BIOGEOCHEMISTRY OF BRACHIOPOD INTRACRYSTALLINE PROTEINS AND AMINO ACIDS

Derek Walton

Thesis submitted for the degree of Doctor of Philosophy

Department of Geology and Applied Geology
University of Glasgow
June 1992

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

For my Mum and Dad, and for Caroline

Acknowledgements

This project was funded by a studentship from the U.K. Natural Environment Research Council (GT4/89/GS/42), which I gratefully acknowledge. The project also benefited from the financial support of the Department of Geology and Applied Geology, University of Glasgow, who provided funds for the chemicals required to carry out the analyses.

This thesis has depended on the support of a great many people throughout the U.K. and around the world. I would especially like to thank my supervisor, Dr. Gordon B. Curry, for his support and enthusiasm throughout the course of the project, and also for critically reading the first draft of this thesis. My lab. colleagues, Dr. Maggie Cusack, Dr. Rosalind Quinn, Dr. Kazuyoshi Endo (Yoshi), Mrs. Ikbal al-Rikabi, Heather Clegg, and Lynda Mitchell, are also deeply thanked, both for their stimulating ideas in the lab., and also for bearing with my sense of humour throughout the project. The technical support, in preparing samples for analysis, of Sandra MacCormack is also gratefully acknowledged, as it would not have been possible for this project to be completed on time if it were not for her help. Also in the Department of Geology and Applied Geology, I would like to thank Mr. Roddy Morrison who could find me anything I needed for my project, and Professor B.E. Leake for allowing me to use the facilities of the Department. Douglas MacLean is thanked for the photography necessary throughout this project. Sir Alwyn Williams is thanked for allowing me to occupy one of the visitor offices in his suite, which was especially important during the writing of this thesis, and also for allowing me to use the photograph reproduced in **figure 2.5**.

Dr. Maggie Cusack, who has added more to this project than she will ever really know, is particularly thanked for her patience in teaching biochemistry to a 'dumb' geologist, for critically reading the first draft of this thesis and for her continuing advice on all matters. Thank you Maggie.

From my field expeditions, I would like to thank the people of New Zealand in general who made me welcome wherever I went. In particular, I would like to thank Professor Doug Campbell and Dr. Daphne Lee (Department of Geology, University of Otago) for advice and for office space in the Department. At the Portobello Marine Laboratory, University of Otago, I would like to thank Dr. Keith Probert for office and living space, and Chris Spiers, captain of the *R.V. Munida*, for assistance in collecting Recent brachiopods. I would also like to thank Drs. Daphne Lee (University of Otago), Charles Thayer (University of Pennsylvania) and Melissa Rhodes (Academy of Natural Sciences, Philadelphia) for providing samples of Recent brachiopods. My friends Ed Wright (Dunedin) and the Emmett Family (Auckland) are thanked for their hospitality during my visit. I would also like to acknowledge the support of the NERC in providing

me with funds to collect the samples.

The project depended on an Applied Biosystems 420H amino acid analyser. From Applied Biosystems, I would like to thank all members of the service team who kept the analyser up and running, especially Ivan Gatensby who taught me how to use and maintain the instrument and Sandra Allen who constantly listened to my problems with the instrument on the telephone. Dr. Phil Jackson (Applied Biosystems) is thanked for his advice on amino acid chemistry, and for supplying the photographs of the amino acid analyser used in **Chapter Three**.

I would like to thank Dr. Howard Armstrong (University of Durham) and Professor J.R. Cann (University of Leeds; both formerly at the Department of Geology, University of Newcastle upon Tyne) for stimulating my interest in Molecular Palaeontology, and in research in general. My flat mates at 308 (1/2), Gian Rizzi, Mark Osborne and Steven Penman, are thanked for their company and sense of humour. In the Department of Geology and Applied Geology, I would like to thank Dr. Lindsay Ferguson, Joe Crummy, Simon Stromberg, Dr. George Bowes, Dr. Mark Wilkinson and Dr. Andy Mullis, as well as my lab. and geological flat mates, for making the Department a more pleasant place in which to work. The leaders and boys of the 191st Glasgow Scout Group are also thanked for the effect they had on this project, by taking me away from it for a few hours each week.

I would like to thank my Mum and Dad, Frank and Elizabeth Walton, for always encouraging me to do what I thought was for the best, and to my brothers Steve and Stuart for their continual support. Finally, I would like to thank Caroline for her continuing support, for keeping my feet on the ground, for critically reading the drafts of this thesis, and for allowing my second love to come between us for the last three years. Roll on 10th April 1993.

Declaration

The material presented in this thesis summarises the results of three years of research carried out in the Department of Geology and Applied Geology of the University of Glasgow, under the supervision of Dr. Gordon B. Curry. This study is based on my own independent research and any previously published or unpublished results of other researchers used in this thesis have been given full acknowledgement in the text.

Derek Walton
June 1992.

"We need to move away from the 'stamp collecting'
phase of molecular palaeontology
to a more systematic survey"

Dr. G.A. Dover
Phil. Trans. R. Soc. Lond B., **333**, p.427.

