indicates that these two residues are localised close to each other in the native tetramer. It is clear that the C- terminal structure is conducive to forming a complex such as the

H3 - H4 tetramer. This domain has also been implicated as being responsible for the first degree of DNA compaction, viz. the folding of the DNA around the nucleosome (Econtre and Parello, 1988; Ausio *et al.*, 1989).

Despite the high content of arginine and lysine residues in the N- terminal, the stability of the nucleosome is unaffected if the terminal is removed by trypsin digestion. The ability of the trypsinized H3 to form a reconstituted nucleosome is also unaffected. It is possible that the N- terminal of H3 plays a role in the regulation of the accessibility of the nucleosomal DNA (Encontre and Parello, 1988). Figure 3.4, indicates that the first 20 residues of the N- terminal are the most conserved domain of the protein. However, such a high degree of conservatism does not implicate this domain as a regulatory structure. It is the variant terminal domains of H2A and H2B that have led to the proposal that these regions may perform some regulatory functions. It is possible that the H3 N-terminal may play a role in establishing higher order chromatin structures. e.g. the 30 nm fiber.

The partially solved primary structure of *Olisthodiscus luteus* H3 is shown in figure 3.4. Of the peptides sequenced, there is 100 % homology to corresponding peptides from human, bovine, mouse and sea star H3. The first 20 amino acids of the *Olisthodiscus luteus* H3 N- terminal are conserved to all other organism, as is the cystiene at residue 110. As is observed in other histone H3s, lysine at position 9 was found to be 83 % methylated (DeLange *et al.*, 1973; Hayashi *et al.*, 1982).

A consensus sequence for plant H3 proteins has been devised by Wu *et al.* (1989). This consensus sequence shows four positions that appear to be different in all known plant and all known animal structures. In animal H3 sequences, residues 41, 53 and 96 have been identified as tyrosine, arginine and serine respectively. In plants, residues 41, 53 and 96 appear to be phenylalanine, lysine and alanine respectively. Residue 90 in plants is more variable than the same position in the animal histone H3s. In all animal structures, this residue has been found to be methionine, except for H3.3 variants that have a glycine. In plants, four different amino acids of different characteristics may be found, i.e. alanine and leucine (both hydrophobic) serine (polar) or a basic arginine. Of these residues, only residue 53 is found in a sequenced domain of *Olisthodiscus luteus* H3. At this point, *Olisthodiscus luteus* has an arginine, homologous to animals.

All the peptides sequenced of the *Olisthodiscus luteus* H3 are homologous to highly conserved domains of the molecule. An evolutionary comparison by the GCG program would be of little value. However, by inspection, it can be observed that the *Olisthodiscus luteus* H3 does not appear to be as homologous to algae (*Volvox*), plants or fungi as it is to higher eukaryotic invertebrates and sea urchins.

```
                    1                                                  50
Tetrahymena         ..........  ..v.......  ..........  vs.......k  ..........
Sea Urchin1         ..........  ..........  ..........  ..........  ..........
O. luteus           ..........  ..........  ..........  .
Human               ..........  ..........  ..........  ..........  ..........
Fruit Fly           ..........  ..........  ..........  ..........  ..........
Sea Urchin2         ..........  ..........  ..........  ..........  ..........
Coral               ..........  ..........  .....a....  ..........  ..........
Bread Tree          ..........  ..........  ..........  ..........  ..........
Maize               ..........  ..........  ..........  ..........  ..........
Pea                 ..........  ..........  ..........  ..........  ..........
Alfalfa             ..........  ..........  ..........  t.........  ..........
Caen.elegans        ..........  ..........  ..........  .s........  ..........
Frog                ..........  ..........  v....k.c..  ..........  ..........
Cress               ..........  ..........  ..........  .........t  .......t.
Volvox              ..........  ..........  .......-t.  ..........  ..........
Yeast3              ..........  ..........  .s.....a..  ..........  ..........
Aspergillus         ..........  ..........  .s.....a..  s.........  .k........
Neurospera          ..........  ..........  .s........  s.........  .k........
Yeast4              ..........  ..........  .s........  s.........  .k........
CONSENSUS           ARTKQTARKS  TGGKAPRKQL  ATKAARKSAP  ATGGVKKPHR  YRPGTVALRE

                    51                                                 100
Tetrahymena         ..k...t...  ..........  .....m.m.n  .i....q...  ..........
Sea Urchin1         ..........  ..........  ..........  ..........  .....s....
O. luteus           ..........  ..........  ...                     ...
Human               ..........  ..........  ..........  ..........  .....c....
Fruit Fly           ..........  ..........  ..........  ..........  .....s....
Sea Urchin2         ..........  ..........  ..........  ..........  .....s....
Coral               ..........  ..........  ..........  ..........  .....s....
Bread Tree          ..........  ..........  ..........  .........a  ..........
Maize               ..k.......  ..........  ..........  .........a  ..........
Pea                 ..k.......  ..........  ..........  .........s  ..........
Alfalfa             ..k.......  ..........  ..........  ......h...  ..........
Caen.elegans        ..........  ...ra.....  ..........  ..........  ..........
Frog                ..........  ..........  ..........  .....r....  .....s....
Cress               ..k.......  ..........  ..........  .........a  ..........
Volvox              ..k.......  ..........  ..........  ......q...  ..........
Yeast3              ..........  ..........  ..........  .........g  .....v....
Aspergillus         ..........  ..........  .........s  .........g  ....sv....
Neurospera          ..........  ..........  .........s  .........g  l...sv.s..
Yeast4              ..........  ..........  ..........  .........g  ....sv....
CONSENSUS           IRRYQKSTEL  LIRKLPFQRL  VREIAQDFKT  DLRFQSSAVM  ALQEAAEAYL
```

```
                101                                                135
Tetrahymena     ..........  ....r.....  t..m.....  ....f
Sea Urchin[1]   ..........  ..........  ..........  .....
O. luteus       ..........  ..........  .
Human           ..........  ..........  ..........  .....
Fruit Fly       .a........  ..........  ..........  .....
Sea Urchin[2]   .r........  ..........  ..........  .....
Coral           ..........  ..........  ..........  .....
Bread Tree      ..........  ..........  ..........  .....
Maize           ..........  ..........  ..........  .....
Pea             ..........  ..........  ..........  .....
Alfalfa         ..........  ..........  ..........  .....
Caen.elegans    ..........  ..........  ..........  .....
Frog            .a........  ..........  ..........  .....
Cress           ..........  ..........  ..........  .r...
Volvox          ..........  ..........  ..........  .....
Yeast[3]        .s........  ...g.....q  ...m.....l  ....s
Aspergillus     .s........  .........q  s........l  ....s
Neurospera      .s........  .........q  s........l  ....n
Yeast[4]        .s.......a  .........q  k...k....l  ....s
CONSENSUS       VGLFEDTNLC AIHAKRVTIM PKDIQLARRI RGERA
```

Key to superscripts:
1: Sand Sea Urchin
2: Purple Sea Urchin
3: Fission Yeast
4: Baker's Yeast

FIGURE 3.4 Comparison of histone H3 sequences aligned for maximum homology by the GCG program. H3 sequences from the following sources were compared (common name and H3 type, proper name, accession number for "Swissprot" GCG data bank): *Tetrahymena pyriformis* H3.1 (P15512); Sand Sea Urchin H3 (100% homologous to Common Sea urchin and Sea Star) *Psammechinus miliaris* (P02298); *Olisthodiscus luteus* H3; Human H3.1 (100% homologous to Mouse and Bovine) *Homo sapien* (P16106); Fruit Fly H3 *Drosophila melanogaster* (P02299); Purple Sea Urchin H3 *Strongylocentrotus purpuratus* (P06352); Staghorn Coral H3 *Acropora formosa* (P22843); Altenstein's Bread Tree H3 *Encephalartos altensteinii* (P08903); Maie H3 *Zee mays* (P05203); Pea H3 *Pisum sativum* (P02300); Alfalfa H3.2 *Medicago sativa* (P11105); *Caenorhabditis elegans* H3 (P08898); African Clawed Frog H3.2 *Xenopus laevis* (P02302); Mouse-ear Cress H3.1 *Arabidopsis thaliana* (P05329); *Volvox carteri* H3 (P08437); Fission Yeast H3.1 *Schizosaccharomyces pombe* (P09988); *Aspergillus nidulans* H3 (P23753); *Neurospera crass* H3 (P07041) and Baker's Yeast H3 *Saccharomyces cerevisiae* (P02303).

3.6 *Olisthodiscus luteus* Histone H4

The H4 histones are the smallest and the most conserved of all the histones. The H4 molecules of human are 100 % homologous to bovine, pig, mouse, chicken, duck, frog, trout and clam worm. Consequently only human H4 has been shown in the sequence alignment (figure 3.5). Similarly, the H4s from wheat, maize, pea and mouse-ear cress are all homologous. Only wheat has been shown in the figure. In almost all organisms of which the primary structure of H4 has been solved, the length of the molecule was found to be 102 amino acids long.

Bavykin *et al.* (1985) determined through chemical cross-linking experiments that the two H4 proteins bind to the DNA at a distance of 1,5 helical turns either side of the dyad axis. The point of cross-linking was found to be the highly conserved histidine at residue 18. (Ebralidse *et al.*, 1988). This histidine is found in the most basic domain of the H4: Lys-Arg-His-Arg-Lys-Val-Leu-Arg. This domain is not conserved in the *Tetrahymena* H4s, but the distribution of basic amino acids remains conserved.

As briefly discussed in section 3.5., the two H4 proteins are bound to two H3 histones to form a tetramer. Böhm *et al.* (1977) established that residues 38 - 102 are involved with the formation of this tetramer. The first 20 residues of the N- teminal were shown to be unimportant for the formation of this complex (Econtre and Parello, 1988). This is to be expected, as this domain was shown to bind to the DNA as discussed above. This domain has been proposed to be "fixed" in the nucleosomal structure, unlike the mobile terminal domains of H2A, H2B and H3 (Smith and Rill, 1989). The non-mobile, highly conserved nature of the N-terminal probably reflects the importance of its function i.e. playing a central role in the formation of the H3 - H4 tetramer and in binding to the DNA to form a nucleosomal precursor.

Böhm *et al.* (1977) established that the last 18 residues of the C- terminal were essential for the formation of the tetramer. However, this domain was later proposed to bind to the H2B during octamer formation (Martinson *et al.*, 1979). The importance of this domain is reflected in its conserved nature. An evolutionary pressure appears to be present to maintain its primary structure - not only to facilitate binding to H2B, but also, apparently, to help maintain the

structure, or folding, of the central domain which is the binding site to histone H3. The latter finding has been deduced from the evidence that no stable tetramer is formed once the C- terminal of H4 has been removed (Böhm, 1977). Unlike the other histones, H4 N- and C- terminals are more highly conserved across the evolutionary table than the central domain.

The complete primary structure of the *Olisthodiscus luteus* H4 has been solved. Only two differences emerge in primary structure when compared to human (and most other vertebrates') H4s. Lysine at residue 21 (figure 3.5.) has changed to arginine in the *Olisthodiscus luteus* sequence. Similarly, lysine at residue 78 is also arginine in the case of the algae. Arginine at position 21 is shared only by *Tetrahymena* H4.1. Arginine at position 78 is homologous to wheat, maize, pea, physarum, volvox and *Tetrahymena*. The N- and C- terminals of the *Olisthodiscus luteus* H4 are 100 % conserved when compared to a wide range of organisms across the evolutionary table.

Modified amino acids were found at 3 positions in the *Olisthodiscus luteus* H4 protein. Residue 1 is assumed to be acetylated serine because i) the N- terminal is blocked to Edman degradation, ii) the residue is "unblocked" by HFBA treatment and iii) acetylated serine as the first residue of H4 is highly conserved. Allis *et al.*, (1985) determined that newly synthesized H4 is diacetylated and proposed that this plays a conserved and highly significant role in histone deposition. Acetylation was also found at lysine (residue 16). Acetylation at this residue is not unique. However, for the *Olisthodiscus luteus* H4, acetylated lysine accounts for 90 % of the lysine. DeLange *et al.* (1969) have found that only 50 % of the lysines are acetylated in calf H4 at this position. It may be significant that this modification occurs at a very basic sequence: (16) AcLys-Arg-His-Arg-Arg (20), effectively destroying the basicity of the lysine. This may provide a means by which the degree of H4 binding to DNA could be regulated.

In most organisms studied, lysine at residue 20 occurs in a methylated form - usually dimethyl lysine (DeLange *et al.*, 1969; Ogawa *et al.*, 1969). This is not possible for *Olisthodiscus luteus* H4 as this residue is arginine. It is possible that the function of this modification at lys 20 may be achieved by methylation of another lysine. Lysine at residue 79 in the *O.luteus* H4 was found to be 100 % dimethylated. This appears to be a unique site for modification. If the structure of H4 in the tetramer is such that positions 20 and 79 are close

together in three-dimensional space, it is possible that dimethylation at Lys 79 may achieve the same effect as the normally conserved dimethylation at Lys 20. However, this is speculation and may well be a completely independent event.

The *Olisthodiscus luteus* H4 is highly conserved when compared with both vertebrates and higher plants (figure 3.5.). Such conservatism of H4 does not lend itself well to evolutionary studies, especially comparisons of lines that have diverged more recently. The GCG distance program finds that the *Olisthodiscus luteus* H4 is most closely related to both wheat and human H4s (Table 3.4).

```
                1                                                  50
Caen.elegans    ..........-.........-.........-.........-.........
Sea Urchin      ..........-.........-.........-.........-.........
Human           ..........-.........-.........-.........-.........
Wheat           ..........-.........-.........-.........-.........
O. luteus       ..........-.r.......-.........-.........-.........
Fruit Fly       t.........-.........-.........-.........-.........
Slime Mold      ..........-.........-.........-.........-.....nt..
Volvox          ..........-.........-.........-.........-.........
Aspergillus     ..........-.........-.........-.........-.....a...
Neurospora      t.........-.........-.........-.........-.....a...
Yeast1          ..........-.........-.........-.........-.........
Yeast2          ..........-.........-.........-.........-.....a...
Tetrahymena     -ag.......-..v......k-rsnkas.e..-.........-.....s...
Oxytricha       a....v...y-..v.....tk-.s.k.t.m..-.........-.....s...
CONSENSUS       SGRGKGGKGL-GKGGAKRHSR-KVLRDNIQGI-TKPAIRRLAR-RGGVKRISGL

                51                                                 100
Caen.elegans    ..........-.........-...c.....-.........-.........-...
Sea Urchin      ..........-.........-...c.....-.........-.........-...
Human           ..........-.........-.........-.........-.........-...
Wheat           ..........-.........-.....r..-.........-.........-...
O. luteus       ..........-.........-.....r..-.........-.........-...
Fruit Fly       ..........-.........-.........-.........-.........-...
Slime Mold      .........t-.........-.....r..-.........-.........-...
Volvox          .....t....-n........s-.....r..-.........-.........-...
Aspergillus     .........t-...g.....-.........-...s.....-.........-...
Neurospora      .........t-...g.....-.........-...s.....-.........-...
Yeast1          ....v.a...-s...s....s-.........-...s.....-.........-...
Yeast2          ......a...-l.........-.........-...s.....s.....i..-...
Tetrahymena     ....s.q...-s.........-.....r..-.........-.........-...
Oxytricha       ......n..r-s........s-.........-.........-.........-...
CONSENSUS       IYEETRGVLK-VFLENVIRDA-VTYTEHAKRK-TVTAMDVVYA-LKRQGRTLYG-FGG
```

Key to superscripts:
1: Baker's Yeast
2: Fission Yeast

FIGURE 3.5 Comparison of histone H4 sequences aligned for maximum homology by the GCG program. H4 sequences from the following sources were compared (common name and H4 type, proper name, accession number for "Swissprot" GCG data bank): *Caenorhabditis elegans* H4 (P18678); Sand Sea Urchin H4 (100% homologous to Painted and Purple Sea Urchins) *Psammechinus miliaris* (02306); Human H4 (100% homologous to Bovine, Pig, Rat, Mouse and Chicken) *Homo sapien* (P02304); Wheat H4 (100% homologous Maize, Pea and Mouse-Ear Cress) *Triticum aestivum* (P02308); *Olisthodiscus luteus* H4; Fruit Fly H4 *Drosophila melanogaster* (P02307); Slime Mold H4 *Physarum polycephalum* (P04915); *Volvoc carteri* H4 (P08436); *Aspergillus nidulans* H4 (P23750); *Neurospora crassa* H4 (P04914); Baker's Yeast H4 *Saccharomyces cerevisiae* (P02309); Fission Yeast H4 *Schizosaccharomyces pombe* (P09322); *Tetrahymena pyriformis* H4.1 (P02310) and *Oxytrich nova* H4 (P18836).

O.luteus	H4
Caenorhabditis elegans	0.9806
Sea Urchin	0.9806
Human	0.9903
Wheat	0.9903
Fruit Fly	0.9806
Slime Mold	0.9612
Volvox	0.9612
Aspergillus	0.9612
Neurospera	0.0515
Baker's Yeast	0.9417
Fission Yeast	0.9706

TABLE 3.4 Comparison of the complete *O.luteus* H4 protein compared to sequences aligned in Figure 3.5. and as discussed in section 3.1.2. 100 % homology has a value of 1. Proper names and "Swissprot" accession numbers are as given in the legend of Figure 3.5.

3.7 Evolution of the *Olisthodiscus luteus* Histones

The primary structure of conserved, ubiquitous proteins have been proposed to reflect evolutionary relationships among various taxonomic groups. Phylogenetic trees, constructed after analysis of cytochrome c proteins, were found to be consistent with the fossil record and morphological data (Dickerson, 1971; Schwartz and Dayhoff, 1978).

The dating of the phylogenetical trees, i.e. the determination of the time when two divergent lines branched, is achieved by comparing with paleontological data. The accuracy of dating divergent evolutionary lines is therefore dependant on the accuracy of the fossil record. The validity of phylogenetic trees determined from protein primary structure analysis, is again, dependent on the molecular clock hypothesis. The molecular clock, or rate constancy hypothesis (Wilson *et al.*, 1977), implies that for a given protein, the rate of amino acid substitutions remains constant when averaged over a sufficiently long time, and that the substitutions are independent of the generation time of the organism. After studying substitution rates in the bacteria *Escherichia coli*, the results of Hartl and Dykhuizen (1979) appear to support this hypothesis. However, Wu and Li (1985) found higher rates of substitution in rodents than in humans after comparing coding regions of 11 genes. Whether this is due to the generation time difference between these two mammals, or as a result of some other influence still remains to be determined. The validity of the molecular clock remains a controversial issue. For the purpose of this study, the rate constancy hypothesis has been accepted and applied to the *O.luteus* histones.

Histone proteins appear to be suitable for phylogenetical studies. They are highly conserved and are found in all but a few eukaryotes.

Phylogenetic analysis utilizing histone primary structures has been previously conducted. Although the results are generally consistent with similar analyses performed using cytochrome c sequences (Dickerson, 1971), some startling exceptions have arisen: Yeast H3 (Brandt and von Holt, 1982) appears to have diverged 4 000 million years ago, while *Tetrahymena* H4 appears to have diverged 6 000 million years ago (Glover and Gorovsky, 1979). Rodrigues (1985) calculated that animals and plants diverged 3 000 million years ago, in contrast to the 1 200 million years ago deduced from cytochrome c data (Dickerson, 1971). These results are clearly incompatible with the postulate that all life had, at some stage, a

common ancestor called the progenotes, which may have existed some 3 500 million years ago (Woese and Fox, 1977; Woese *et al.*, 1990). Various explanations for these anomalies, which consistently place divergence times earlier than previously thought, have been proposed (Brandt and von Holt, 1982; Goodman, 1981).

It is possible that the molecular clock is an invalid concept and that generation time has an effect on the rate of nucleotide substitution rate (Brandt and von Holt, 1982). The possibility also exists that the molecular clock has slowed down in more recent times, or has undergone a series of acceleration and deceleration phases (Goodman, 1981).

A third possibility exists, which appears to have received little attention in the literature. The organisation of nuclear DNA varies throughout the eukaryotic world in terms of size, percentage of unique DNA, tertiary packing and the amount of active chromatin. It is possible that some organisms may tolerate more mutations in their histone sequences than another organism which requires a higher stringency for their survival. If this is the case, the rate constancy hypothesis cannot be generally applied. The function of the histones may be conserved, but it is possible that the specificity is not. An organism that tolerates more mutations will therefore appear to have diverged earlier than it did compared to an organism that has a lower PAM.

Finally, the possibility that eukaryotic life is in fact older than previously thought does exist. If this is the case, the divergence time of 4 000 million years ago as calculated for organisms such as yeast (Brandt and von Holt, 1982) maybe nearer to the correct time than currently believed.

It is clear that determining the divergence time of *O.luteus* histones from its ancestral line has to be treated with some caution. However, an indication of the ancestral line should become apparent.

The method used here for constructing phylogenetic trees is the difference matrix method described by Dickerson (1971) for cytochrome c. The matrix of differences was determined after the relevant sequences had been aligned by the GCG program. Differences occurred between two sequences when amino acids at a given position differed, or when one sequence contained a gap at that position. Differences were recorded as a percentage of the protein length compared. The resulting

differences where then corrected for the possibility of superimposed mutations. This was achieved by applying the expression $N_{corr} / 100 = -\ln[1 - N_{raw} / 100]$, where N_{raw} represents the uncorrected percentage difference and N_{corr} the corrected percentage difference (Dickerson, 1971).

H3 and H4 are highly conserved proteins. The entire H4 sequence and all domains homologous to the sequenced peptides of H3 were used for sequence comparisons. The H1 and H2A peptides sequenced were homologous to the globular domain and were all also used for determining differences. As H2B is made up of distinct domains of different functions (see section 3.4), only the conserved domain between residues 57 and 154 (figure 3.6.) was used for comparison.

The corrected difference matrix of H1, H3, H4, H2A and H2B are shown in tables 3.5 to 3.9, respectively. By inspection, it is clear that comparisons of H1, H3 and H4 sequences would provide little useful information in constructing a phylogenetic tree. The H1 sequence, for the domain compared, shows over 100 % corrected sequence differences and is therefore not conserved enough to make useful comparisons. The histones H3 and H4 exhibit the antithesis. The sequences are too highly conserved - there are no differences between the compared domains of calf, human, rat and *O.luteus* H3!. Clearly such conservatism provides little grounds for establishing divergence.

Phylogenetic trees were constructed from the corrected difference matrix of H2A and H2B. This was achieved by determining the average differences between the divisions of species compared, i.e. mammals, aves, pisces, echinodermata, mollusca, insecta and plantae (Tables 3.10 and 3.11).

Assuming the molecular clock hypothesis is correct, a time scale may be defined in terms of the percentage of accepted point mutations (UEP) accumulated in 10^8 years (Doolittle, 1979). The calibration of UEP units to years is derived from paleontological studies. The dates used are as follows:

i) Divergence of mammals from reptiles and birds at 300 million years ago (Young, 1962).

ii) Divergence of higher vertebrates and fish at about 400 million years ago (Young, 1962).

iii) Divergence of vertebrates and insecta at 600 million years ago (Young, 1962).

Using the number of corrected differences that correspond to the above dates, a "best straight line" graph (y = bx + c) was constructed where y = the number of

MOUSE	1	0														
BOVINE	2	0	0													
RABBIT	3	6	6	0												
RAT	4	0	0	6	0											
DUCK	5	0	10	6	10	0										
CHICKEN	6	3	13	8	13	2	0									
TROUT	7	22	22	24	22	22	19	0								
SEA URCHIN	8	94	94	94	94	94	94	94	0							
SURF-CLAM	9	122	122	130	122	122	130	110	99	0						
MUSSEL	10	104	104	104	104	94	94	84	89	94	0					
MIDGE	11	75	75	75	75	84	84	56	75	129	84	0				
FRUIT FLY	12	94	94	94	94	99	99	56	94	122	89	47	0			
O LUTEUS B	13	104	104	99	104	104	104	89	56	104	84	89	89	0		
O LUTEUS A	14	99	99	94	99	99	99	104	53	94	89	75	80	64	0	
MAIZE	15	137	137	137	137	145	145	67	110	71	80	99	104	99	94	0
PEA	16	129	129	122	129	153	153	116	84	129	122	94	99	71	64	47	0
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

TABLE 3.5 Corrected Percentage Difference Matrix for Histone H1. Values are mutation distances as determined by amino acid sequence comparisons of H1 proteins from the following sources (common name, proper name in italics, GCG-"Swissprot" Accession number): Mouse *Mus musculus* P158641; Bovine *Bos taurus* P02253; Rabbit *Oryctolagus cuniculus* P02251; Rat *Rattus norvegicus* P15865; Duck *Anas platyrhynchos* P09426; Chicken *Gallus gallus* P09987; Rainbow Trout *Salmo Gairdneri* P06350; Purple Sea Urchin *Strongylocentrotus purpuratus* P07796; Atlantic Surf Clam *Spisula solidissima* P22975; Californian Mussel *Mytilus californianus* P22974; Midge *Chironomus thummi thummi* P21895; Fruit Fly *Drosophila Hydei* P17268; *Olisthodiscus luteus* H1b; *Olisthodiscus luteus* H1a; Pea *Pisum sativum* P08283; Maize *Zea mays* P23444.

HUMAN	1	0										
FRUIT FLY	2	3	0									
SAND SEA URCHIN	3	0	3	0								
PURPLE SEA URCHIN	4	1	3	1	0							
O LUTEUS	5	0	3	0	1	0						
ALFALFA	6	3	5	3	4	3	0					
CRESS	7	1	4	1	3	1	1	0				
PEA	8	1	4	1	3	1	1	0	0			
MAIZE	9	1	4	1	3	1	1	0	0	0		
VOLVOX	10	8	10	8	9	8	8	7	7	7	0	
YEAST	11	9	10	9	9	9	11	11	11	11	18	0
		1	2	3	4	5	6	7	8	9	10	11

TABLE 3.6 Corrected Percentage Difference Matrix for Histone H3. Values are mutation distances as determined by amino acid sequence comparisons of H3 proteins from the following sources (common name, proper name in italics, GCG-"Swissprot" Accession number): Human (100% homology to Bovine and Mouse) *Homo sapien* P16106; Fruit Fly *Drosophila melanogaster* P02299; Sand Sea Urchin *Psammechinus miliaris* P02298; Purple Sea Urchin *Stronglocentrotus purpuratus* Po6352; *Olisthodiscus luteus*; Alfalfa *Medicago sativa* P11105; Mouse Ear Cress; *Arabidopsis thaliana* P05330; Garden Pea *Pisum sativum* P02300; Bakers Yeast *Saccharomyces cerevisiae* P02303; Maize *Zea mays* P05203; *Volvox carteri* P08437.

CAENORHABDITIS	1	0												
SEA URCHIN	2	0	0											
HUMAN	3	1	1	0										
WHEAT	4	3	3	2	0									
O LUTEUS	5	3	3	2	2	0								
FRUIT FLY	6	2	2	1	3	3	0							
SLIME MOLD	7	5	5	4	3	4	5	0						
VOLVOX	8	5	5	4	3	4	5	5	0					
ASPERGILLUS	9	8	8	7	8	9	8	7	10	0				
NEUROSPORA	10	9	9	8	9	10	7	8	11	1	0			
BAKERS YEAST	11	9	9	8	9	10	9	11	8	7	8	0		
FISSION YEAST	12	10	10	9	10	11	10	11	11	7	8	8	0	
TETRAHYMENA	13	25	25	24	23	22	24	23	22	28	28	28	30	0
		1	2	3	4	5	6	7	8	9	10	11	12	13

TABLE 3.7 Corrected Percentage Difference Matrix for Histone H4. Values are mutation distances as determined by amino acid sequence comparisons of H4 proteins from the following sources (common name, proper name in italics, GCG-"Swissprot" Accession number): *Caenorhabditis elegans* P18678; Sand Sea Urchin *Psammechinus miliaris* P02306; Human *Homo sapiens* P02304 (100% homologous to Bovine, Mouse, Rat, Pig, Chicken, Duck, Frog, Trout); Wheat *Triticum aesitivum* P02308 (100% homologous to Maise, Pea and Mouse Ear Cress); *Olisthodiscus luteus*; Fruit FLy *Drosophila melanogaster* P02307; Slime Mold *Physarum polycephalum* P04915; *Volvox carteri* P08436; *Aspergillus nidulans* P23751; *Neurospora crassa* P04914; Bakers Yeast *Saccharomyces cerevisiae* P02309; Fission Yeast *Scizosaccharomyces pombe* P09312; *Tetrahymena pyriformis* P02310

HUMAN	1	0														
MOUSE	2	2	0													
RAT	3	0	2	0												
CHICKEN	4	0	2	0	0											
DUCK	5	0	2	0	0	0										
TROUT	6	4	4	4	4	4	0									
SEA URCHEN (P)	7	7	7	7	7	7	7	0								
SEA URCHIN (PU)	8	9	9	9	9	9	5	5	0							
SEA URCHIN (SA)	9	5	5	5	5	5	2	9	7	0						
MIDGE	10	7	7	7	7	7	7	7	5	9	0					
FRUIT FLY	11	7	7	7	7	7	7	7	7	5	9	0				
O LUTEUS	12	9	9	9	9	9	9	11	9	5	15	15	0			
WHEAT	13	23	23	23	23	23	23	21	21	25	15	15	32	0		
PEA	14	21	21	21	21	21	17	21	21	21	15	15	23	17	0	
VOLVOX	15	15	15	15	15	15	17	17	19	17	17	17	15	23	19	0
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

TABLE 3.8 Corrected Percentage Difference Matrix for Histone H2A. Values are mutation distances as determined by amino acid sequence comparisons of H2A proteins from the following sources (common name, proper name in italics, GCG-"Swissprot" Accession number): Human *Homo sapien* P02261 (100% homologous to Bovine); Mouse *Mus musculus* P22752; Rat *Rattus norvegicus* P02262; Chicken *Gallus gallus* P02263; Muscovy Duck *Cairina moschata* P13912; Salmon Trout *Salmo gairdneri* P02276; Painted Sea Urchin *Lytechinus pictus* P09589; Purple Sea Urchin *Strongylocentrotus purpuratus* P16886; Sand Sea Urchin *Psammechinus miliaris* P02271; Midge *Chironomus thummi thummi* P21896; Fruit Fly *Drosophila melanogaster* P02267; *Olisthodiscus luteus*; Wheat *Triticum aestivum* P02276; Pea *Pisum sativum* P25470; *Volvox carteri* P16866.

HUMAN	1	0													
MOUSE	2	1	0												
DUCK	3	5	6	0											
CHICKEN	4	4	5	1	0										
SEA URCHIN (AN)	5	27	27	27	26	0									
SEA URCHIN (PU)	6	27	27	27	25	11	0								
SEA URCHIN (SA)	7	14	16	16	14	18	18	0							
STARFISH	8	14	14	16	14	18	18	13	0						
MIDGE	9	12	12	13	12	22	22	14	9	0					
FRUIT FLY	10	10	10	11	10	19	19	14	7	3	0				
LIMPET	11	11	11	12	11	21	18	16	9	5	4	0			
O LUTEUS	12	23	24	24	23	27	21	23	23	23	21	22	0		
WHEAT	13	36	37	34	33	46	44	36	32	30	28	31	31	0	
VOLVOX	14	31	31	33	31	40	37	34	34	31	30	33	34	26	0
		1	2	3	4	5	6	7	8	9	10	11	12	13	14

TABLE 3.9 Corrected Percentage Difference Matrix for Histone H2B. Values are mutation distances as determined by amino acid sequence comparisons of H2B proteins from the following sources (common name, proper name in italics, GCG-"Swissprot" Accession number): Human *Homo sapien* (100% homologous to Bovine) P02278; Mouse *Mus musculus* P10853; Muscovy Duck *Cairina moschata* P14001; Angulate Sea Urchin *Parechinus angulosus* P02287; Spiny Starfish *Marthasterias glacialis* Po2283; Midge *Chironomus thummi thummi* P21897; Fruit Fly *Drosophila melanogaster* P02283; Sandpaper Limpet *Patella granatina* P02284; *Olisthodiscus luteus*; Wheat *Triticum aestivum* P05621; *Volvox carteri* P16868

MAMMALS	1	0							
BIRDS	2	0,6	0						
FISH	3	4,0	4,0	0					
ECHINODERMATA	4	7,1	7,1	4,7	0				
INSECTA	5	7,1	7,1	7,1	7,1	0			
O LUTEUS	6	9,0	9,0	9,0	8,4	14,8	0		
PLANT	7	22,1	22,1	20,0	22,0	14,8	27,7	0	
ALGAE	8	14,8	14,8	16,9	17,5	16,9	14,8	21,0	0
		1	2	3	4	5	6	7	8

TABLE 3.10 Average Percentage Difference Matrix of H2A.
The H2A amino acid sequences compared in Figure 3.4 were clustered in defined groups and the average differences between one group and all the others was determined.

MAMMALS	1	0							
BIRDS	2	5,3	0						
MOLLUSCA	3	10,9	11,5	0					
INSECTA	4	10,9	11,5	4,8	0				
ECHINODERMATA	5	20,7	20,7	15,7	15,9	0			
O LUTEUS	6	23,1	23,7	21,8	21,8	23,4	0		
WHEAT	7	35,5	33,3	31,2	24,6	39,5	31,2	0	
ALGAE	8	31,2	31,4	32,6	30,5	36,3	34,0	25,7	0
		1	2	3	4	5	6	7	8

TABLE 3.11 Average Percentage Difference Matrix of H2B.
The H2B amino acid sequences compared in Figure 3.5 were clustered in defined groups and the average differences between one group and all the others was determined.

million years since divergence, c = y'- bx', x = number of corrected differences per 100 residues, y' and x' are mean values of x and y, and b = [y(x - x')] / [(x - x')]^2.

The graph for data from both H2A and H2B is shown in figure 3.6. It would be expected that the graph of average rates of change would intersect the graph at 0,0. For both sets of data presented here, the "best fit line" intersects the y-axis at a positive value. Several explanations are possible:

i) The number of calibration points are to few too be statistically accurate. This will improve as more paleontological data becomes available.

ii) The population of histone sequences studied is too small, e.g. only one pisces histone sequence was available for comparison.

iii) The molecular clock is not regular, but rather operates in a step-like fashion. A period of sudden change is followed by a period of no change. The mean change over a long period would be linear.

iv) The molecular clock is not linear, but follows random periods of acceleration and deceleration.

The unit evolutionary period (UEP) is defined as the time required for 1 % acceptable amino acid changes to occur. For the domains studied here, it was found that the UEP for H2A is 68,7 ± 23 X 10^6 years and 60,6 ± 26 X 10^6 years for H2B. The phylogenetic tree, calibrated to time using the UEP values, is shown for H2A in figure 3.7. and for H2B in figure 3.8.