Contents

List of figures

List of tables

Abbreviations

APS	Ammonium persulphate
Bis	N, N'-methylene bisacrylamide
CBB-R250	Coomassie Brilliant Blue R250 protein stain
DIEA	Diisopropylethylamine
DNA	Deoxyribonucleic acid
EDTA (Na_2, K_3)	Ethylenediaminotetraacetic acid (disodium, tripotassium salt)
g	grams
hplc	high pressure (or performance) liquid chromatography
fplc	fast protein liquid chromatography
kDa	kilodaltons
mg	milligrams
mL	millilitre
mm	millimetres
mM	milliMolar
Ma	Million years
MilliQ™	deionised pure water
ng	nanograms
nm	nanometres
nmol	nanomoles
PAGE	Poly acrylamide gel electrophoresis
PCA	Principal components analysis
PITC	phenylisothiocyanate
pmol	picomoles
RNA (m, r, t)	Ribonucleic acid (messenger, ribosomal, transfer)
SDS	Sodium dodecyl sulphate
TCA	Trichloroacetic acid
TEMED	N, N, N', N'-tetramethylethylene-diamine
TFA	Trifluoroacetic acid
Tris	[hydroxymethyl]-aminomethane
UPGMA	Unweighted pair-group method using arithmetic averages
UV	ultraviolet light
γ-ABA	γ-aminobutyric acid
μg	micrograms
μL	microlitre

μm micrometres

Three letter and one letter codes of amino acids

Amino Acid	3 letter	1 letter	Amino Acid	3 letter	1 letter
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Apartic Acid	Asp	D	Methionine	Met	M
Asparagine	Asn	N	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic Acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Summary

Amino acids were released from a range of fossil and Recent samples, sediments and fingertips. Fingertips contain up to 100 times the concentration of amino acids in the fossils, and sediments up to 10 times the concentration. However, the distribution of amino acids in these samples were significantly different from those in fossil samples, and were easily discriminated using statistical methods.

Incarcerated molecules (protein, lipids and carbohydrates) from Recent samples were released from the calcium carbonate of the shell into solution. The protein fraction of the samples were purified to homogeneity using SDS PAGE, and the separated proteins analysed by partial *N*-terminal sequencing and amino acid analysis. The crude extract was also fractionated by reverse phase liquid chromatography. Following the partial characterisation of these intact, extant molecules, attempts were made using similar techniques to fractionate the bulk organic extract from related fossils. This only resulted in broad bands or peaks of low molecular weight compounds at low concentrations, rather than the sharp peaks which resulted from the analysis of Recent extracts, and it was not possible to purify the molecules using these methods. It was also concluded that during the filtration stage of preparation the majority of the molecules were lost.

A different method of preparation was applied to fossil samples, which did not include any concentration or filtration steps, and the free amino acids and small peptides which are present in the acid soluble fossil extracts were therefore also quantified. Amino acid analysis of these extracts from fossils revealed that the proteins which were originally present within the shells have undergone natural hydrolysis reactions (cleavage of peptide bonds as a result of time or heat) and are highly degraded, indicating why the separations by biochemical techniques could not be applied to fossil samples. By 0.2 Ma, up to 80% of the amino acids are present in the free state, indicating that the majority of the peptide bonds in the sample have been broken, leading to the production of larger numbers of small peptides. Some peptides remain, indicated by increased concentrations of amino acids following acid hydrolysis. Proteins from intracrystalline sites within fossils are therefore in a poor state of preservation, and it is unlikely that it will be possible to separate and concentrate these molecules in order to complete primary sequence analysis. The peptide bonds break rapidly in the samples in response to the action of temperature, the presence of water, and the nature of the residues present on either side of the peptide bonds.

Individual amino acids also undergo degradative reactions in the fossil record, and the molecules may be grouped in terms of their stability. The decomposition of most amino acids may be described by exponential or logarithmic curves, indicating a rapid rate of decomposition in the fossil record. The amino acids are generally degraded more rapidly in the free state than when they are bound into peptides. Degradation products may be non-amino acids, non-standard amino acids or proteinogenic amino acids, depending

on the original molecule which has decomposed. The advantage of utilising intracrystalline molecules is that the degradation products of both proteins and amino acids remain within the shell and are not leached out (as is the case for intercrystalline molecules) ensuring that any preserved taxonomic signal remains within the shell.

Despite the severe degradation of the molecules, taxonomic information is still preserved within the samples, although at a lower level of discrimination than in the Recent, and this information may be revealed by statistical analyses. No homogenisation of the amino acids in the sample had taken place, and that subordinal level discrimination is possible to at least 0.4 Ma. Older samples showed merging of samples from similar orders, although different orders and phyla could easily be discriminated by this method. When samples are analysed together from all horizons using these statistical methods, samples could still be discriminated to the ordinal level, no matter what the age of the sample, indicating that although proteins and amino acids were decomposed rapidly, some taxonomic signal remains in the shell. The amino acid content of fossils can also be used for several other applications, including dating, correlation and palaeoenvironmental analyses.

The amino acid composition of the insoluble fraction of the shell shows no enrichment in unstable or basic amino acids, and the proportion of the sample which is insoluble also show no direct relationship with the age of the sample. In all cases, the amino acid concentration of the insolubles was less than that of the soluble fraction.

Chapter One: The molecular approach to palaeontology

The molecular approach to palaeontology

1.1 Introduction

The explosion of research in biochemistry and biotechnology has implications throughout scientific research. The explanation and understanding of the principles of heredity and speciation are especially important in biology, and, as a consequence, in palaeontology. Whereas species definition and characterisation in biology is increasing using molecular information, this is not the case in palaeontology. Previous work has described and attempted to characterise molecules trapped in fossils (**section 1.5**), although this has followed no systematic pattern. In the present study, the main aim was to analyse fossil samples from a young, well dated succession which also have extant representatives, in order to give a quantitative and qualitative assessment of the geochemistry of proteins and amino acids and the taxonomic information which may be gained from these molecules, and the role which organic molecules may play in geology in general.