The general topology of these reconstructions compare favorably with that of McLaughlin and Dayhoff (1973) reconstructed from cytochrome c data, with a few exceptions. Comparing the evolutionary tree of histone H2A and cytochrome c , the late divergence of aves and mammals are in agreement, as is the early divergence of insecta from the animalia line. Unfortunately, McLaughlin and Dayhoff (1973) do not include the echinodermata in their phylogenetic tree. The H2A tree places the echinodermata divergence from the animilia line slightly more recently than the insecta. The cytochrome c data places the divergence of algae to be on the plant line soon after the divergence of animals and plants. Contrary to this, the H2A phylogenetic reconstruction places the algae (volvox) divergence along the animal line. Even more surprising is the placing of the algae *Olisthodiscus*

luteus divergence closer to the time that insecta diverged, than to the time volvox diverged. According to the H2A reconstruction, plants and animals diverged 1 774 million years ago. This is slightly earlier than the 1 300 million years proposed by Dickerson (1971), but much later than the 3 000 million years proposed by Rodrigues (1985). *O.luteus* appears to have diverged from the animal line 954 million years ago - far more recently than the divergence of plants and animals.

Similarly, when the H2B and cytochrome c trees are compared, both place the divergence of aves and mammals as a comparatively recent event. According to the H2B tree, mollusca and insecta diverged from a common ancestor, but the cytochrome c data place the divergence of molluscs as being more recent than the early diverging insecta. The placing of mollusc divergence in the H2B tree is probably an artifact due to only one mollusc sequence being compared.

Unlike the H2A tree, the H2B tree places the divergence of volvox along the plant line. This is in keeping with the cytochrome c data. Again, however, the divergence of *Olisthodiscus luteus* is placed along the animal line, slightly earlier than the divergence of the echinodermata. According to the H2B evolutionary tree, plants and animals diverged 1 975 million years ago, and *O.luteus* diverged from higher animals 1 382 million years ago. As the H2A and H2B proteins evolved concurrently in an organism, an average value for the two would be statistically more accurate. This study finds that plants and animals diverged 1 875 million years ago and *O.luteus* diverged from higher animals about 1 382 million years ago.

Algae may be defined as: "simple, photosynthetic plants with unicellular organs of reproduction" (Abercrombie *et al.*, 1980). It is surprising to find that the H2A and H2B phylogenetic trees both place the divergence of *O.luteus* along the animal line and that this divergence took place far more recently than the divergence of plants and animals, as well as volvox (algae) and plants. The evolutionary trees are derived from average values within a phylum. More surprising is the fact that the most similar sequences to *O.luteus* H2A and H2B proteins (section 3.3 and 3.4) are those from the echinodermata. Sand sea urchin (*Psammechinus miliaris*) H2A was found to be the most similar to that of *O.luteus* H2A, with a corrected percentage difference of 5,31 %. For H2B, *O.luteus* was found to be most similar to purple sea urchin (*strongylocentrotus purpuratus*) with a corrected percentage difference of 20,53 %. It appears that both sets of data indicate *O.luteus*

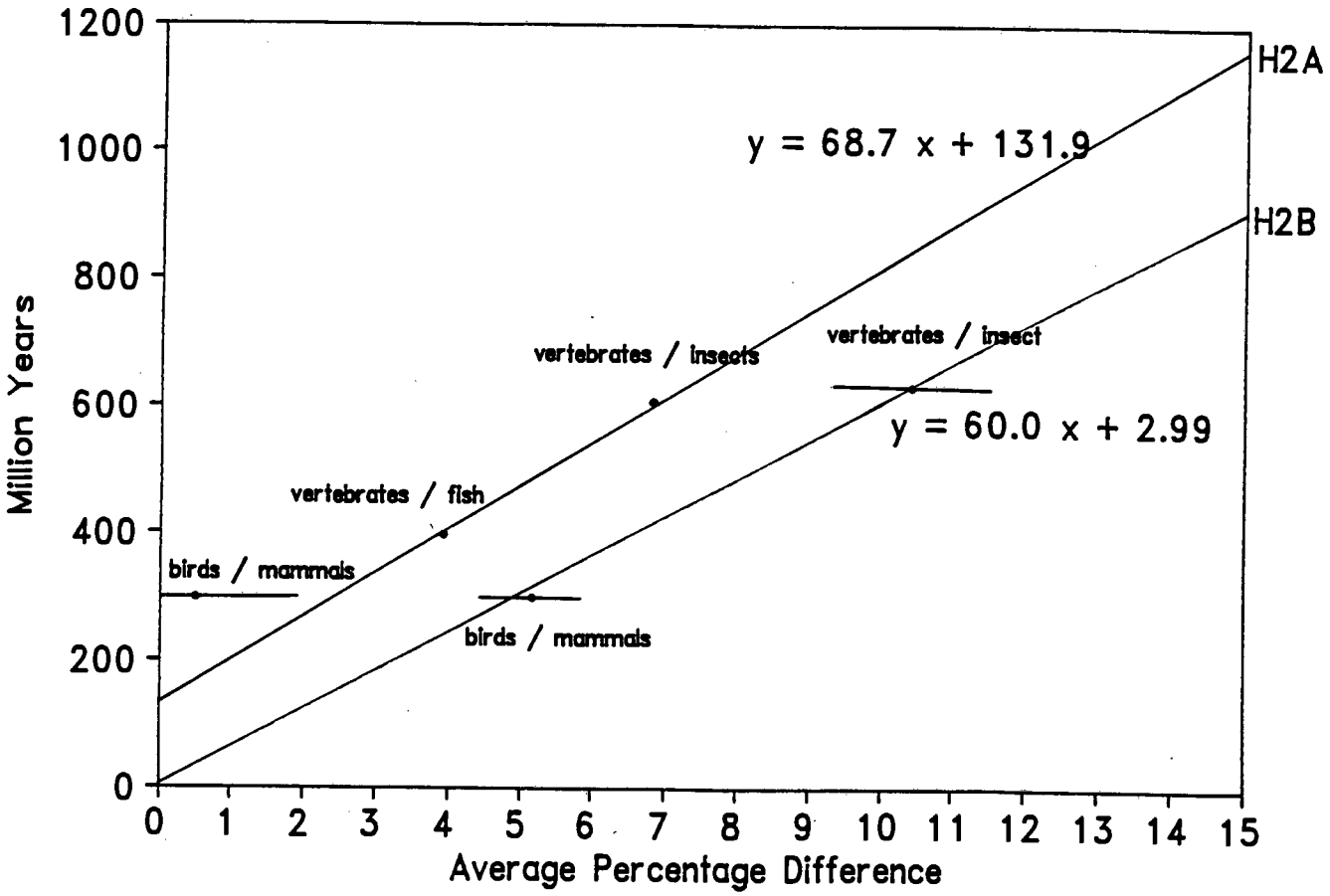

FIGURE 3.6: "Best Straight Line" graph of the accepted point mutations for both H2A and H2B plotted with respect to million years. Bar length represents standard deviation within the sample population.

FIGURE 2.1a: Isolated *O.luteus* nuclei (Section 2.2.) viewed under a Nikon Diaphot-TMD Microscope fitted with a TMD-EF Flourescent attachment. Samples were stained with Hoescht stain prior to photographing. (Magnification X 200)

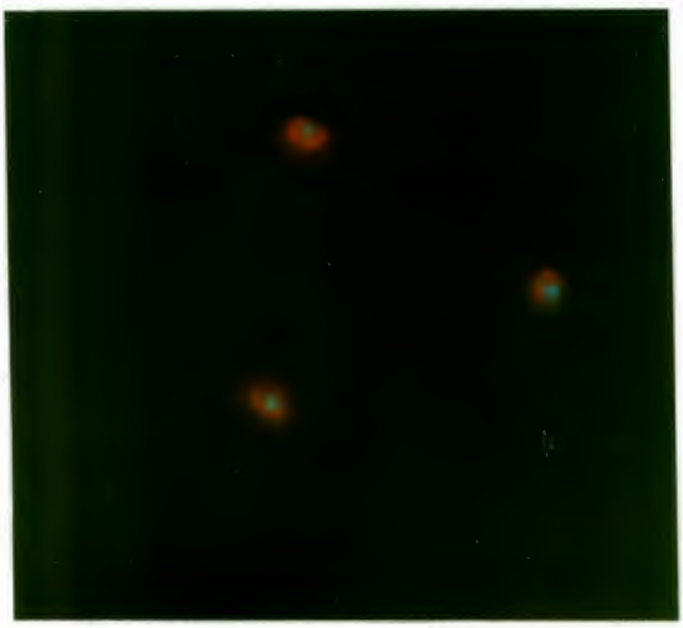

FIGURE 2.1b: Intact *O.luteus* cells, fixed with 0.1% gluteraldehyde and viewed as for Figure 2.1a. (Magnification X 200)

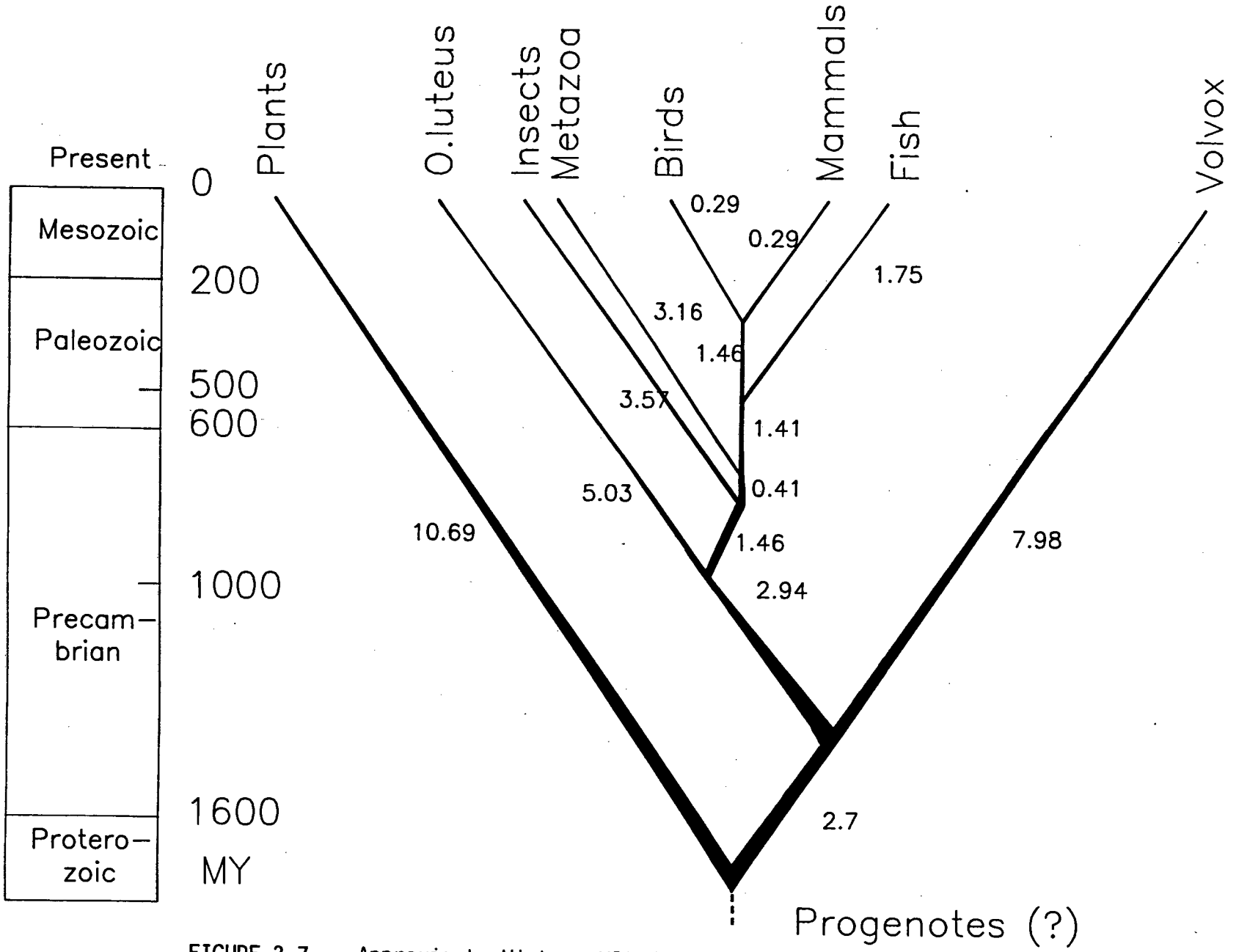

FIGURE 3.7: Approximate Histone H2A derived phylogenetic tree. Leg lengths have been deduced from the average difference matrix shown in Table 3.10..

MY = million years from the present

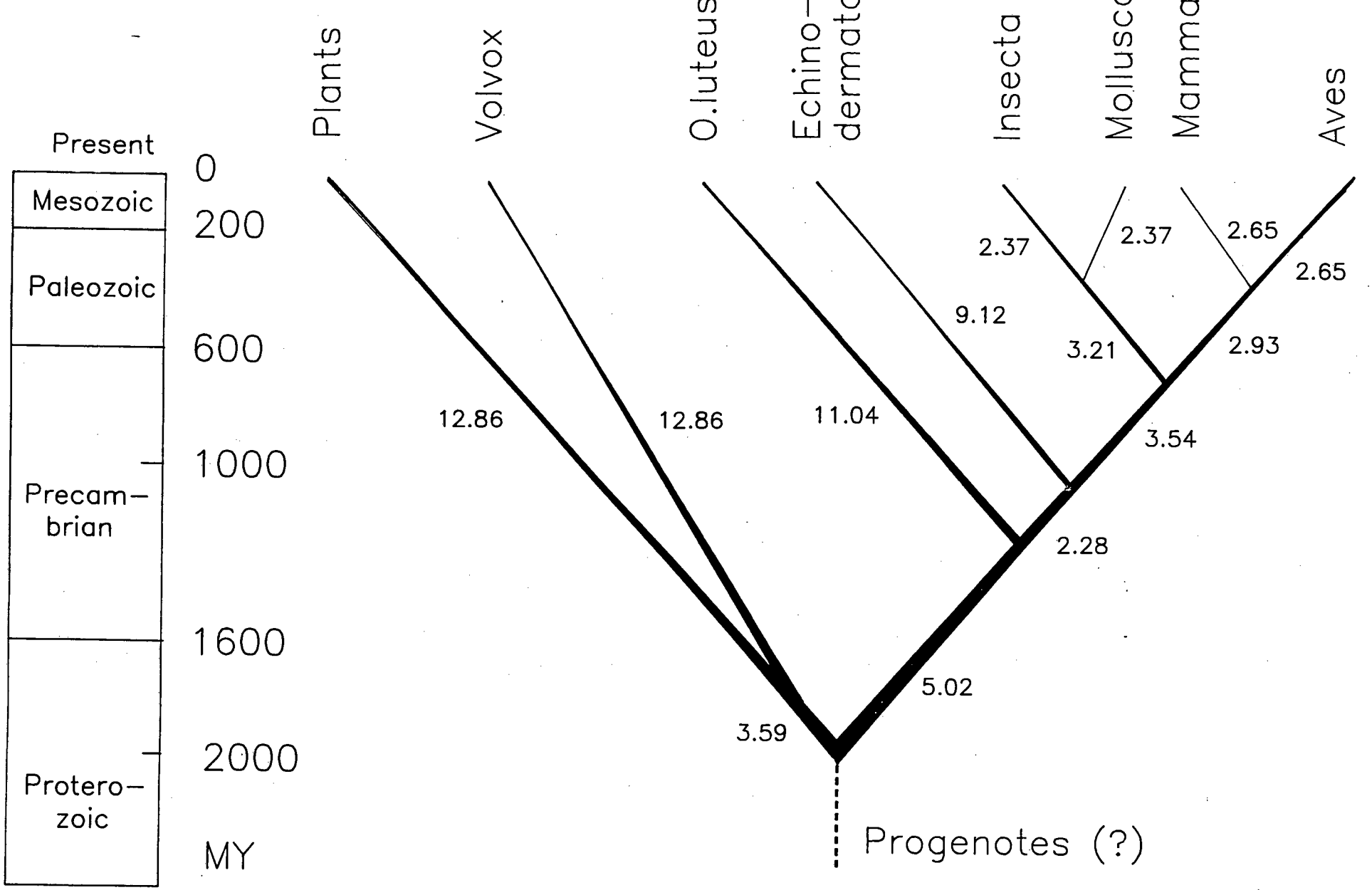

FIGURE 3.8: Approximate Histone H2B derived phylogenetic tree.
Leg lengths have been deduced from the average difference matrix shown in Table 3.11..
MY = million years from the present.

histones (and presumably the organism) to be more closely related to echinodermata than the average percentage matrix would suggest. This apparent anomaly may be explained by several possibilities:-

i) Convergent Evolution

The possibility does exist that both the *O.luteus* and the echinodermata have similar requirements of their genetic material. As DNA packing is mediated by the histones, convergent evolution could be expected. However, considering the morphology and life cycles of the two organisms, this would seem unlikely. Furthermore, electron micrographs of condensed chromatin of *O.luteus* and echinodermata appear distinctly different. This suggests that the DNA accessibility, replication and packing requirements are different and that divergent evolution has occurred.

ii) Horizontal Gene Transfer

A horizontal transfer of the histone genes may have taken place from an echinodermata ancestor to the *O.luteus* ancestor some 960 million years ago. As little is known about the algal gene organization and the arrangement of histone genes and promotors, the likelihood of the occurrence of this phenomenon cannot be discussed at this stage. Horizontal Gene Transfer remains a possibility as to the origin of the *O.luteus* histones H2A and H2B.

iii) Common Ancestor of *O.luteus* and Echinodermata Nuclei

There is little evidence to suggest that the *O.luteus* ancestor later diverged to form the echinodermata, losing its chloroplasts in the process. However, as will be discussed below, there is some evidence to suggest that the echinodermata ancestor gave rise to the *O.luteus* nucleus.

By comparison of sequences of the small-subunit ribosomal RNA coding regions of several chlorophytes and chrysophytes, Gunderson *et al.* (1987) proposed that chrysophytes represent a lineage that diverged from protists shortly before the divergence of animals, plants and fungi, the latter three diverging almost simultaneously. This early divergence of chrysophytes from plants is supported by findings of Markowicz and Loiseaux-de Goer (1991).

These authors did comparisons of the 16S rRNA gene sequences from plastids, which led them to suggest that chromophyte plastids genomes have a composite, common phylogenetic origin with at least two ancestors. The ancestors implicated are the cyanobacteria (blue-green algae) and a ß-proteobacteria (purple bacteria). In contrast, chlorophytes (green algae and land plants) plastid genome ancestor appears to be a cyanobacteria, but distinctly different from that ancestor of the chrysophyte genome. The gene coding for Rubisco (ribulose-1,5-biphosphate carboxylase) small subunit of the proposed chrysophyte *O.luteus* was indeed found to be closely related to the chemautrophic ß-proteobacterium *Alcaligenes eutrophus*, as well as that of the brown algae *Fucus* (Boczar *et al.*, 1989). This close relationship between the plastids of *O.luteus* and brown algae has been confirmed in this laboratory by sequence comparisons of cytochrome C553 purified from *O.luteus* and those found in the "Swissprot" data-bank. (Manuscript in preparation.) *O.luteus* cytochrome C553 was found to be closely related to that of the brown alga *Alaria esculenta* and the yellow-green alga *Bumilleriopsis filiformis*.
It is clear that the chrysophyte lineage is distinctly different from that of terrestrial plants and green algae. The classification of some algae as plants is a simplification that does not reflect the true history of the organism.

The endosymbiotic origin of plastids in modern eukaryotic cells is now a well-accepted concept (Margoliash and Fitch, 1968; Schwemmler and Schenk, 1980; Gray *et al.*, 1984). However, the study of the plastids, as performed above, gives information regarding the history of the plastid only, and not necessarily of the whole organism. It has been suggested that organisms that have chloroplasts surrounded by two double membranes arose from the endosymbiosis of an alga by a non-photosynthetic host (Gibbs, 1962; Whatley and Whatley, 1981). This phenomenon was demonstrated for the alga *Cryptomonad* by Douglas *et al.* 1991. The authors showed this organism to be: "evolutionary chimaeras of two phylogenetically distinct unicellular organisms."

Olisthodiscus luteus has two sets of double membranes surrounding its chloroplasts (Gibbs, 1962). This could suggest that the alga is also a result of two endosymbiotic events. The primary symbiotic event might have been the inclusion of a photosynthesizing ß-proteobacteria into a non-photosynthesizing eukaryotic host cell forming an ancestor of the brown algae. The secondary symbiotic event would be the symbiosis of the photosynthesizing brown algae ancestor (or brown

algae-like), into a non-photosynthesizing eukaryotic host. This study, from data derived from sequencing the conserved histone proteins, strongly suggests the secondary non-photosynthesizing host to be from the echinodermata lineage.

That *O.luteus* arose from an endosymbiotic event some time after the divergence of animals, fungi and plants, as previously proposed for the origin of chrysophytes (Gunderson *et al.*, 1987), may be met with some scepticism. However, one could speculate on the plausibility of this proposal. Both the *O.luteus* and the echinodermata are marine organisms. Both lack a cell wall, which is presumably conducive to the formation of a symbiotic relationship. Algae, when conditions are favourable, are capable of forming "blooms" - a high concentration of the organism in an environment. Similarly, to ensure the survival of the externally fertilising echinodermata, the marine enviroment at spawning time would be rich in gametes and embryos. Should such favorable conditions be met, it is foreseeable that a brown alga could come into close contact with an echinodermata cell or cells. One could speculate the following scenario. The divergence of the *O.luteus* line from the echinodermata line about 900 million years ago, as determined by histone sequence comparisons, presumably would indicate the time when the algal cell was endosymbiotically taken up into the echinodermata host. Presumably, the two organisms would have lived symbiotically for some time prior to this endosymbiosis event. During this time, mutations in the genome of the two eukaryotic cells could provide a vast pool of variants, which as long as the symbiotic relationship was maintained, would be lethal to neither organism. If this was the case, then it would be expected that genes that promote and enhance the symbiosis would also be selected for. If such a scenario did exist, it would not be difficult to speculate that the increasing symbiotic relationship would eventually lead to the incorporation of one cell into another - the ultimate symbiotic relationship.

The fact that *O.luteus* does not have two nuclei such as the dinoflagellates, or has no nucleomorph as do the *Cryptomonads* (Douglas *et al.*, 1991), does not disprove the validity of the above hypothesis. One of the nuclei may have been lost after or just prior to the endosymbiotic event. This, however, would have required the transfer of those genes that made the symbiotic relationship possible, to the surviving nucleus. It is also possible that the two nuclei fused to form a "macro" nucleus with subsequent gradual loss of superfluous DNA. This might explain why *O.luteus*, a relatively simple organism when compared with higher

animals, contains a comparable amount of DNA (1,66 X 10^{-12}g DNA per cell, of which 34 % is unique) when compared with higher eukaryotes (Ersland and Cattolico, 1981). This value is almost double the amount of DNA found in sea urchin sperm cells. Elson *et al.* (1954) determined that the sea urchin *Strongylocentrotus purpuratus* containes 0,77 X 10^{-12} g DNA per cell and that *Lytechinus miliaris* contains 0,87 x 10^{-12} g per cell. Ersland and Cattolico (1981) established that the genome of *O.luteus* consists of 4 % foldback, 58 % repetitive and 34 % single copy sequences. A 3000 nucleotide length fragment of DNA allowed these authors to establish a Cot value of 100 M.s. Ersland and Cattolico, (1981), observe that none of these values are unique to this algae, and that they are comparable to values found in animals.

The most important aspect upon which the above hypothesis is based, is the finding that the histones of *O.luteus* are similar to those of the echinodermata, and that the two diverged at a time far more recently than previously thought for algae. A reader might, and correctly so, point out that evolutionary studies using data from histone sequences, have on occasion, been at odds with other data. Such discrepancies have been seen in wheat, *Tetrahymena* and yeasts, as discussed above (Rodrigues, 1985; Glover and Gorovsky, 1979; Brandt and von Holt, 1982). Although these organisms are all believed to have diverged early, the histone sequence data suggest that they diverged much earlier than had previously been thought. This might be explained by an assumption that the histone clock had been running faster early on in evolution. A survey through the literature does not show any discrepancy between more recently diverging lines. Consequently, more faith can be placed in results indicating a more recent divergence, such as haves been found in this study. Alternatively, it is possible that the alga *O.luteus* histone molecular clock is unique, in that it has slowed down and consequently the histones appear to be more recently diverging. This does not explain, however, the similarity between the algal histones and the echinodermata histones. The slowing down of the *O.luteus* clock, or the speeding up of all other histones' molecular clocks, does not seem likely if one considers the conservation of the histones in general.

The aim of this study was to characterize the histones of the *Olisthodiscus luteus* algae and to try to determine something of its evolutionary past so as to assist in its classification. The former task has been successfully executed. The latter however, has shown that the classification of some algae to be an almost

impossible task. As pointed out by Assali *et al.* (1991): "...phycologists often thought that chromophytes actually include different eukaryotic lineages, (which) form an artificial group..." This study strongly supports this view. As for the classification, distinction has to be made between the host cell and the chloroplast history and characteristics. This study suggests the origin of the *O.luteus* to be much more recent than that of other algae and plants. If this speculation proves to be correct, then the jump from the animal lineage to another independent lineage should be reflected in the organism's classification. It is not yet known whether all chrysophytes originated from a single ancestry secondary-endosymbiotic event (irrespective of when), or whether they arose independently from similar events throughout evolutionary time. If the latter is found to be the case, the classification of the chrysophyte might well be better achieved by the historical method of morphological comparisons, creating an artificial group having only the method of the evolution in common, rather than all having a common ancestor.

This study not only finds the question of evolution and classification of *Olisthodiscus luteus* that needs to be re-addressed, but also the question of genetic transfer between two independant organisms.

CHAPTER 4

MATERIALS AND METHODS

4.1 Solutions and Buffers

4.1.1 Buffers

All chemicals used were reagent grade unless otherwise stated. Provasoli's Enriched Seawater (PES) (Wynne, 1978) consisted of the following, made up from stock solutions. Stock solutions were made up in twice distilled H_2O. Solution concentrations are given in parentheses. 10 mL $NaNO_2$ (35 g/100 mL), 10 mL Na_2glycerophosphate (5 g/100 mL H_2O), 10 mL Vitamin B_{12} (1 mg/100 mL H_2O), 10 mL Thiamine (50 mg/100 mL H_2O), 10 mL Biotin (0,5 mg/100 mL H_2O), 100 mL Tris-HCl, ph 7,8 (5 g/100 mL H_2O), 250 mL Fe (as EDTA 1:1 molar ratio) (351 mg $Fe(NH_4)_2SO_4.6H_2O$ and 300 mg Na_2EDTA / 250 mL H_2O), 250 mL PII Trace Metal solutionn (1,14g H_3BO_3 49 mg $FeCl_3.6h_2O$, 164 mg $MnSO_4.4H_2O$, 22 mg $ZnSO_4.7H_2O$, 48 mg $CoSO_4.7H_2O$, 1 g Na_2EDTA / 1000 mL H_2O). The enriching media was made up to 1 250 mL and stored at 4^oC. Media was added to filtered Atlantic seawater at a ratio of 20 mL media to 1 000 mL seawater. Typically, 100 L of enriched seawater was prepared in 20 X 5 L Schott bottles. Enriched seawater was autoclaved prior to inoculation.

O-3 Medium (McIntosh and Cattolico, 1978) was prepared as follows: 102 g NaCl, 25 g $MgSO_4$, 2,125g $MgCl_2$, 5,65 g $CaCl_2.2H_2O$, 3,8 g KNO_3, 0,25 g KH_2PO_4, 0,15 g $NaHCO_3$, 3,8 mL Stock Solution A, 3,8 mL Stock Solution B, 9,5 mL 1M Tris-HCl ph 7,6, 0,38 mL Vitamin B_{12} (0,1 mg/mL H_2O) made up to5 L with twice distilled H_2O. Stock solution A consisted of 18,6 g Na_2EDTA, 2,4 g $FeCl_3.6H_2O$ in 1 000 mL H_2O. Stock solution B consisted of 40 mg $ZnCl_2$, 600 mg H_3BO_3, 15 mg $CoCl_2.6H_2O$, 40 mg $(NH_4)_6MoO_{24}.4H_2O$ in 1 000 mL twice distilled water. O-3 medium was autoclaved in 5 L Schott Bottles prior to inoculation.

Nuclei Medium A was similar to that used by Nothacker and Hilderbrandt (1985). It consisted of 0,5 M hexyleneglycol, 0,5 % Nonidet P-40, 0,5 mM $CaCl_2$, 15 mM $MgCl_2$,

5 mM EGTA,1 mM PMSF, 0,5 % (v/v) thiodiglycol and 15 mM Tris-HCl, ph 7,3. Nuclei Medium B contained the same as Nuclei Medium A except it contained no Nonidet P-40. 1 M sucrose solution was prepared as for Nuclei Medium B but with the addition of 1 M sucrose. Hexylene glycol was added to stabilize the nuclei membrane (Wray *et al.*, 1977). Nonidet P-40 was found to be more efficient in dissolving cellular and chloroplast membranes, but at 0,5 % (v/v) left the nuclei membrane intact. PMSF was added as protease inhibitor (Panyim and Chalkey, 1976). Thiodiglycol was added to protect the methionines from oxidation by acting as a radical scavenger (Zweidler, 1978). All buffers were made up 24 hours before needed and stored at $4^{o}C$.

4.1.2 Enzymes

The following enzymes were used during the course of this study:
Staphylococcus aureus V8 endoproteinase (Boehringer Meinheim)
Endoproteinase-Asp (Boehringer Meinheim)
Endoproteinase-Arg (Boehringer Meinheim)
Endoproteinase-Lys (Boehringer Meinheim)

4.2 Isolation of *Olisthodiscus luteus* basic nuclear proteins

4.2.1 Culturing of *Olisthodiscus luteus*

Olisthodiscus luteus was obtained as a gift from Professor R.N. Pienaar (University of the Witswaterand). 250 mL of stationary phase culture was used to seed a 5 L Schott bottle containing PES or O-3 Medium. Constant aeration was achieved by the use of domestic aquarium pumps connected in series. Typically, twenty 5 L Schott bottles were inoculated at one time (figure 4.1). Growth was allowed to take place at $22^{o}C$ on a 12 hour dark / 12 hour light regime. Light was

provided by a double bank of 40 W GEC "cool white" fluorescent tubes. Cells were harvested after a cell count of 3-5 X 10^5 cells/mL was achieved.

4.2.2 Cell Harvest

Cultured algal cells were pelleted by means of a Sorvall SS34 rotor and a Szents-Gyorgyi and Blom continuous flow adapter in a Sorvall RC2-B centrifuge set at 700 g at 4^{o}C. Typically 50 L of culture was centrifuged at an adjusted flow rate that ensured a clear supernatant. All subsequent isolation procedures were performed at 4^{o}C.

Figure 4.1: Cultures of *Olisthodiscus luteus*. 100 L of culture maintained at 22^{o}C with constant aeration. Light provided by 40 W GEC "Cool White" flourescent tubes.

4.2.3 *Olisthodiscus luteus* Nuclei Isolation

Nuclei isolation of pelleted *O.luteus* was achieved by a modified method of Rizzo and Bourghardt (1983) utilizing a buffer system similar to that of Nothacker *et al.* (1985) described above. Cells were lysed in 200 mL Nuclear Medium A and subjected to 15 strokes of a loose dounce homogenizer. The suspension was allowed to stir at 4^{o}C for 1 hour and subsequently centrifuged at 755 g for 15 minutes in a fixed angle rotor. The pellet was resuspended in 40 mL of Nuclear Medium A by 5 strokes of a loose fitting homogenizer and layered over 4 X 20 mL 1 M sucrose solutions. Sedimentation of crude nuclei was achieved by differential centrifugation in a swing-out rotor at 7970 g for 15 minutes. The pellet was again resuspended in Nuclear Medium A and again subjected to the 1 M sucrose differential centrifugation step. Finally the nuclei were washed by slowly resuspending in 100 mL Nuclear Medium B and pelleted by centrifugation at 480 g for 10 minutes.

Nuclei, isolated by the method above,were on occasion further purified. This was achieved by utilizing the self generating density gradient properties of Percoll and modification of Nothacker *et al.*'s (1985) nuclei isolation procedure. Prior to the second differential centrifugation step, crude nuclei were slowly resuspended in 3 mL Nuclear Medium A with the aid of a loose fitting dounce homogeniser. This was then brought up to 70 mL with 25 % Percoll prepared in Medium A. After mixing by inversion, nuclei were separated from other cellular debris by centrifugation in a Type 60ti Beckman rotor at 48 000 g for 35 minutes at 0 - 4^{o}C. The nuclear band close to the base of the tube was then gently aspirated off by means of a pasteur pipette. The nuclei were washed in 100 mL Nuclear Medium B and gently pelleted by centrifugation at 480 g for 10 minutes. Nuclei were used immediately for basic nuclear protein extraction. On no occasion were isolated nuclei stored.

4.3 Basic Nuclear Protein Extraction

Several methods of basic nuclear protein extraction from *O.luteus* nuclei, as isolated by the method described in section 4.2.3., were attempted and compared.

4.3.1 2M NaCl Extraction

This extraction method is based on that of Ohlenbusch *et al.*, 1967. Essentially, to total nuclear isolate, 5 mL extraction buffer consisting of 15 mM Tris-HCl, ph 7,3, 5 mM EDTA, 1 mM PMSF and 2 M NaCl was added. The nuclei were homogenized by subjecting the suspension to a Polytron homogeniser for 1 minute. Large, heavy insoluble material was removed by centrifuging the sample in a desk top centrifuge for 10 minutes. The supernatant underwent further centrifugation in a type SW 65L ti Beckman rotor at 45 000 rpm for 16 hours at $4^{o}C$ in order to pellet the DNA. The resulting supernatant was dialysed for 12 hours against two change of 5 L 0,1 mM PMSF in glass distilled water prior to lypholisation.

4.3.2 Hydrochloric Acid Extraction

Extraction of basic nuclear proteins by employing hydrochloric acid was attempted (Kossel, 1884; Johns, 1964). To total nuclei isolated from 5 L culture, 10 mL of 0,25 M HCl was added. The pellet was resuspended by 5 strokes of a loose fitting dounce homogeniser and the basic proteins allowed to extract from the nuclei overnight at $4^{o}C$. Following extraction, the suspension was centrifuged in a Sorvall SM 24 rotor at 10 000 rpm for 10 minutes. The basic nuclear protein containing supernatant was dialysed against 5 L 0,2 mM PMSF in twice glass distilled water for 12 hours. The basic proteins were then lypholised and stored at $-20^{o}C$.

4.3.3 Sulphuric Acid Extraction

The H_2SO_4 procedure for extraction of basic histone proteins has been described by Fambrough and Bonner (1966). The same protocol was employed here in an attempt to extract basic proteins. 2 mL of 0,4 M H_2SO_4 was added to washed and pelleted *O.luteus* nuclei. After resuspension of the nuclei by using a loose fitting dounce homogeniser, the basic proteins were allowed to extract on ice for 60 - 90 minutes. The histone depleted nuclei were pelleted by centrifugation in a desk top centrifuge set at full speed for 10 minutes. The supernatant was carefully decanted and mixed with 8 - 10 mL ice cold acetone (analytical grade). The basic protein precipitate was allowed to form overnight at -20^{o}C. The precipitate was pelleted by centrifugation at 20 000 g in a fixed angle rotor for 10 minutes, washed in 10 mL ice cold acetone before being pelleted again. The precipitation was then dried in a desiccator and stored at -20^{o}C.