Molecular palaeontology is an integration of molecular biology and geology, studying organic molecules to gain information regarding the past. The chapter introduces the concept of molecular palaeontology, and reviews the benchmark papers in the study of the biochemistry of fossils. This chapter also acts as an introduction to the organic molecules examined in this study, and discusses their *in vivo* function and chemistry. Finally, the flow of genetic information, present in the DNA of an organism, is considered in terms of translation to proteins via protein manufacture, and how this affects the corresponding structure and chemistry of the molecule. This is intended only to serve as an introduction, as numerous texts cover the subject in much greater detail. A large part of this section is adapted from Stryer (1988). Finally, the chapter deals with the basic principles of numerical taxonomy, and introduces the statistical techniques used in the study.

Immediately prior to the beginning of this project, the Department of Geology and Applied Geology at the University of Glasgow completed the building of a Molecular Palaeontology laboratory, comprising preparation rooms, a clean room (which is up to 10 000 times cleaner than average laboratories) and a complete range of biochemical equipment, ready to begin direct characterisation of organic molecules useful to geology.

At the same time, research into indirect methods of molecular analysis via immunology were ongoing in the department (e.g. Quinn, 1990; Endo and Curry, 1991; Endo, 1992), providing the basis for further, more direct, study of the molecules. Interdisciplinary research was further advanced by increased contact with other departments and the appointment of a biochemist to study the intracrystalline molecules released from extant brachiopods.

As a geologist, confronted with biogeochemical investigation, a period of theoretical and practical retraining was necessary before research could begin. The basics of biochemistry and of the operation and maintenance of the relevant equipment were all

necessary facets of the project. In order to extract relevant information from the fossils, analytical methods were adapted from the literature and new methods were derived (**Chapter Three**) before investigation into the molecules could begin.

1.2 Molecular Palaeontology

Palaeontology is the study of past life, and, as such, must utilise as many techniques as possible in order to gain a view of this. Recent incremental increases in the sensitivity of biochemical instrumentation has led to a better understanding of the processes involved in the biology of organisms. The advances in both organic geochemistry and molecular biology have important implications for palaeontology (Curry, 1987). Large accumulations of organic matter in the lithosphere have long been recognised in the form of economic deposits such as oil and coal and, on average, 2% of the mass of sedimentary rocks is due to organic matter, although most of these degraded deposits do not yield molecules with information of any use in phylogenetic investigation. The amount of carbon stored in the lithosphere since photosynthesis began may be up to 10 000 times that in the present day biomass (Ourisson *et al.*, 1984), representing an enormous potential for the recovery of information.

The term "Molecular Palaeontology" was originally coined by Calvin (1968) and applied to the study of biological marker compounds in the organic geochemistry of oils and coals. The term was redefined by Runnegar (1986) to include molecules sourced directly from living and fossil organisms (as opposed to having been released from the organism on death or decomposition). Runnegar (1986) also defined the three main areas where molecular biology impinges on palaeontology:

(i) The traditional study of fossil molecules and their degradation products. This should include **any** organic molecule formed during the life of an organism, which has then been preserved in the fossil record, no matter how it has been degraded through time.

(ii) The role of biopolymers in the construction of the mineral and carbohydrate skeletons of fossils (the field of biomineralisation).

(iii) The study of historical information which may be derived from the primary structures of nucleic acids and proteins of **living** organisms.

In such a relatively new field, the terminology of molecular palaeontology has not yet become consistent throughout the literature (e.g. Summons, 1988). In this study, terminology is based on that proposed by Runnegar (1986). The term "fossil molecules" will be used in reference to the original molecules which have undergone fossilisation, whereas "molecular fossils" refers to molecules extracted from extant organisms and used to provide information regarding the past, in a way similar to the so-called living fossils *Lingula* and *Latimeria* (coelacanth). In the literature, these terms have been interchanged

with incredible frequency.

Fossil molecules have been studied for some time; more than 50 years ago, it was suggested that a vanadyl porphyrin, a common molecule in oil shales, was the degradation product of chlorophyll *a*, a molecule found in all photosynthetic cells (Treibs, 1934). However, it was not until Ekstrom *et al.* (1983) determined the crystal structure of the molecule from a Cretaceous oil shale that this was confirmed. This was a major advance in the study of organic geochemistry and also for molecular palaeontology, although it also sounded a warning; the molecule had undergone degradative reactions, changing the nature of the molecule by substituting the central magnesium ion for a vanadyl ion (**figure 1.1**). Molecules within the fossil record could therefore still be recognised, even though they had undergone some degradation.