4.3.4 Protamine Displacement of Basic Nuclear Proteins.

Basic nuclear proteins were isolated by a modification of the method of van der Westhuizen and von Holt (1971). Washed nuclei, isolated from 50 L of culture, were dissolved in 10 mM Tris-HCl, ph 7,4, 2 M NaCl, 0,2 mM PMSF to a final DNA concentration of 4 mg/mL. The solution was allowed to equilibriate on ice for 30 - 60 minutes. An equal volume of the same buffer but which contained protamine (Merk), at a concentration of 20 mg/mL, was added. The two solutions were mixed and dialysed against 10 mM Tris-HCl, ph 7,4; 0,15 M NaCl; 0,2 mM PMSF at 4^{o}C for 16 hours with one change of buffer after 6 - 8 hours. The resulting precipitated deoxyribonucleoprotamine complex was pelleted by centrifugation at 10 000 g for 20 minutes. The basic proteins and excess protamine containing supernatant was concentrated to 3 - 5 mL by ultrafiltration through a PM-10 membrane. Subsequent fractionation of protamine and extracted protein was achieved by gel filtration over a column of Sephadex G-50 (3 X 90 cm) equilibriated and eluted with 10 mM Tris-HCl, ph 7,4; 0,15 M NaCl at 4^{o}C. Eluted fractions were monitored by absorbance at 230 nm. Protein containing fractions were pooled and dialysed

against 0,1 mM PMSF at 4°C overnight prior to lypholysation. Dried fractions were stored at -20°C until needed.

4.4 Analysis of Extracted Nuclear Proteins by SDS-PAGE

Extracted *O.luteus* nuclear proteins were characterized by SDS-PAGE according to the method of Laemmli (1970) with some modifications.

4.4.1 Solutions for SDS-PAGE and Gel Casting

The following stock solutions were prepared:
Solution A: Running gel buffer containing
1,125 M Tris-HCl, ph 8,8; 0,3 % (w/v) SDS.
Solution B: Stacking gel buffer containing
0,375 M Tris-HCl, ph 6,8; 0,3 % (w/v) SDS.
Solution C: 10 X concentrated Electrode Buffer containing
0,25 M Tris-HCl, ph 8.3; 1 % (w/v) SDS; 1,92 M glycine.
Solution D: Acrylamide solution containing
30 % (w/v) acrylamide and 0,8 % (w/v) bis-acrylamide.
Solution E: Sample Buffer containing
0,0625 M Tris-HCl, ph 6,8; 2 % (w/v) SDS; 10 % (w/v) glycerol; 5 % (v/v) mercaptoethanol and 0,005 % bromophenol blue.
Solution F: 10 % ammonium persulphate.

All stock solutions were prepared in twice glass distilled water. If protein electroelution was to be performed, all solutions were prepared using analytical grade chemicals.

Slab SDS gels (25 X 12.5 X 0,5 cm) were cast as follows:
Running gel: 15 mL Solution D, 10 mL Solution A, 4,7 mL H_2O, 25 μL TEMED and 0,4 mL Solution F.
Stacking gel: 2 mL Solution D, 5 mL Solution B, 5,7 mL H_2O, 20 μL TEMED, 0,3 mL Solution F.

Gel solutions were briefly degassed prior to the addition of the TEMED. Polymerisation was allowed to occur for about 1 hour.

4.4.2 Sample Preparation

Samples, in plastic Eppendorf vials, were dissolved to the required concentration with sample buffer. After being heated to 100^{o}C for 2 minutes, samples were rapidly cooled on ice, microfuged and loaded onto the gel. Generally 5 - 20 μL volumes were loaded.

4.4.3 Electrophoresis

Electrophoresis was allowed to occur at 95 - 100 V for 16 hours

4.4.4 Staining and Destaining of gels

Gels were stained for about 1 hour with 0,25 % (w/v) Coomassie Brilliant Blue (Merk) in 9 % (v/v) glacial acetic acid, 45 % technical methanol and water, whilst under continuous agitation. Destaining was achieved by placing the gel in 0,9 M acetic acid and 40 % methanol in water, covering the gel with absorbent tissue paper to absorb stain, and agitated until the background was clear. Protein bands, being stained prior to electroelution were stained and destained in solutions prepared with analytical grade reagents only.

4.5 Electroelution of Electrophoresed Proteins

Stained protein bands were electroeluted as follows. The gels were allowed to destain until the protein bands were only faintly visible. The required bands were excised using a scalpel blade - with care taken to avoid contamination with

proteases. The excised gel strip was soaked in a solution of 0,9 M acetic acid and 0,25 M CTAB. The soaked gels were placed in the sample chamber of a Schleicher and Schull electroeluter and covered with the same buffer. The whole unit was placed in a flat bed electrophoresis tank and filled with 0,9 M acetic acid. Electroelution was performed at 100 V for 4 - 6 hours. The eluted protein was recovered from the protein trap and lypholised. Excess stain and CTAB was removed by resuspending the protein in 100 µL methanol and 1 mL Acetone. The precipitated proteins were recovered by centrifugation, lypholised and stored at -20^{o}C.

4.6 Protease Inhibitor Assay

Early on in this study it became obvious that the isolated proteins were being degraded, presumably by a proteolytic enzyme endogenous to *O.luteus*. Several common protease inhibitors were assayed to determine their effectiveness against this proteolysis. Sea urchin sperm histones were used as the substrate, while *O.luteus* cell lysate was used as the proteolytic factor.

4.6.1 Protease Inhibitor Stock Solutions

Protease inhibitor stock solutions were made up as follows. The serine protease inhibitors BZA and PMSF (Fahrney and Gold, 1963) were dissolved in DMSO to a final concentration of 100 mM. TLCK (Shaw *et al.*, 1965) and TPCK (Schoellman and Shaw, 1963), both serine and thiol protease inhibitors were made up to 1 M solutions with DMSO. The metaloprotease inhibitors EGTA and EDTA were each dissolved in H_2O to a final concentration of 250 mM whilst the ph was constantly adjusted to ph 7,3 with 1 M NaOH. pCMB, also shown to be effective as a thio protease inhibitor (Hellerman and Perkins, 1934; Finkle and Smith, 1958), was prepared by dissolving in 0,04 M NaOH to a final concentration of 1 mM. A 100 mM stock solution of leupeptin was prepared in 50 mM Tris-HCl, ph 7,2. Leupeptin has been shown to act as a thiol protease inhibitor (Umezawa, 1976). The acid protease inhibitor pepstatin (Umezawa, 1976) was dissolved in DMSO to a final concentration of 100 mM. Sodium bisulphate and sodium molybdate (Hazato and Murayama, 1981) were both made up to 10 mM final concentration with twice glass distilled water.

4.6.2 The Protease Inhibitor Assay

200 mL of *O.luteus* culture (3 X 10^5 cells/mL) was pelleted in a desk top centrifuge for 10 minutes. The pellet was resuspended in 5 mL 50 mM Tris-HCl, ph 7,6, and the cells then ruptured by 15 strokes of a tight-fitting dounce homogeniser. The resulting lysate was stored at 4^oC until required.

Sea urchin sperm histones (a gift from S. Schwager) were dissolved in twice distilled water to a final concentration of 10 mg/mL.

18 Eppendorf vials were prepared as follows (refer to Table 4.1):
Tubes 2 - 18 all contained 20 μL of the sea urchin histone solution. Tubes 4 - 18 also contained 80 μL cell lysate. Tube 6 to 16 each contained a different protease inhibitor. Tube 6 contained 0,4 μL PMSF; tube 7, 0,4 μL BZA; tube 8, 4 μL TPCK (stock solution diluted 1:100); tube 9, 4 μL TLCK (stock solution diluted 1:100); tube 10, 20 μL pCMB; tube 11, 0,4 μL leupeptin; tube 12, 0,4 μL pepstatin, tube 13,4 μL sodium molybdate; tube 14, 1,6 μL EDTA; tube 15 μL EGTA and tube 16, 4 μL sodium bisulphate. Tubes 1 and 5 contained all inhibitors in amounts equivalent to those aliquoted in tubes 6 - 16. Tubes 1, 3 - 16 were made up to 200 μL final volume with 10 mM Tris-HCl, ph 7,6, resulting in a final inhibitor concentration of 0,2 mM except for pCMB (final concentration of 10^{-3} mM) due to its low solubility in H_2O. Tube 17 contained 100 mL 10 mM sodium citrate, ph 5,5, and tube 18 contained 100 μL 10 mM Tris-HCl, ph 9. Tubes 1 and 2 were frozen, while all other tubes were incubated at 37^{o}C for 16 hours.

After 16 hours, all tubes were brought to 4^{o}C and microfuged for 15 minutes. The supernatants were carefully aspirated off and lypholised. The dried material was then dissolved in 20 μL SDS-PAGE sample buffer and analysed by SDS-PAGE as described in section 4.4.

Tube	Substrate (μL)	Tris-HCl 10 mM pH 7.6 (μL)	Lysate (μL)	Inhibitor	Inhibitor amount from stock (μL)[1]	Treatment
1	-	98,8	80	all	*	frozen
2	20	-	-	-	-	frozen
3	20	180,0	-	-	-	37° C
4	20	100,0	-	-	-	37° C
5	20	78,8	80	all	*	37° C
6	20	99,6	80	PMSF	0,4	37° C
7	20	99,6	80	BZA	0,4	37° C
8	20	96,0	80	TPCK	4,0 (1:100)	37° C
9	20	96,0	80	TLCK	4,0 (1:100)	37° C
10	20	80,0	80	pCMB	20,0	37° C
11	20	99,6	80	leupeptin	0,4	37° C
12	20	99,6	80	pepstatin	0,4	37° C
13	20	96,0	80	Na_2MoO_4	4,0	37° C
14	20	98,4	80	EDTA	1,6	37° C
15	20	98,4	80	EGTA	1,6	37° C
16	20	96,0	80	$NaHSO_4$	4,0	37° C
17	20	100,0[2]	80	-	-	37° C
18	20	100,0[3]	80	-	-	37° C

Substrate: 10 mg/mL Total Sea urchin sperm histones
* Same quantities as for tubes 6 - 16
1 Final concentration 0,2 mM except for pCMB (final concentration: 10^{-3} mM)
2 10 mM sodium citrate pH 5,5
3 10 mM Tris-HCl pH 9,0

TABLE 4.1: Summary of Protease Inhibitor Assay

4.7 Fractionation of extracted *Olisthodiscus luteus* Nuclear Proteins

4.7.1 Molecular Exclusion Chromatography in the presence of Sodium Chloride

The extracted nuclear proteins were fractionated by molecular exclusion chromatography over a column of Bio-Gel P-60 as described by Böhm *et al.* (1973), and von Holt and Brandt (1977). Samples were dissolved in 50 mM NaCl, 6 M urea and 1 % (v/v) mercaptoethanol and left at room temperature for 1 - 2 hours to allow complete denaturation to take place. The sample was then loaded onto the Bio-Gel P-60 (M) column (3 X 100 cm) at $4^{o}C$. The protein fractions were eluted with 0.02 N HCl, 50 mM NaCl and 0,1 mM PMSF. Typically, one run consisted of 200 fractions of 1,2 mL/fraction. Eluted fractions were monitored for protein content by measuring absorbance at 230 nm. The fractions corresponding to a protein containing peak were combined and dialysed against 0,1 mM PMSF at $4^{o}C$ for 16 hours before being lypholised. Dried samples were then analysed by SDS-PAGE prior to further purification by high performance liquid chromatography.

4.7.2 Reverse-Phase High Performance Liquid Chromatography (HPLC)

Final *O.luteus* nuclear basic nuclear protein purification was achieved by HPLC. The same system has been utilised in this study for the fractionation of peptides generated from enzymatic or chemical cleavages of isolated proteins. For both procedures, an acetonitrile-water mixture was used as the solvent system (Ryeszotarski and Mauger, 1973).

Molnar and Horvath (1977) illustrated the importance of ion-pairing of an ionic solute and the resulting effects on chromatographic behavior. In this study, it was found that heptafluorobutyric acid (HFBA) (Bennet *et al.*, 1980), used as the ion-pairing agent in an acetonitrile-water mixture, suitable for the fractionation of proteins extracted from the algal nucleus. For the purification of peptides generated by fragmentation of these proteins, trifluoroacetic acid (TFA) (Bennet *et al.*, 1980) was found to be, on the whole, superior.

The fractionation and purification of nuclear extracted proteins by HPLC was found to be so efficient, that partial purification by Bio-Gel P-60 was later deemed unnecessary.

The use of a perfluorinated carboxylic acid-acetonitrile-water solvent system for fractionation of peptides and proteins has many advantages. The solvents are easily and reproducibly prepared and stored. As a solvent system, they act as good solubilising agents for hydrophobic peptides. At the concentrations used (≈ 0.1 % HFBA or TFA and 70 % acetonitrile) the mixture exhibits minimal absorption at 230 nm, allowing accurate detection of protein or peptide peaks by monitoring at this wavelength. As all the buffer constituents are volatile, pure dried proteins can be achieved by simply lypholysing the eluted 230 nm absorbing peaks.

4.7.2.1 High Performance Liquid Chromatography Apparatus

A Waters Associates High Performance Liquid Chromatography System was used for all HPLC purification procedures. The system included two M600 A solvent delivery pumps, a M660 solvent programmer, a U6K Manual Sample Injector and a M440 Fixed Wavelength Ultra-Violet absorbance detector set for detection at 229 nm.

4.7.2.2 HPLC Column

A home packed HPLC column was used for purification of both proteins and peptides. The column (4 X 250 mm) was packed with analytical Vidac protein C_4 (10 μm) fully endcapped phase. The column was stored in 50 % (v/v) Acetonitrile when not in use.

4.7.2.3 Solvent Preparation

Trifluoroacetic acid and Heptafluorobutyric acid were purified by refluxing over CrO_3, distilled, passed over neutral aluminium oxide and redistilled. The acids were stored in sealed vessels and brought to room temperature prior to unsealing.

4.7.2.4 Solvent Gradient Programme

The same solvent gradient programme was used for both buffer system. After sample injection, buffer A was pumped isocratically for 10 minutes to allow elution of non-binding material such as urea and mercaptoethanol. Sample elution was achieved by generating a linear gradient (0-100%) of buffer B over a period of 60 minutes. Flow rate was maintained at 0,7 mL/min at all times. The column was prewashed utilizing the same gradient programme and solvents, ending in a 15 minute isocratic flush of solvent A. This served to equilibriate the column and elute any contaminants in the system.

4.7.2.5 Sample Preparation

Lypholized protein samples were dissolved in 6 M urea,
1 % Mercaptoethanol in water (200 µL - 500 µL) and allowed to denature for 1 hour at room temperature. After acidifying with HFBA (0,1 % v/v) the sample was clarified by centrifugation. Generally, the total volume was loaded on to the column.
For enzyme generated peptides, the total (generally 500 µL) enzyme reaction was loaded after being acidified with TFA (0,1 % final concentration) and clarified by centrifugation. In the case of chemical cleavage reactions, samples were first lypholised and then redissolved in 0,1 % (v/v) TFA (200 µL), centrifuged and loaded onto the column. Eluted samples, collected manually, were placed under a stream of nitrogen to remove the volatile acetonitrile and HFBA / TFA. Samples were then lypholised and stored at -20^{o}C until required.

4.8 Production of Peptides

4.8.1 Enzymatic Cleavages of Proteins

4.8.1.1 *Staphylococcus Aureus* V8 Protease Digestion

Staphylococcus aureus V8 protease cleaves peptide bonds specifically on the carboxyl side of glutamic acid or aspartic acid residues (Drapeau *et al.*, 1972) with optimum activity at pH 4,0 and pH 7,8. The enzyme specificity can be changed such that only glutamyl bonds are hydrolysed by selection of buffer composition. (Houmard and Drapeau, 1972). In ammonium acetate buffer (pH 4,0) only glutamyl bonds, with the exception of asp-gly bonds, are hydrolised.
For all protein digestion experiments, 0,5 mg of protein was dissolved in 200 µL 0,05 M ammonium acetate, ph 4,0. *Staphylococcus aureus* V8 protease (25 µg) was added to the sample and digestion allowed to proceed at 37^{o}C in a water bath for 16 - 18 hours. The sample was occasionally agitated.

If a large peptide could not be fully sequenced and if it was found to contain a single aspartic residue in the sequenced domain, cleavage at this residue was performed. The cleavage was achieved by dissolving the peptide in 400 µL 0,05 M K_2HPO_4, pH 7,0, and 20 µg *Staphylococcus aureus* V8 protease (Houmard and Drapeau, 1972). Digestion of the peptide was allowed to occur at 37^{o}C as described above. The protease generated peptides were fractionated by HPLC.

4.8.1.2 Endoproteinase Asp-N Protease Digestion

Endoproteinase ASP-N (Boehringer Mannheim), isolated from a *Pseudomonas fragi* mutant (Noveau & Drapeau, 1979) is a metallo protease which is claimed to cleave exclusively at the peptide bond N- terminally of aspartic and cysteic acid residues. The enzyme specificity is maintained in phosphate, acetate or Tris buffers at pH 6,0-8,5. (Drapeau, 1980).
Generally 0,5 mg protein was dissolved in 200 µL 10 mM

Tris-HCl pH 7,5. 2 μq of endoproteinase Asp-N was added and incubated at 37^0C for 18 hours in a water bath with occasional agitation. The generated peptides were fractionated by HPLC.

4.8.1.3 Endoproteinase Arg-C

Endoproteinase Arg-C (Boehringer Mannheim) isolated from submaxillaris glands of mice, cleaves specifically at arginine residues (Levy *et al.*, 1970). Typically 0,5 mg of protein was dissolved in 10 mM Tris-HCl (pH 8) and 10 μg endoproteinase Arg-C added. Digestion was allowed to proceed at 37^0C in a water bath for 3,5 hours with occasional agitation. The resulting peptides were fractionated by HPLC.

4.8.1.4 Endoproteinase Lys-C

Endoproteinase Lys-C (Boehringer Mannheim) is a serine protease which specifically hydrolyses bonds at the C- terminal side of lysine residues. (Jekel *et al.*, 1983) 0,5 mg protein was dissolved in 200-500 μL 0,1 M NH_4HCO_3, pH 8,3 to which 5 μg protease was added and digestion allowed to occur for 12-16 hours at 37^0C in a water bath with occasional agitation. Fractionation of resulting peptides was achieved by HPLC.

4.8.2 Chemical Cleavage of Proteins

4.8.2.1 Cyanogen Bromide

Cyanogen Bromide (CnBr) cleaves peptide bonds at the C- terminal side of methionine residues (Gross and Witkop, 1961). The method followed here is similar to that described by Gross (1967). To 0,5 mg of protein dissolved in 500 μL 70 % (v/v) redistilled formic acid, 0,5 mg CnBr was added and dissolved by gentle agitation. The solution was kept under nitrogen for 6 hours. A further ± 0,2 mg

CnBr was added and cleavage allowed to take place for a further 12 hours. The volatile reagents were removed by placing under a stream of nitrogen for 20-30 minutes. Finally 3 volumes of distilled water was added prior to being lypholysed. The resulting peptides were fractionated by HPLC.

4.8.2.2 Heptafluorobutyric Acid

While investigating the stability of peptide bonds under sequencing conditions Brandt *et al.* (1982) observed that exposure of a protein to HFBA at 50-55^{0}C for a period of time resulted in two types of cleavages - at asp residues and at serine residues. This cleavage has been utilized to unblock N- terminal peptides in cases where the blocking group is assumed to be acetylated serine. The same method has also been employed to generate a limited number of peptides from intact protein. Protein or peptide was dissolved in 0,5 mL HFBA and the glass vial sealed under nitrogen. Cleavage was allowed to occur for 30 hours in a water bath maintained at 50^{0}C. After cleavage, the HFBA was evaporated by placing under a gentle stream of nitrogen. The resulting peptides were separated by HPLC and subjected to gas phase sequencing.

4.9 Peptide nomenclature

For ease of discussion, a peptide nomenclature system was devised which describes the peptides parent protein, the type of cleavage that produced it and finally, the number of the peak (numbering from early to late eluting) of the chromatogram resulting from peptide fractionation. The abbreviation of proteolytic methods employed is as follows:

V8 = *Staphylococcus aureus* V8 protease raise with digestion carried out at pH 4.
V8H = *Staphylococcus aureus* V8 protease digesting substrate at pH 7,0.
ED = Endoproteinase Aspartic -N (cleaving N- terminally to aspartic acid).
ER = Endoproteinase arginine (cleaving C- terminal side of arginine).
Ek = Endoproteinase lysine (cleaves C- terminal side of lysine residue).

For peptides produced by chemical cleavage methods, the following shorthand applies:

Cn = Cyanogen Bromide cleavage (as described in section 4.8.2.1).

HF = Heptafluorobutyric acid cleavage as described in (section 4.8.2.2).

A peptide labelled D.ER.6 therefore refers to the 6th eluting peptide when protein D is fractionated by HPLC after being subjected to endoproteinase - arginine digestion.

4.10 Sequence Analysis

Sequence analysis by manual gas-phase Edman degradation (Edman, 1950; Brandt and Frank, 1988) or by utilizing an automated gas-phase sequencer (Brandt *et al.*, 1984) was performed. For both methods, reagent and solvent purification as well as sample preparation, remained the same.

4.10.1 Sequence Reagents and Solvents

Phenylisothiocyanate	(Merck)
Heptafluorobutyric acid	(3M)
Trifluoroacetic acid	(Sigma)
Heptane	(Merck)
1 - Chlorobutane	(Merck)
Benzene	(Merck)
Trimethylamine	(Merck)
Ethylacetate	(Merck)
Trimethylamine	(Merck)
Methanol	(Carlo Erba)
Polybrene	(Pierce)

Most reagents (sequencing grade) were found to be of sufficient purity and was not further purified. Similarly the solvents (analytical grade) required no further treatment. However, 1 - chlorobutane, benzene, trifluoroacetic acid and heptafluorobutyric acid were further purified by the method of Hunkapiller *et al.* (1983).

Trimethylamine was purified according to (Brandt *et al.*, 1984). Using an all glass distillation apparatus, trimethylamine was prepared from the hydrochloride by the slow addition of a concentrated NaOH solution. The condenser and the receiving bottle were cooled to $-25^{0}C$. Following this the collected

trimethylamine was refluxed over phthalic anhydride at room temperature and redistilled. Finally, the TMA was diluted to 25% (v/v) with distilled H_2O.

4.10.2 Sample Preparation

Protein or peptide sample's to be sequenced either by manual gas-phase or automated gas phase sequencing protocols, were prepared in the same manner - by spotting onto a glass fibre disc prepared with Polybrene (Brandt and Frank, 1988).

4.10.2.1 Preparation of Glass Filter Disc's and Application of Protein Carrier

Glass filter discs (Whatman GF/C) were punched from the larger discs using a cutting tool of 0,8 centimeter diameter. Approximately 50 discs were prepared at one time. The cut filters were soaked in 30-40 mL neat TFA for at least 2 hours so as to remove any contaminants.

Excess TFA was removed by washing filters four times with 30-50 mL 1-chlorobutane. The discs were dried in vacuo before being positioned into glass manual sequencing glass funnels (1 cm diameter, 2 cm length tapering to 0,2-0,4 cm diameter) or into Eppendorf vial (1,5 μL) "funnels". Eppendorf funnels were constructed by simply tearing off the caps and nipping off the sealed bottom. At this stage Polybrene was applied to the disc. Polybrene (Pearce) is the synthetic polymer hexadimethrine bromide formed from N,N,N',N',-tetramethylhexamethylenediamine and trimethylene bromide. Using Polybrene as a protein carrier, less than 20 nmoles of protein or peptide has been successfully sequenced in the spinning cup sequenator (Tarr *et al.*, 1978; Klapper *et al.*, 1978). Due to its hydrophobic as well as hydrophilic domains Polybrene acts as an anchor for both hydrophobic and hydrophilic peptides.

Due to Polybrene's hydroscopic, nature a substantial amount of water may be retained on the disc during the course of the Edmans degradation protocol. This retained water, with TMA form hydroxyl ions which react with the PITC resulting in

the formation of aniline. Aniline in turn reacts with PITC to form DPTU - which is partially retained by the Polybrene. The DPTU is extracted when the PTH-amino acid and can frustrate positive identification of PTH-amino acid. For this reason, only 0,25-1,0 mg polybrene in 10 µL of H_2O is applied to the disc. After the Polybrene has dried the glass filter discs were subjected to a "cleaning cycle" before use. The discs, placed in the sequencing vials, were exposed to TFA vapours in a desiccator for 15 minutes. After briefly drying the discs under vacuum for 2 minutes, the discs were washed with 3 x 200 µL ethyl acetate and 2 x 200 µL 1 - chlorobutane. Finally the discs were dried under vacuum and stored until needed.

4.10.2.2 Sample Application

100 pmole - 5 nmole of peptide or protein to be sequenced was pipetted evenly over a prepared disc. The applied sample was then dried under a vacuum for 10-15 minutes. To reduce contaminating amino acids that may be present on the disc or in the sample, the loaded disc was subjected to a cleavage cycle described under section 4.10.3.2. As the automatic gas phase sequencer starts its program with a TFA cleavage cycle, prior to PITC coupling, discs destined for the sequencer were loaded without performing a manual TFA cleavage cycle.

4.10.3 Manual Gas Phase Sequencing

Manual gas phase sequencing was performed by a method developed in this laboratory and described by Brandt and Frank (1988). The method is useful in that N-terminal amino acids of up to twelve samples can be determined simultaneously. Further more the disc size employed in the manual sequencing protocol is identical as to that of the automatic gas phase sequencer. This allows one to assay several peptides N- terminal amino acids and then transfer a selected peptide or protein to the automatic gas phase sequencer for further analysis. Three distinct stages are performed during the manual sequencing protocol i.e. coupling, cleavage and conversion.

4.10.3.1 Coupling

15 µL of PITC solution (5% (w/v) in heptane) was carefully applied evenly over the disc surface. The glass columns containing the discs was placed in a small desiccator and purged with nitrogen. A 1,5 mL Eppendorf tube containing 1 mL of TMA (10% (v/v) TMA; 0,1 mg/1mL ninhydrin in H_2O) was introduced into the desiccator and stabilized. A slight vacuum was pulled before sealing the desiccator in order to accelerate the TMA evaporation. The entire desiccator and contents was incubated at 50^0C for 20-40 minutes - usually by submerging into a 50^0C water bath. After incubation, the Eppendorf containing the base was removed and the volatiles removed by inducing a vacuum for 5 minutes. The filters were washed by placing the funnels over test tubes and sequentially passing 2 x 200 µL heptane, 2 x 200 µL ethyl acetate and 1 x 200 µL 1-chlorobutane through the filters by pipetting. The sequencing discs were then briefly dried (5 minutes) under vacuum. It has been observed that a longer drying period results in a less efficient cleavage cycle.

4.10.3.2 Cleavage

The columns containing the sequencing discs were placed in a desiccator similar to that of above. An Eppendorf vial containing 0,5 mL TFA was introduced. After purging the desiccator with N_2, a vacuum was introduced for 1 minute. As for coupling incubation, the desiccator was incubated at 50^0C for 15 minutes. After this period, the tube containing TFA was removed and the desiccator evacuated for 2 minutes. The ATZ derivatives were extracted from the filter discs by placing the glass columns over 5 mL glass test-tubes and eluting the derivatives with 2 x 200 µL 1-chlorobutane. The filter discs were dried down under vacuum and either stored, used for another cycle of manual sequencing or loaded onto the automatic gas phase sequencer. The ATZ derivatives were dried down in preparation for conversion by either placing in a desiccator under vacuum or by placing under a stream of nitrogen.

4.10.3.3 Conversion

The ATZ derivatives were converted to PTH-amino acids by adding 50 µL TFA (20% (v/v) in H_2O) to each tube, purging with N_2, stopping the tube and incubating at 80^0C for
10 - 15 minutes with the aid of a heating block. After evaporating the TFA / H_2O in vacuo, the PTH-amino acids were dissolved in a suitable solvent and identified.

4.10.4 Automated Gas-Phase Sequencing

This laboratory designed and build a custom designed gas-liquid solid phase sequencer as described below. All automated sequencing procedures were performed on this machine.

4.10.4.1 The Gas-Phase Sequencer

Due to the extremely high costs of acquiring a gas-liquid solid phase sequencer, this laboratory undertook to construct one. The construction of the sequencer was essentially as described by Hewick *et al.* (1981). However, a major modification to the Hewick *et al.* model may be found in the reagent and solvent delivery system. The valve system, also designed in this laboratory by Brandt *et al.* (1984), was devised to convert a model 890 Beckman spinning cup sequenator into a "gas-phase" sequencer. The advantages of this system being the ease and cheapness of manufacture and replacement, the absence of dead pockets and the small dead volumes that occur. The schematic flow diagram of the sequencer is shown in figure 4.2.

4.10.4.2 The Sequencer Programme

The reagents and solvents used is given in Table 4.2. The complete sequencing program controlled by a dedicated BBC-Computer, is shown in Table 4.3. The temperature of the sequencing disc cartridge was set at $44^{0}C$ whilst the conversion flask was set at $50^{0}C$. Two different nitrogen pressures where used for transfer of reagents and solvents, at a low pressure of 5 k Pa and a high pressure of 15 k Pa.

A high pressure nitrogen line of 15 - 20 k Pa was used for drying down samples in the conversion flask. Each Edman degradation cycle starts with a cleavage step, followed by a coupling step. The cleavage cycle triggers the conversion step of the previous extracted ATZ-derivatives. Initially PTH-amino acid were analysed

independently as described in section 4.10.5.1 prior to the development of an "on-line" system.

Reagents and Solvents used in the Gas-Phase Sequencer	
Reagent / Solvents	Volume / cycle
R1: 5 % (v/v) phenylisothiocyanate in n-heptane	72 μL
R2: 25 % (v/v) trimethylamine in water	37,1 cc*
R3: trifluoroacetic acid	87,9 cc*
R4: 25 % trifluoroacetic acid	700 μL
R5: 1 nm norleucine in ethyl acetate	93 μL
S1: Benzene	1525 μL
S2: 0.001 % (w/v) dithioerythritol in Ethyl Acetate	1296 μL
S3: 1-chlorobutane	1400 μL
S4: ethyl acetate	347 μL
N1 & N6: low pressure nitrogen	
N2 & N4: high pressure nitrogen	

* As measure by a Gilmont flowmeter

TABLE 4.2 Solvents and reagents used in the Gas-Phase Protein Sequencer

Cleavage and Coupling Loop of Sequencing Program				
Step	Solvent/ Reagent	Event	Time(s)	Volume
1	N1	Low N_2 pressure in all lines	4	
2	R3	TFA gas delivered to cartridge	999	73 cc
3		Conversion loop triggered	1	
4	R3	TFA gas delivered to cartridge	200	14 cc
5	N2	N_2 flushes lines and cartridge	60	
6	S3	1-chlorobutane dissolves and delivers ATZ derivatives to conversion flask	10	100 µL
7		Pause	30	
8	S3	As for step 6	30	300 µL
9		Pause	30	
10	S3	As for step 6	30	300 µL
11		Pause	30	
12	S3	As for step 6	30	300 µL
13	N2	N_2 flushed thru cartridge and flask	180	
14	R2	TMA vapours flush thru cartridge	100	2 cc
15	R1	PITC container pressurized	10	
16		Waste open	5	
17	R1	PITC delivered to Cartridge	2	24 µL
18	N1	Low pressure N_2 flush thru cartridge	10	
19	N2	High pressure N_2 flush	85	
20	R2	TMA vapour flushed thru cartridge	500	12 cc
21		waste open		
22	R1	PITC container pressurized	5	
23	R1	As for step 17	2	24 µL
24	N1	As for step 18	10	
25	N2	As for step 19	60	
26	R2	As for step 20	500	12 cc
27		Waste open	5	
28	R1	As for step 17, but waste closed	10	24 µL
29	N1	As for step 18	10	
30	N2	As for step 19	85	
31	R2	As for step 20	500	12 cc
32	N2	High pressure N_2 clears R1 to waste	100	
33	S1	Benzene flows thru disc to waste	8	98 µL
34		Pause	30	
35	S1	As for step 33	60	714 µL
36		Pause	30	
37	S1	As for step 33	60	714 µL
38		Pause	40	
39	S2	Ethyl acetate thru disc to waste	140	1008 µL
40		Pause	30	
41	S2	As for step 39	40	288 µL
42	S3	1-Chlorobutane thru disc to waste	40	400 µL
43	N1	Low pressure N_2 flush to dry disc	20	
44	N2	High pressure N2 flush to dry disc	250	
		END		

Steps 1 - 5 results in cleavage of the ATZ derivative.
Steps 6 - 13 transfers derivative to conversion chamber.
Steps 14 - 32 results in the coupling of PITC to the peptide in the presence of the buffer TMA.
Steps 33 - 44 washes the disc and ends with a drying cycle.

TABLE 4.3 The Cleavage and Coupling Loop of the Sequencer program

The Conversion Loop of the Sequencing programme				
Step	Solvent/ Reagent	Event	Time(s)	Volume
1	N5	High pressure N_2 to dry conversion flask	10	
2	R5	Norleucine container pressurized	10	
3	R5	Norleucine delivered to flask	8	93 cc
4	N4	High pressure N_2 flush thru flask	10	
5	N4	As for step 4	175	
6	R4	25 % TFA transfer to flask	70	700 µL
7		Incubation at 50°C	420	
8	N4	High pressure N_2 expels TMA	250	
9	N4	As for step 4	300	
10	N4	As for step 4	100	
11	N4	As for step 4	50	
12		Pause	300	
13	N4	As for step 4	50	
14	N4	As for step 4	200	
15	N4	As for step 4	700	
16	S4	Ethyl acetate transferred to flask	17	77 µL
17	N4	High pressure N_2 clears lines	6	
18	N4	High pressure N_2 bubbles thru flask	10	
19		Pause	5	
20	N4	High pressure N_2 pressurizes flask	4	
21		Pause	10	
22		Rheodine valve activated	4	
23		Pause	10	
24	N4	High pressure N_2 expels solution to fraction collector*	150	
25	N4	High pressure N_2 clears line	30	
26		Fraction collector advanced#	1	
27	S4	As for step 16	17	77 µL
28	N4	As for step 18	4	
29	N4	As for step 18	10	
30	N4	High pressure N_2 expels solution to liquid waste	50	
31	N4	High pressure N_2 clears solvent lines	50	
32	N4	As for step 30	26	
33	N4	As for step 30	24	
34	S4	As for step 16	60	270 µL
35	N4	As for step 20	20	
36	N4	As for step 18	5	
37	N4	As for step 30	50	
38	N4	As for step 31	70	
39	N4	As for step 30	50	
40	N4	High pressure N^2 dries flask	100	
41	N4	High pressure N_2 dries lines	100	
42	N4	High pressure N_2 dries lines thru flask to fraction collector	550	
43	N4	As for step 42 END	200	

* For the "On Line" PTH-Amino Acid identification, transfer of the PTH-Amino Acid solution is straight to the HPLC Column

\# For the "On Line" system, this step would activate the HPLC program.