A class of lipids, known as hopanoids, were first described from the fossil

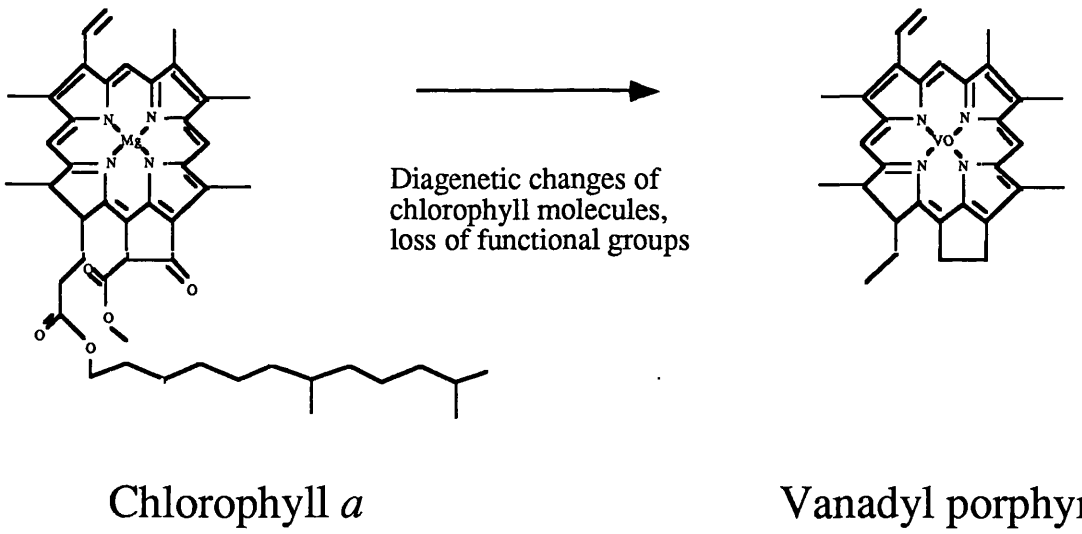


Figure 1.1: Conversion of the biological molecule chlorophyll *a* to the geochemical marker compound vanadyl porphyrin. The degradation of the molecule takes place as a stepwise process, but the final product may be identified as being sourced from the parent (after Ekstrom *et al.*, 1983).

record, where they occur in large concentrations (Ourisson *et al.*, 1979). As these molecules are only soluble in a mixture of polar and non-polar solvents (e.g. chloroform and methanol), it was not until a specific search for them was made that they were found within the cell membranes of living bacteria. In this way, these molecules have an analogue with "conventional" palaeontology, as *Latimeria* was described from the fossil record prior to its discovery in the extant fauna.

There are four main classes of molecules present within the cells of organisms,

namely nucleic acids, proteins, carbohydrates and lipids. The study of fossil molecules relies on finding a balance between the stability of molecules over time, and the information which they carry regarding the genetics of the fossil organism (Runnegar, 1986). Information-rich molecules such as DNA and RNA are inherently unstable due to the presence of sugars in the molecular backbone (**section 1.4**), and hence are unlikely to be preserved in the fossil record, unless some form of exceptional preservation occurs (e.g. Golenberg *et al.*, 1990). The information-poor carbohydrates are unlikely to be preserved for the same reason.

Lipids are far more stable in geological environments, although they are also information-poor molecules. Proteins tend to be degraded more rapidly than lipids (Meyers *et al.*, 1980; Benner *et al.*, 1984; Hedges *et al.*, 1985), but at rates significantly slower than the nucleic acids. As proteins contain information derived from RNA templates, which were in turn derived from DNA, they represent a class of information rich molecules with some potential for fossilisation. The stability of these molecules, and a review of the literature dealing with degradation will be considered later (**Chapters Six and Seven**).

1.3 Molecular function and chemistry

1.3.1 Amino acids

Amino acids are naturally occurring organic acids, ubiquitous in living organisms, and are the building blocks of proteins. All amino acids contain a hydrogen atom, a carboxyl group (-COOH), an amino group ($-NH_2$) and a side chain (denoted by R), bonded to a single carbon atom (known as the α-carbon; **figure 1.2**). In all amino acids these four parts of the molecule are bonded to the α-carbon in a tetrahedral array. This conformation causes optical activity in the molecules; the amino acids (with the exception of glycine) can occur as isomers which are mirror images of each other, known as the D-

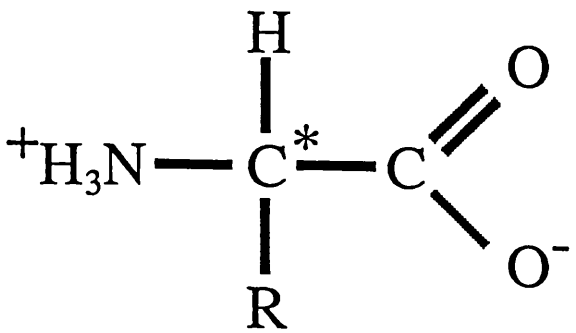


Figure 1.2: General formula for the amino acids. All amino acids contain an amino group ($-NH^+_3$), a hydrogen atom, a carboxyl group ($-COO^-$) and a side chain (R) bonded covalently to the alpha carbon (*).

and L- isomers, and which will rotate a beam of plane polarised light in different directions. Only L-isomers normally occur in biological systems, although some bacteria do contain D-isomers. Over time, however, these undergo racemisation reactions (for a review, see Schroeder and Bada, 1976), whereby L-isomers revert to the D- form and eventually an equilibrium (D:L ≈ 1:1) is attained.

The side chains of amino acids differ in size, shape, charge, hydrogen bonding capacity and chemical reactivity. In proteins, twenty types of side chain are commonly found; and proteins from all species, from bacteria to humans, contain these same twenty 'proteinogenic' amino acids. The structure and nature of these side chains determines the conformation and reactivity of the protein. Five amino acids, glycine (denoted by Gly, or G), alanine (Ala, A), valine (Val, V), leucine (Leu, L) and isoleucine (Ile, I) have aliphatic side chains ranging in size from a single hydrogen atom in Gly to C_4H_9 in Leu and Ile (**figure 1.3**). This type of side chain is hydrophobic, which aids the stabilisation of the three dimensional structure of the water soluble proteins, and is also linear, which allows compact structures to be developed.

Proline (Pro, P) also has an aliphatic side chain, but this differs from the above

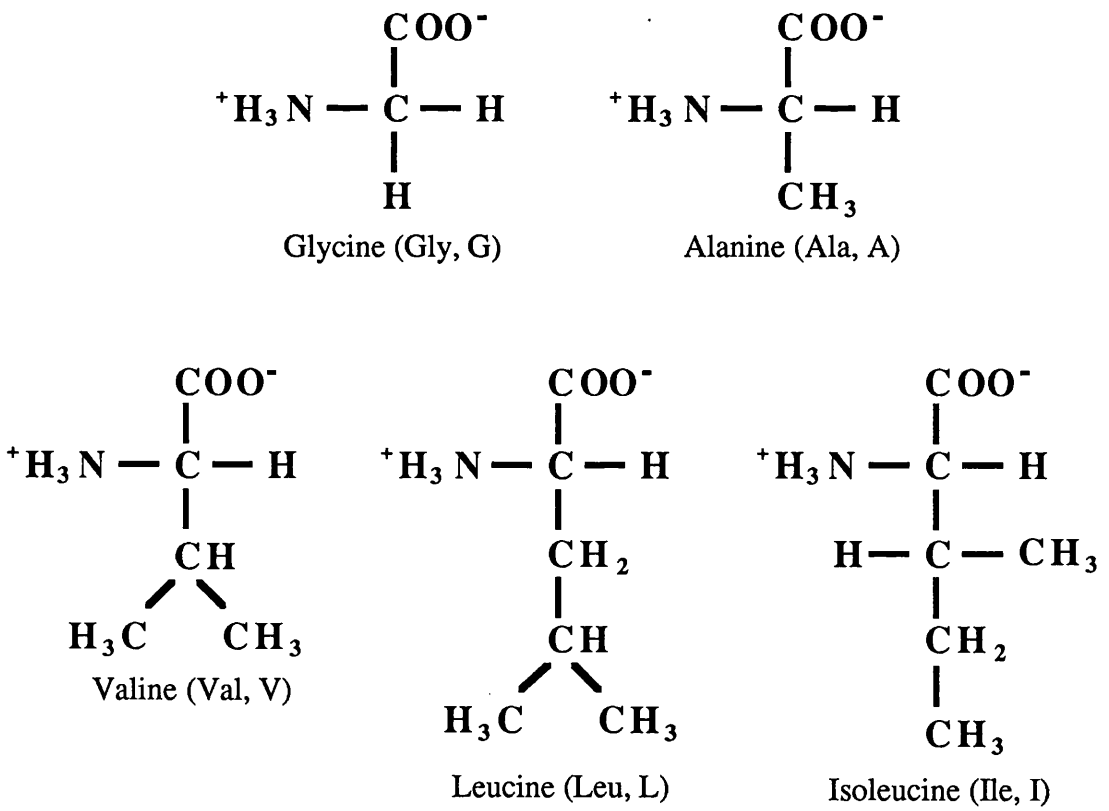
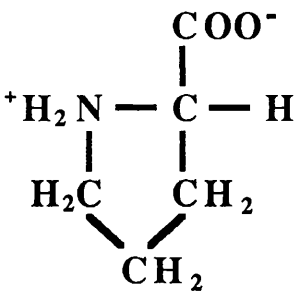


Figure 1.3: Structural diagrams of the amino acids which contain hydrophobic aliphatic side chains.

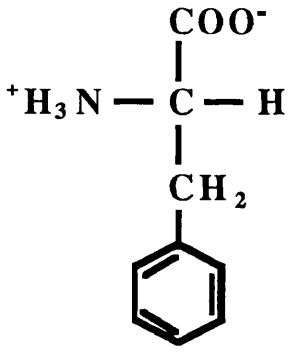
by bonding both to the carbon atom and amino group, resulting in the formation of an aliphatic ring (**figure 1.4**) which may affect the protein structure (polypeptide chains often bend at Pro residues). Aromatic ring structures are found in the side chains of three amino acids (**figure 1.5**), phenylalanine (Phe, F), tyrosine (Tyr, Y) and tryptophan (Trp, W). The side chains of Trp and Phe are highly hydrophobic, whereas that of Tyr contains a hydroxyl group (-OH) which is active, and which reduces the hydrophobicity. The aromatic rings of these three amino acid side chains contain delocalised electrons in pi-clouds (above and below the aromatic ring) which enables electron transfer and interactions with other pi-clouds.

Aliphatic hydroxyl groups (**figure 1.6**) are contained in the side chains of

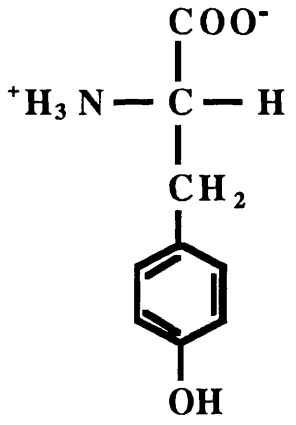


Proline (Pro, P)

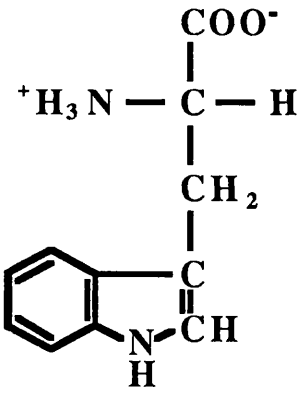
Figure 1.4: Structural diagram of the amino acid Pro, which differs from the remaining amino acids by the formation of an aliphatic ring in the side chain. This structure is important in protein architecture as protein chains often fold at Pro residues.