Steps 1 - 5 introduces the internal standard, norleucine, to the AZT derivative solution and dries down the sample.
Steps 6 - 15 allows conversion of the AZT derivatives to PTH-Amino Acid, followed by drying of the solution.
Steps 16 - 26 resuspends the PTH-Amino Acids and transfers the solution to the fraction collector.
Steps 27 - 43 washes and dries the conversion flask and lines in readiness for the following conversion cycle.

TABLE 4.3 continued. The Conversion Loop of the Gas-Phase sequencer program.

4.10.5 Identification and Quantitation of Phenylthiohydantion Amino Acid

During the early stages of this study phenylthiohydantion (PTH) amino acids were identified by reverse phase high performance liquid chromatography by manually loading samples onto a Hewlett-Packard 1084A high performance liquid chromatograph (section 4.10.5.1.). Later an "on line" system was acquired which is discussed in section 4.10.5.2.

4.10.5.1 "Off-Line" Identification and Quantitation of PTH-Amino Acids by Reverse Phase HPLC

PTH-amino acids were identified in a Hewlett-Packard 1084A high performance liquid chromatograph on a 5 μm particle size Spherisorb ODS2 reverse phase column at 55^{0}C. The ternary solvent system which achieved isocratic separation of the PTH-amino acid was that of Lottspeich (1980). The solvent mixture consisted of 68,5 % (v/v) 0,01 M sodium acetate in glass distilled water with acetic acid to ph 5,2, 31,5 % (v/v) acetonitrile and 0,5 % (v/v) dichloroethane. The solvent temperature was maintained at room temperature 23^{0}C.

Dried down PTH amino acid samples from the sequencer was resuspended in 100 μL eluting buffer. Generally, 20 μL was injected onto the column at a flow rate of 0,5 mL/min. The flow rate was increased to 1 mL/min over 20 minutes. The full time of the programme was 29 minutes. Elution of PTH-amino acids was monitored at 254 nm. Loading one fifth of the total sample volume resulted in the quantity of the internal standard phenylthiohydantion molecule being 1 nmol. The peak areas were integrated and quantified from the chromatogram by a dedicated computer. Peak identification was achieved by comparison of retention times of a standard PTH-amino acid mixture chromatographed under the same conditions.

4.10.5.2 "On-line" Identification and Quantitation of PTH-Amino Acids

By means of a slight modification of the sequencer program an "on-line" system of identifying and quantitating the PTH-amino acids was achieved. The PTH-amino acids, dissolved in elution buffer, was pumped directly from the conversion flask to the HPLC sample loading loop. This system proved to be highly reproducible as well as time saving. The same principle of PTH-AA separation by reverse phase HPLC used in the "off-line" system (section 4.10.5.1.) was used here. It was found that the column (5 μm Spherisorb ODS2) at 55^{o}C and the buffer system of Lottspeich (1980) need not change under this new procedure.

Solvent flow through the HPLC column was achieved by a Shimadzu LC - 9A liquid chromatography pump set at a constant flow rate of 0,5 mL/min. Eluting PTH-AA were detected by a Waters 490E programmable multiwave detector (figure 4.3). The wavelengths used was 268 nm as well as 313 nm to aid identification of PTH-serine and PTH-threonine. Data collected was sent to a linked autochrom CS1 unit for storage and finally collected by an 4 channel APEX chromotography workstation. The workstation was operated from an AT-computer running an Autochrom Inc. APEX software (model M625-1) program which identified, integrated and quantitated the peaks of the PTH-AA elution chromatogram.

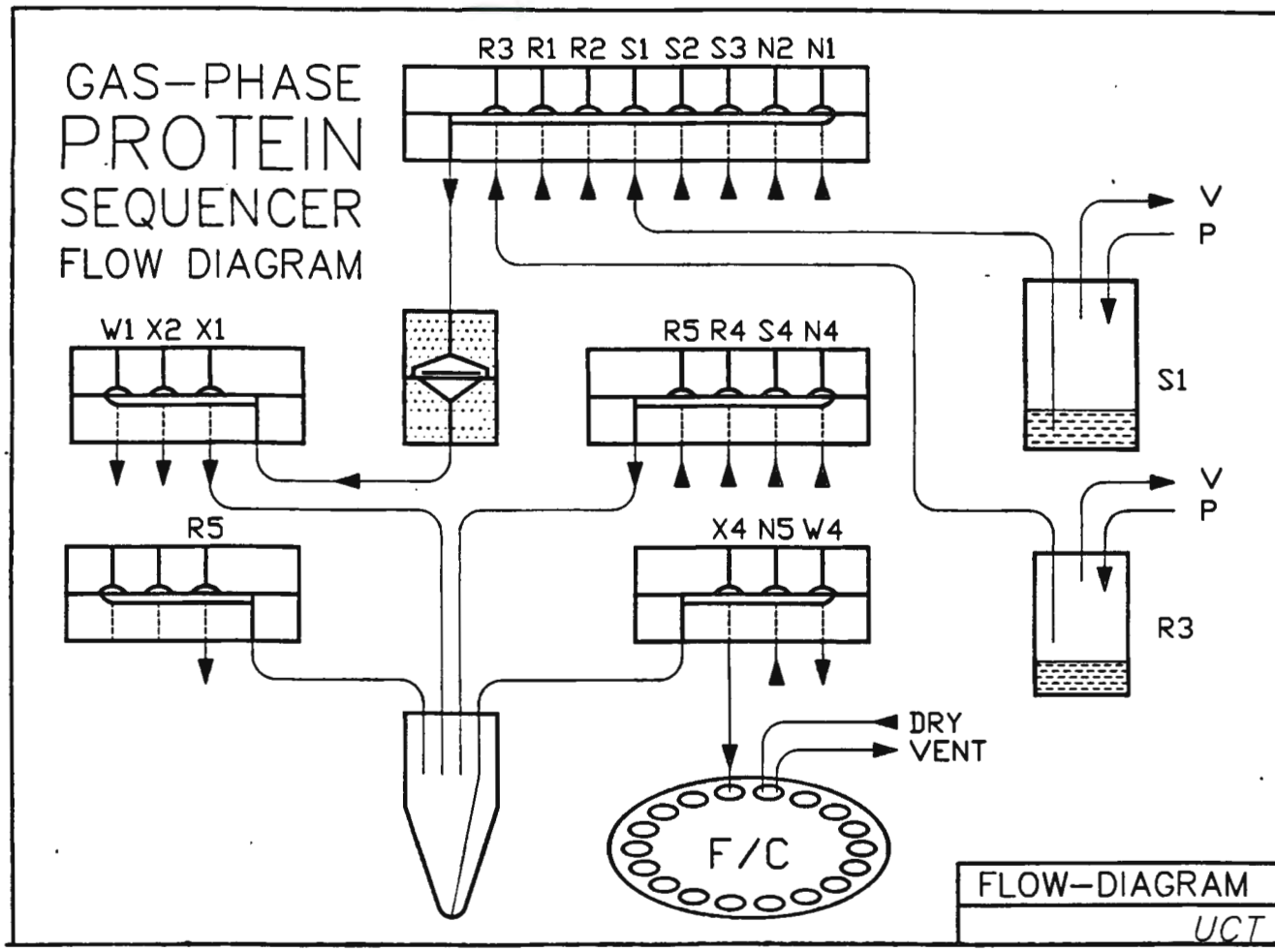

Figure 4.2: Schematic flow diagram of the home-made gas-phase sequencer. Valve labels refer to Table 4.3. V = vent; P = pressure; F/C = fraction collector.

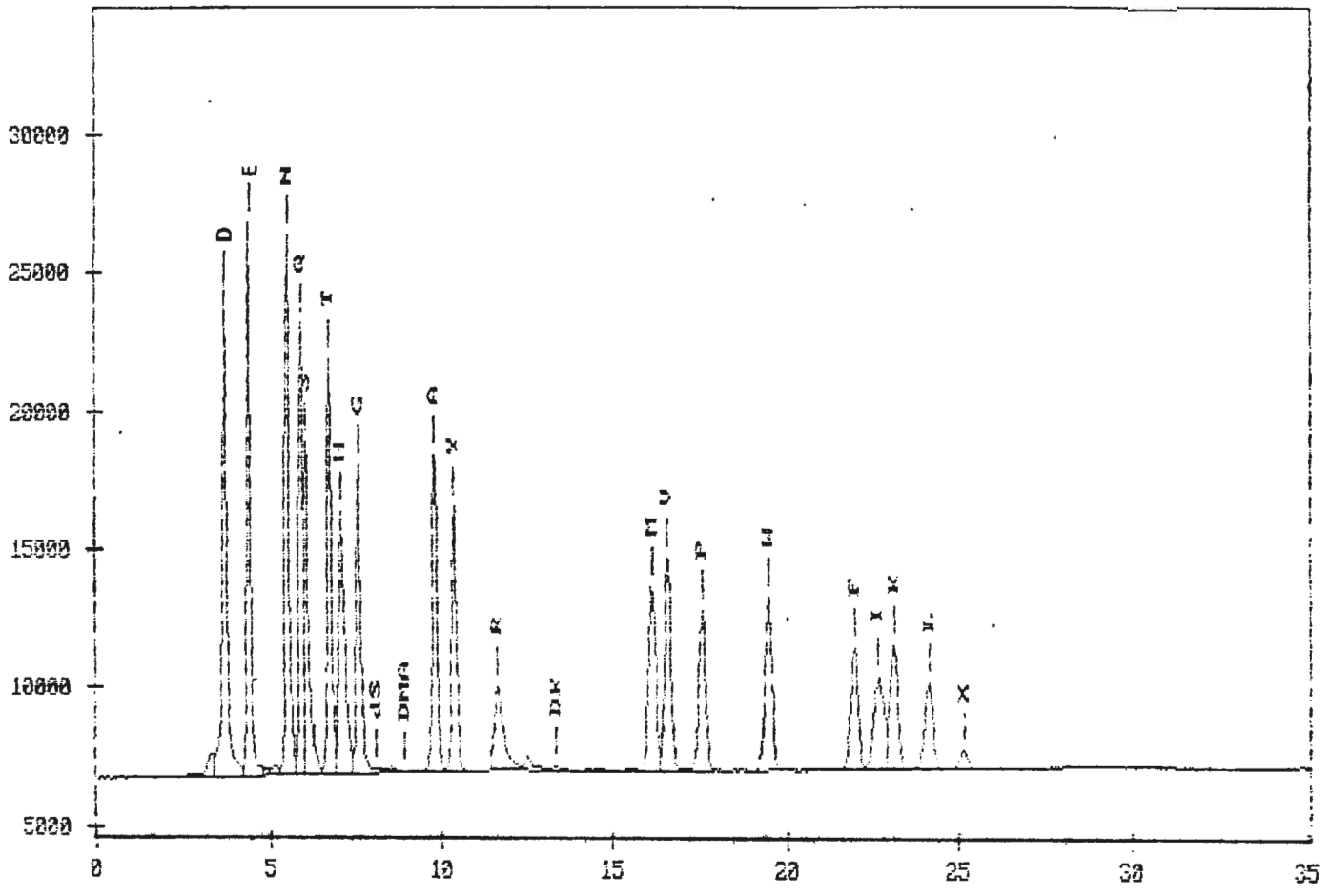

Figure 4.3: HPLC trace of PTH- amino acid standard (100 pmol each amino acid). The trace is from the "on-line" detection system (section 4.10.5.2) using a "Waters" 490 E programmable detector and an APEX 4 channel workstation. Detection wavelength is at 268 nm. Peak height measured in μV deflection and time in minutes.
HPLC column: 5 μm Spherisorb OD52 at 55^{o}C
Elution Buffer: 68,5 % (v/v) 0,01 M sodium acetate in glass distilled water with acetic acid to pH 5,2; 31,5 % (v/v) acetonitrile and 0,5 % (v/v) dichloroethane.
Linear flow programme: Sample was injected at a flow rate of 0,5 mL/min. Flow rate increased to 1 mL/min over 20 minutes. Run length was 20 minutes.
Internal standard (X): Norleucine

REFERENCES

Abercrombie, M., Hickman, C.J. and Johnson, M. (1980) *The Penguin Dictionary of Biology, 7th Edition*, Penguin Books, Harmondsworth, 14

Allan, J., Hartman, P.G., Crane-Robinson, C. and Aviles, F.X. (1980) *Nature (London)*, 288, 675 - 679

Allan, J., Mitchell, T., Harborne, N., Böhm, L. and Crane-Robinson, C. (1986) *J. Mol. Biol.*, 187, 591 - 601

Allfrey, V., Faulner, R.M. and Mirsky, A.E. (1964) *Proc. Natl. Acad. Sci. USA.*, 51, 786 - 794

Allis, C.D., Chicoine, L.G., Richman, R. and Schulman, I.G. (1985) *Proc. Natl. Acad. Sci. USA.*, 82, 8048 - 8052

Assali, N-E., Martin, W.F., Sommerville, C.C. and Loiseaux-de Goër, S. (1991) *Plant Molec. Biol.*, 17, 853 - 863.

Ausio, J., Dong, F. and van Holde, K.E. (1989) *J. Mol. Biol.*, 206, 451 - 463

Bannon, G.A., Bowen, J.K., Yao, M-C. and Gorovsky, M.A. (1984) *Nucleic Acids Res.*, 12, 1961 - 1975

Bavykin, S.G., Usachenko, S.I., Lishanskaya, A.I., Shick, V.V., Belyavsky, A.V., Undritsov, I.M., Strokov, A.A., Zalenskaya, I.A. and Mirzabekov, A.D. (1985) *Nucleic Acids Res.*, 13, 3439 - 3459

Bennet, H.P.J., Browne, C.A. and Solomon, S. (1980) *J. Liquid Chromatog.*, 3, 1353 - 1365

Boczar, B., Delaney, T. and Cattolico, R.A. (1989) *Proc. Natl. Acad. Sci. USA.*, 86, 4996 - 4999

Bogorad, L. (1975) *Nature (London)*, 188, 891 - 898

Böhm, E.L., Crane-Robinson, C. and Sautiére, P. (1980) *Eur. J. Biochem.*, 106, 525 - 530

Böhm, E.L., Hayashi, H., Cary, P.D., Moss, T., Crane-Robinson, C. and Bradbury, E.M. (1977) *Eur. J. Biochem.*, 77, 487 - 493

Böhm, E.L., Sautiére, P., Cary, P.D. and Meader, D. (1988) *Biochim. et Biophys. Acta.*, 956, 224 - 231

Böhm, E.L., Strickland, W.N., Strickland, M., Thwaits, B.H., van der Westhuizen, D.R. and von Holt, C. (1973) *FEBS lett.*, 34, 217 - 221

Bradbury, E.M. and Rattle, H.W. (1972) *Eur. J. Biochem.*, 27, 270 - 281

Brandt, W.F. and Frank, G. (1988) *Anal. Biochem.*, 168, 314 - 323

Brandt, W.F. and von Holt, C. (1982) *Eur. J. Biochem.*, 121, 501 - 510

Brandt, W.F., Henschen, A. and von Holt, C. (1980) *Hoppe-Seyler's Z. Physiol. Chem.*, 361, 943 - 952

Brandt, W.F., Henschen, A. and von Holt, C. (1982) In: *Methods in Protein Sequence Analysis* (Ed: Elzinga, M.), Humana Press, Clifton, New Jersey, 101 - 110

Brandt, W.F., Rodrigues, J. and von Holt, C. (1988) *Eur. J. Biochem.*, 173, 547 - 554

Brandt, W.F., Alk, H., Chauhan, M. and von Holt, C. (1984) *FEBS lett.*, 174, 228 - 232

Brown, W. (1990) *New Scientist*, 127, 30

Camerini-Otero, R.D. and Felsenfeld, G. (1977) *Proc. Natl. Acad. Sci. USA.*, 74, 5519 - 5523

Carter, N. (1937) *Arch. Protistenk.*, 90, 1 - 68

Cattolico, R.A. (1978) *Plant Phys.*, 62, 558 - 562

Cattolico, R.A., Boothroyd, J. and Gibbs, S. (1976) *Plant Phys.*, 57, 497 - 503

Certa, U. and von Ehrenstein, G. (1981) *Anal. Biochem.*, 118, 147 - 154

Choe, J., Kolodrubetz, P. and Grunstein, M. (1982) *Proc. Natl. Acad. Sci. USA.*, 79, 1484 - 1487

D'Anna, A.J. and Isenberg, I. (1974) *Biochemistry*, 13, 4992 - 4997

Daban, J.R. and Cantor, C.R. (1982) *J. Mol. Biol.*, 156, 771 - 789

De Lange, R.J., Fambrough, D.M., Smith, E.L. and Bonner, J. (1969) *J. Biol. Chem.*, 244, 319 - 334

De Lange, R.J., Hooper, J.A. and Smith, E.L. (1972) *Proc. Natl. Acad. Sci. USA.*, 69, 882 - 884

De Lange, R.J., Hooper, J.A. and Smith, E.L. (1973) *J. Biol. Chem.*, 248, 3261 - 3274

De Lange, R.J., Williams, L.C. and Martinson, H.G. (1979) *Biochemistry*, 18, 1942 - 1946

Dickerson, R.E. (1971) *J. Mol. Evol.*, 1, 26 - 45

Doolittle, R.F. (1979) In: *The Proteins* (Eds: Neurath, H. and Hill, R.L.) Academic Press, New York, 1979., 4, 30

Douglas, S.E. and Turner S. (1991) *J. Mol. Evol.*, 33, 267 - 273

Douglas,E.S., Murphy, C.A., Spencer, D.F., Gray, M.W. (1991) *Nature (London)*, 350, 148 - 151

Drapeau, G.R. (1980) *J. Biol. Chem.*, 255, 839 - 840

Drapeau, G.R., Boily, Y. and Houmard, J. (1972) *J. Biol. Chem.*, 247, 6720 - 6726