Phenylalanine (Phe, F)



Tyrosine (Tyr, Y)

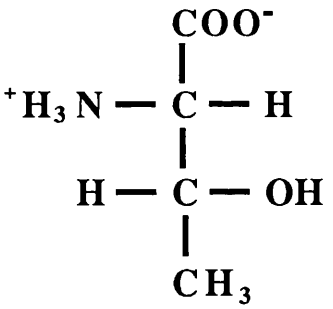


Tryptophan (Trp, W)

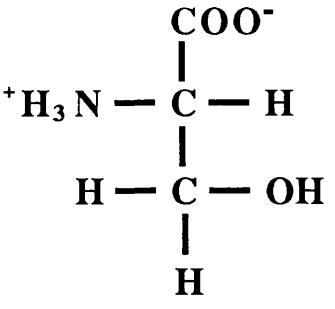
Figure 1.5: Structural diagrams of the amino acids which contain aromatic rings in their side chains. Of these, Trp is not identified in amino acid analysis due to acid attack on the side chain ring (**section 3.7.2**)

serine (Ser, S) and threonine (Thr, T). These hydroxyl groups render these amino acids more reactive and hydrophilic than others with aliphatic side chains. Methionine (Met, M) and cysteine (Cys, C) contain sulphur atoms in the side chain (**figure 1.7**), which are hydrophobic. The position of Cys is important in protein architecture, as the molecules are highly reactive and form disulphide bonds with other Cys molecules.

The remaining amino acids have side chains which are charged. The basic

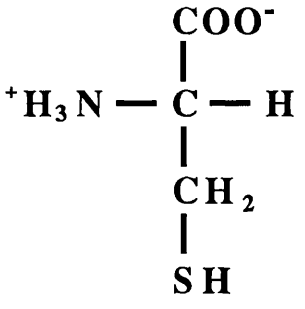


Threonine (Thr, T)

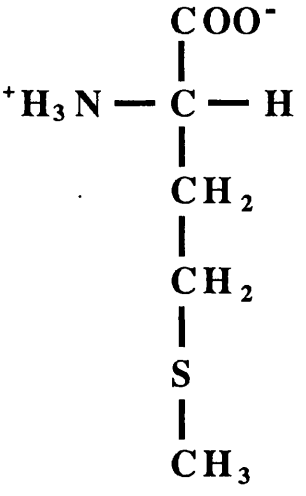


Serine (Ser, S)

Figure 1.6: Structural diagrams of the amino acids which contain hydroxl groups in their aliphatic side chains, rendering the side chains more reactive and hydrophobic than the other aliphatic side chains.



Cysteine (Cys, C)



Methionine (Met, M)

Figure 1.7: Structural diagrams of the amino acids which contain sulphur atoms within their side chains. These amino acids are difficult to quantify on the amino acid analyser as they are susceptible to oxidation (**section 3.7.2**).

amino acids, lysine (Lys, K), arginine (Arg, R) and histidine (His, H), contain a second amino group in their side chains (**figure 1.8**). These side chains are highly charged and polar, which causes them to be hydrophilic. Lys and Arg are positively charged at neutral pH. His may either be positively charged or uncharged, depending on the local conditions, which is exploited in enzymes which the charge on a His residue may be rapidly changed as an aid to the making and breaking of bonds. Aspartic acid (Asp, D) and glutamic acid (Glu, E) have acidic side chains, which is caused by the presence of a second carboxyl group (**figure 1.9**). These are nearly always negatively charged at neutral pH. The final two proteinogenic amino acids, asparagine (Asn, N) and glutamine (Gln, Q) are uncharged derivatives of the acidic amino acids, and contain a terminal amide group instead of a