Ebralidse, K.K., Gracher, S.A. and Mirzabekov, A.D. (1988) *Nature (London)*, 331, 365 - 367

Econtre, I. and Parello, J. (1988) *J. Mol. Biol.*, 202, 673 - 676

Edman, P. (1950) *Acta. Chem. Scand.*, 4, 283 - 293

Edman, P. (1970) In: *Protein Sequence Determination* (Ed: Needleman, S.B.) Springer - Verlag, Berlin, 1970, 211 - 231

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

Edman, P. and Henschen, A.(1975) In: *Protein Sequence Determination* (Ed: Needleman, S.B.) Springer - Verlag, Berlin, 1970, 231 - 279

Eickbush, T.H. and Moudrianakis, E.N. (1978) *Biochemistry*, 17, 4955 - 4964

Elgin, S.C.R., Schilling, J. and Hood, L.E. (1979) *Biochemistry*, 18, 5679 - 5685

Elson, D., Gustafson,T., and Chargaff, E. (1954) *J. Biol. Chem.*, 209, 285

Ersland, D.R. and Cattolico, R.A. (1981) *Biochemistry*, 20, 6886 - 6893

Ersland, D.R., Aldrich, J. and Cattolico, R.A. (1981) *Plant Phys.*, 68, 1468 - 1473

Fahrney, D.E. and Gold, A.M. (1963) *J. Am. Chem. Soc.*, 85, 997

Fambrough, D.M. and Bonner, J. (1966) *Biochemistry*, 5, 2563 - 2569

Feng, D-F. and Doolittle, R.F. (1987) *J. Mol. Evol.*, 25, 351 - 360

Finch, J.T. and Klug, A. (1976) *Proc. Natl. Acad. Sci. USA.*, 73, 1897 - 1901

Finkle, B.J. and Smith, E.L. (1958) *J. Biol. Chem.*, 230, 669

Franklin, S.G. and Zweidler, A. (1977) *Nature (London)*, 266, 273 - 275

Fusauchi, Y. and Iwai, K. (1983) *J. Biochem.*, 93, 1487 - 1497

Genske, J.E., Cairns, B.R., Stack, S.P. and Landfear, S.M. (1990) *Mol, Cell. Biol.*, 11, 240 - 249

Gibbs, S.P. (1962) *J. Cell. Biol.*, 14, 433 - 444

Gibbs, S.P. (1981) *Ann. N.Y. Acad. Sci.*, 193 - 208

Gibbs, S.P., Chu, L.L. and Magnussen, C. (1980) *Phycologia*, 19, 173 - 177

Glover, C.V.C. and Gorovsky, M.A. (1979) *Proc. Natl. Acad. Sci. USA.*, 76, 585 - 589

Goodman, M. (1981) *Prog. Biophys. Molec. Biol.*, 37, 137 and 156

Gray, M.W., Sankoff, D. and Cedergren, R.J. (1984) *Nucleic Acids Res.*, 12, 5837 - 5852

Gribskov, M. and Burgess, R.R. (1986) *Nucleic Acids Res.*, 14, 6745 - 6763

Gross, E. (1967) *Methods in Enzymology*, 11, 238 - 255

Gross, E. and Witkop, B. (1961) *J. Am. Chem. Soc.*, 83, 1510 - 1511

Gunderson, J.H., Elwood, H., Ingold, A., Kindle, K. and Sogin, M.L. (1987) *Proc. Natl. Acad. Sci. USA.*, 84, 5823 - 5827

Hara, Y., Inouye, I. and Chihara, M. (1985) *Bot. Mag. Tokyo*, 98, 251 - 262

Hartl, D. and Dykhuizen, D. (1979) *Nature (London)*, 281, 230 - 231

Hayashi, H., Nomoto, M. and Iwai, K. (1984) *J. Biochem.*, 96, 1449 - 1456

Hayashi, T., Hayashi, H., Fusauchi, Y. and Iwai, K. (1984) *J. Biochem.*, 95, 1741 - 1749

Hayashi, T., Ohe, Y., Hayashi, H. and Iwai, K. (1982) *J. Biochem.*, 95, 1741 - 1749

Hazato, T. and Murayama, A. (1981) *Biochem. Biophys. Comm.*, 98, 488 - 493

Hellebust, J.A. (1965) *Limnol. Oceanogr.*, 10, 192 - 206

Hellerman, L. and Perkins, M.E. (1934) *J. Biol. Chem.*, 107, 241

Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) *J. Biol. Chem.*, 256, 7990 - 7997

Hewish, D.R. and Burgoyne, L.A. (1973) *Biochem. Biophys. Res. Commun.*, 52, 504 - 510

Hibberd, D.J. (1976) *Bot. J. Linn. Soc.*, 72, 55 - 80

Horgen, P.A. and Silver, J.C. (1978) *Ann. Rev. Microbiol.*, 32, 249 - 284

Horowitz, S. and Gorovsky, M.A. (1985) *Proc. Natl. Acad. Sci. USA.*, 82, 2452 - 2455

Houmard, J. and Drapeau, G.R. (1972) *Proc. Natl. Acad. Sci. USA.*, 69, 3506 - 3509

Hunkapiller, M.W. and Hood, L.E. (1978) *Biochemistry*, 17, 2124 - 2133

Hunkapiller, M.W., Hewick, R.M., Dreyer, W.J. and Hood, L.E. (1983) In: ***Methods in Enzymology.*** **91** (Ed: Hirs, C.H.W. and Timasheff, S.N.) Academic Press, New York (1983), 399 - 413

Isenberg, I. (1978) In: *The Cell Nucleus* (Ed: Busch, H.) Academic Press, New York, 14, 135 - 154

Iwai, K., Hayashi, H. and Ishikawa, K. (1972) *J. Biochem.*, 72, 357 - 367

Jekel, P.A., Weiger, W.J. and Beintema, J.J. (1983) *Anal. Biochem.*, 134, 347 - 354

Johns, E.W. (1964) *Biochem. J.*, 92, 55 - 59

Klapper, D.G., Wilde, C.E. and Capra, J.D. (1978) *Anal. Biochem.*, 85, 126 - 131

Knowles, J.A. and Childs, G.J. (1986) *Nucleic Acids Res.*, 14, 8121 - 8133

Kossel, A. (1884) *Hoppe-Seyler's Z. Physiol. Chem.*, 8, 511 - 515

Laemmli, U.K. (1970) *Nature (London)*, 227, 680 - 685

Laine, B., Sautiére, P. and Biserte, G. (1976) *Biochemistry*, 15, 1640 - 1645

Levy, M., Fishman, L. and Schenkein, I. (1970) *Methods in Enzymology*, 19, 672 - 681

Lewin, B. (1980) In: *Gene Expression, 2nd Edition* (Ed: Lewin, B.) John Wiley and Sons, New York, 350

Lewis, P.N. (1976) *Can. J. Biochem.*, 54, 963 - 970

Lewis, P.N. and Chiu, S.S. (1980) *Eur. J. Biochem.*, 109, 369 - 376

Liao, L.W. and Cole, R.D. (1981) *J. Biol. Chem.*, 256, 3024 - 3029

Lindsey, G.G. and Thompson, P. (1992) *J. Biol. Chem.*, 267, 14622 - 14628

Lottspeich, F. (1980) *Hoppe-Seyler's Z. Physiol. Chem.*, 361, 1829 - 1834

Luck, J.M., Cook, H.A., Eldridge, N.T., Haley, M.I., Kupke, P.W. and Rasmussen, P.S. (1956) *Arch. Biochem. Biophys.*, 65, 449 - 467

Magnussen, C. (1978) *M.Sc Thesis* McGill University, Montreal

Margoliash, E. and Fitch, W.M. (1968) *Ann. N.Y. Acad. Sci.*, 151, 359 - 381

Markowicz, Y. and Loiseaux-de Goër, S. (1991) *Curr. Genet.*, 20, 427 - 430

Martinson, H.G., True, R., Lau, C.K. and Mehrabian, M. (1979) *Biochemistry*, 18, 1075 - 1082

May, G.S. and Morris, N.R. (1987) *Gene*, 58, 59 - 66

McIntosh, L. and Cattolico, R.A. (1978) *Anal. Biochem.*, 91, 600 - 612

McLaughlin, P.J. and Dayhoff, M.O. (1973) *J. Mol. Evol.*, 2, 99 - 116

Mirsky, A.E. and Ris, H. (1951) *J. Gen. Phys.*, 34, 475

Molnar, I. and Howath, C. (1977) *J. Chromatog.*, 142, 623 - 640

Moss, T., Cary, P.D., Abercrombie, B.D., Crane-Robinson, C. and Bradbury, E.M. (1976) *Eur. J. Biochem.*, 71, 337 - 350

Moss, T., Cary, P.D., Crane-Robinson, C. and Bradbury, E.M. (1976b) *Biochemistry*, 15, 2261 - 2267

Müller, K. and Schmitt, R. (1988) *Nucleic Acids Res.*, 16, 4121 - 4136

Müller, K., Linauer, A., Brüderlen, M. and Schmitt, R. (1990) *Gene*, 93, 167 - 175

Newman, S.M. and Cattolico R.A. (1987) *Plant Phys.*, 84, 483 - 490

Newmann, S.M., Derocher, J. and Cattolico, R.A. (1989) *Plant Phys.*, 91, 939 - 946

Noll, M. (1974) *Nature (London)*, 251, 249 - 251

Noll, M. (1977) In: *Nucleic Acids - Protein Recognition* (Ed: Vogel H.J.) Academic Press, New York, 139 - 150

Nomoto, M., Iwai, N., Saiga, H., Matsui, T. and Mita, T. (1987) *Nucleic Acids Res.*, 15, 5681 - 5697

Nothacker, K.D. and Hilderbrandt, A. (1985) *Eur. J. Cell. Biol.*, 39, 278 - 282

Noveau, J. and Drapeau, G.R. (1979) *J. Bacteriol.*, 140, 911 - 916

Ogawa, Y., Quagliarotti, G., Jordan, J., Taylor, C.W., Starbuck, W.C. and Busch, H. (1969) *J. Biol. Chem.*, 244, 4387 - 4392

Ohlenbusch, H.H., Olivera, B.M., Yuan, D. and Davidson, N. (1967) *J. Mol. Biol.*, 25, 299 - 315

Palau, J. and Padrós, E. (1972) *FEBS lett.*, 27, 157 - 160

Panyim, S. and Chalkley, R. (1976) *Arch. Biochem. Biophys.*, 130, 337 - 346

Patterson, G. and van Valkenburg, S. (1990) *J. Phycol.*, *26*, 484 - 489

Patthy, L. and Smith, E.L. (1975) *J. Biol. Chem.*, 250, 1919 - 1920

Quagliarotti, G., Ogawa, Y., Taylor, C.W., Sautière, P., Jordan, J., Starbuck, W.C. and Busch, H. (1969) *J. Biol. Chem.*, 244, 1796 - 1802

Rausch, H., Larsen N. and Schmitt, R. (1989) *J. Mol. Evol.*, 29, 255 - 265

Reith, M.E. and Cattolico, R.A. (1985) *Biochemistry*, 24, 2550 - 2556

Reith, M.E. and Cattolico, R.A. (1985b) *Biochemistry*, 24, 2556 - 2561

Reith, M.E. and Cattolico, R.A. (1985c) *Plant Phys.*, 79, 231 - 236

Reith, M.E. and Cattolico, R.A. (1986) *Proc. Natl. Acad. Sci. USA.*, 83, 8599

Rizzo, P.J. (1980) *Biochim. Biophys. Acta*, 624, 66 - 77

Rizzo, P.J. (1985) *Biosystems*, 18, 249 - 262

Rizzo, P.J. and Burghardt, R. (1980) *Chromosoma*, 76, 91 - 99

Rizzo, P.J. and Burghardt, R.C. (1983) *J. Phycol.*, 19, 348 - 351

Rizzo, P.J., Bradley, W. and Morris, R. (1985) *Biochemistry*, 24, 1727 - 1732

Rodrigues, J de A. (1985) *Ph.D. Thesis*, University of Cape Town

Rodriques, J de A., Brandt, W.F. and von Holt, C. (1979) *Biochim. et Biophys. Acta*, 578, 196 - 206

Ryeszotarski, W.J. and Mauger, A.B. (1973) *J. Chromatog.*, 86, 246 - 249

Sautiére, R., Tyrou, P., Laine, B., Mizon, J., Ruffin, P. and Biserte, G. (1974) *Eur. J. Biochem.*, 41, 563 - 576

Schoellman, G. and Shaw, E. 1963) *Biochemistry*, 2, 252 - 255

Schwartz, R.M. and Dayhoff, M.O. (1978) *Science*, 199, 395 - 403

Schwartz, R.M. and Dayhoff, M.O. (1979) In: *Atlas of Protein Sequence and Structure* (Ed: Dayhoff, M.O.) National Biomedical Researchc Foundation, Washington D.C., 353 - 358

Schwemmler, W. and Schenk, H.E.A. (1980) *Endocytobiology: Endosymbiosis and Cell Biology*, Walter de Gruyter and Co, Berlin

Shaw, E., Mares-Guia, M. and Cohen, W. (1965) *Biochemistry*, 4, 2219 - 2224

Shupe, K., Rizzo, P.J. and Johnson, J.R. (1980) *FEBS lett.*, 115, 221 - 224

Smith, M.M. and Andresson, O.S. (1983) *J. Mol. Biol.*, 169, 663 - 690

Smith, R.M. and Rill, R.L. (1989) *J. Biol. Chem.*, 264, 10574 - 10581

Soto, M., Requera, J.M., Jimenez-Ruiz, A.and Alonso, C. (1991) *Nucleic acids Res.*, 19, 4554

Sperling, R. and Wachtel, E.J. (1981) *Adv. Prot. Chem.*, 34, 1 - 60

Spiker, S. and Isenberg, I. (1977) *Biochemistry*, 16, 1819 - 1826

Strickland, M.S., Strickland, W.N. and von Holt, C (1980) *Eur. J. Biochem.*, 106, 541 - 548

Sures, I., Lowry, J. and Kedes, L.H. (1978) *Cell*, 15, 1033 - 1044

Tarr, G.E., Beecher, J.F., Bell, M and McKean, P.J. (1978) *Anal. Biochem.*, 84, 622 - 627

Thoma, F., Losa, R. and Koller, T. (1983) *J. Mol. Biol.*, 167, 619 - 640

Umezawa, H. (1976) *Methods in Enzymology*, 45, 678 - 695

van der Westhuyzen, D.R. and von Holt, C. (1971) *FEBS lett.*, 14, 333 - 337

van Helden, P.D., Strickland, W.N., Brandt, W.F. and von Holt, C. (1979) *Eur. J. Biochem.*, 93, 71 - 78

Vogel, A.I. (1957) In: *Practical Organic Chemistry, 3rd Edition* (Ed: Vogel, A.I.) Longmans, London, 272 - 273

von Holt, C. (1985) *BioEssays*, 3, 120 -124

von Holt, C. and Brandt, W.F. (1977) *Methods Cell. Biol.*, 16, 205 - 225

von Holt, C., Strickland, W.N., Brandt, W.F. and Strickland, M.S. (1979) *FEBS lett.*, 100, 201 - 217

Wallis, J.W., Hereford, L. and Grunstein, M. (1980) *Cell*, 22, 799 - 805

Weintraub, H. (1984) *Cell*, 38, 17 - 27

Whatley, J. and Whatley, F. (1981) *New Phytology*, 87, 233 - 247

Wilson, A.C., Carlson, S.S. and White, T.J. (1977) *Ann. Rev. Biochem.*, 46, 574 - 639

Wittman-Liebold, B. (1973) *Hoppe-Seyler's Z. Physiol. Chem.*, 354, 1415 - 1431

Wittman-Liebold, B. (1981) In: *Chemical synthesis and sequencing of Peptides and Proteins* (Eds: Liu, T-Y., Schechter, A.N., Heinrikson, R. and Condliffe, P.) Elsevier North Holland, Amsterdam, 75 - 110

Woese, C.R. and Fox, G.E. (1977) *J. Mol. Evol.*, 10, 1 - 6

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) *Proc. Natl. Acad. Sci. USA.*, 87, 4576 - 4579

Woudt, L.P., Pastink, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Mager, W.H. and Planta, R.J. (1983) *Nucleic Acids Res.*, 11, 5347 - 5360

Wray, W.P., Conn, M. and Wray, V.P. (1977) *Methods Cell. Biol.*, 16, 69 - 86

Wu, C-I. and Li, W-H. (1985) *Proc. Natl. Acad. Sci. USA.*, 82, 1741 - 1745

Wu, M., Allis, C.D., Richman, R., Cook, R.G. and Gorovsky, M.A. (1986) *Proc. Natl. Acad. Sci. USA.*, 83, 8674 - 8678

Wu, S-C., Györgyey, J. and Dudits, D. (1989) *Nucleic Acids Res.*, 17, 3057 - 3063

Wynne, M.J. (1978) In: *Introduction to the Algae. Structure and Reproduction* (Eds: Bold, H.C. and Wynne, M.J.) Prentice Hall, London, 577

Yeoman, L.C., Olson, M.O.J., Sugano, N., Jordan, J.J., Taylor, C.W., Starbuck, W.C. and Busch, H. (1972) *J. Biol. Chem.*, 247, 6018 - 6023

Young, J.Z. (1962) In: *The Life of the Vertebrates, 2nd Edition*, Oxford University Press, New York, 389 and 539 - 541

Zentgraf, H., Muller, V. and Franke, W.W. (1980) *Eur. J. Cell. Biol.*, 20, 254 - 264

Zweidler, A. (1978) *Methods Cell. Biol.*, 17, 223 - 233