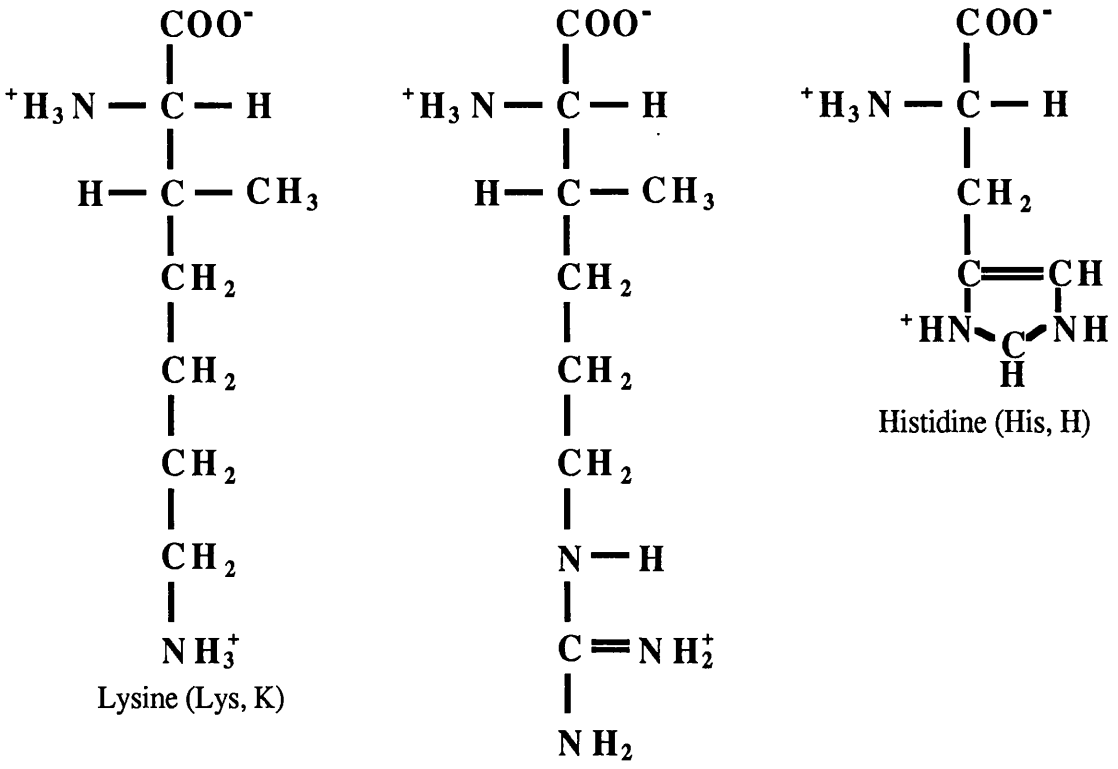


Figure 1.8: Structural diagrams of the amino acids which contain secondary amino groups within their side chains, rendering the side chains hydrophobic.

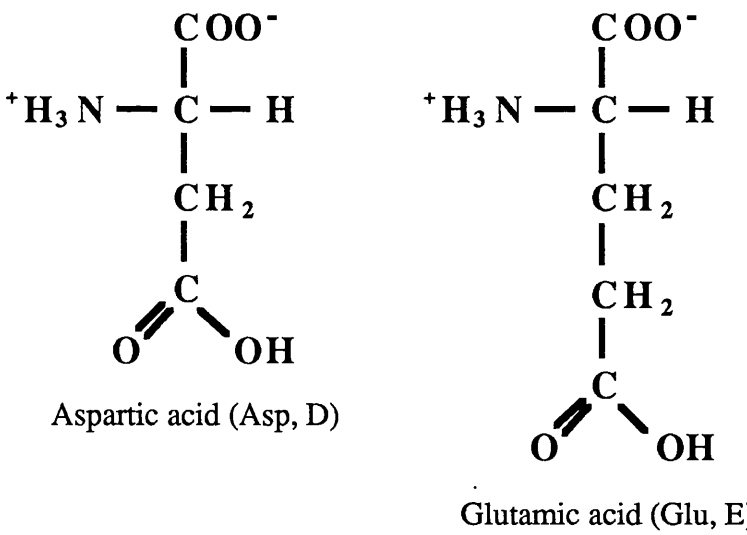


Figure 1.9: Structural diagrams of the amino acids which contain an acidic side chain caused by secondary carboxyl groups in their side chains.

carboxylate group (**figure 1.10**).

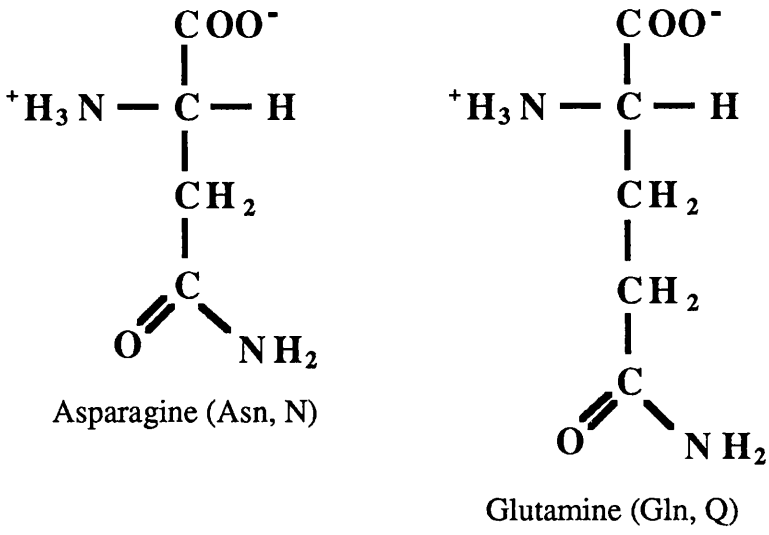


Figure 1.10: Structural diagrams of the uncharged derivatives of Asp and Glu. These amino acids are converted to the acidic representatives by acid hydrolysis (**section 3.7.2**).

1.3.2 Proteins

Amino acids are linked together into polymeric compounds. As they are linked by peptide bonds (**figure 1.11**), resulting chains are known as peptides (with a relative molecular mass of between 100 and 1000, also known as 1 kilodalton, 1 kDa), and the amino acids within the chain are known as residues. Polypeptides are those chains with a molecular weight of between 1 kDa and 10 kDa, a length of approximately 10 to 100 amino acids, and once the chain is larger than this, the molecule is known as a protein (s.s.). The division between polypeptides and proteins is subjective, and both polypeptides and proteins are ubiquitous in living systems. Other polymeric compounds, such as polysaccharides, usually contain a single monomer unit. Proteins differ from these polymers by the fact that any of the twenty protein amino acids can be a monomer unit.

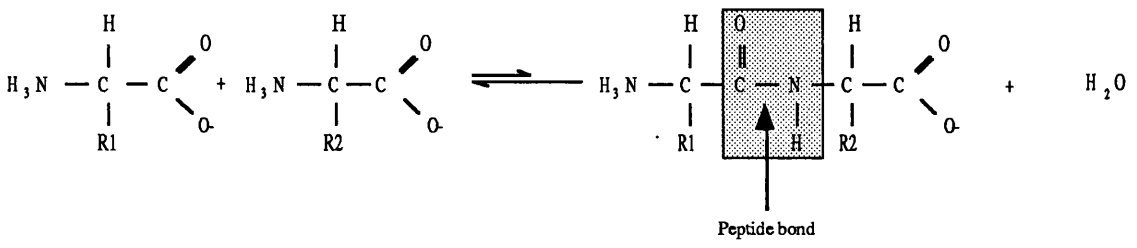


Figure 1.11: Diagram representing the formation of a peptide bond. The equilibrium is pushed to the left, favouring the cleavage of the bond to release free amino acids. This bond is formed on the ribosomes during the process of protein formation.

The amino acids are bonded together by covalent peptide bonds between the α-carboxyl group (that group bonded to the α-carbon atom) and the α-amino group of adjacent amino acids. This linkage by the peptide bond gives rise to a uniform chain known as the protein backbone. The order of the amino acids along the protein backbone defines the primary structure (or primary sequence) of the protein. The bonding peptide group in proteins is both rigid and planar, although there is a large degree of rotational freedom about the covalent bond joining the peptide group to the α-carbon of the adjacent amino acid. The polypeptide chains are able to fold into regular repeating structural motifs known as the α-helix and β-sheet, which are governed by the interactions between the side chains of amino acids (secondary structure). The conformation of proteins is dependent upon the chemical characteristics of the amino acid side chain; hydrogen bonds form between the NH and CO groups in the formation of α-helices and β-sheets. Eleven of the common protein forming amino acids are able to participate in hydrogen bonding. Trp and Arg are hydrogen bond donors only, and Asn, Gln, Ser and Thr are hydrogen bond donors and acceptors. The hydrogen bonding of the acidic and basic amino acids is pH dependent. The overall structure of the protein is known as tertiary structure.

Proteins play crucial roles in almost all biological processes. These may be grouped in terms of function:

(i) **Enzymatic catalysis:** all enzymes are proteins. Enzymes catalyse reactions with an enormous range of complexity and usually increase the rate of reaction in the order of a million fold.

(ii) **Transport and storage:** many small molecules and ions are transported by proteins, for example transport of oxygen by haemoglobin. Iron is transported by transferrin and is stored in the liver as ferritin.

(iii) **Coordinated motion:** proteins are the major component of muscle. The separation and movement of chromosomes during cell replication is produced by contractile assemblages of proteins.

(iv) **Mechanical support:** the strength of materials such as skin and bone is due to the nature of the structural protein, collagen, which they contain.

(v) **Immune protection:** the immune system of the body produces highly specific proteins, antibodies, which are used to recognise and combine with foreign substances within the body.

(vi) **Generation and transmission of nerve impulses:** the response of nerve cells to stimuli is controlled by protein. In the rod cells of the retina, the protein rhodopsin acts as a photoreceptor. Proteins are also present in synapses, where they are responsible for the continued transmission of the nervous impulse.

(vii) **Control of growth and differentiation:** control of the growth and differentiation in higher organisms is controlled by growth factor proteins. The activities of different cells are often coordinated by hormones, of which

many, such as insulin, are proteins.

Proteins are therefore crucial components in many of the fundamental processes of living organisms.

1.4 Formation of molecules and expression of genetic information

1.4.1 DNA and the genetic code

Deoxyribonucleic acid (DNA) carries all genetic information in the eucaryotic cell; i.e. genes are composed of DNA. The structure of this compound, which occurs in all living organisms (except some simple viruses), was described as a double helix by Watson and Crick in 1953. Each helix is composed of four bases (known as nucleotide bases); adenine (A), thymine (T), guanine (G) and cytosine (C), together with phosphates and deoxyribose sugars. The bases from opposite helices form bonded pairs, A-T, G-C and *vice versa*, at the centre of the helix, with the sugars and the bases around the outside.

Replication of DNA occurs in a semiconservative manner, the helix splits and two new strands are added (one to each chain), so the daughter molecules contain one strand each from the parent molecule. It is in this way that the genetic information is passed on between generations of DNA molecules.

DNA is the source of the information for protein synthesis. The sequence of the bases on the DNA molecule determines the sequence of amino acids on the protein. The information contained in this sequence is passed from the DNA to the protein, although there are four nucleotide bases, compared to twenty amino acids used in the manufacture of proteins. By simple calculation, a singlet code (one base = one amino acid) would produce four possibilities, a doublet (two bases = one amino acid) sixteen and a triplet code sixty four. From this, it was deduced (Crick *et al.*, 1961) that an amino acid is coded for by a triplet of nucleotide bases (three bases = one amino acid, known as a codon). The genetic code (**figure 1.12**), as this triplet of nucleotide bases is known, is the key to molecular biology, and is almost universal; differences to this code are only found in some mitochondrial DNA and in ciliates, which branched off the evolutionary tree very early in the history of the Earth. The genetic code in its present form has existed for at least two billion years. From the large number of coding possibilities, compared to the number of amino acids used in proteins, it is obvious that some codes will be duplicate, and that the genetic code is degenerate; i.e. the amino acids are coded for by more than one triplet code. The triplet code acts as a safety mechanism in protein formation, as can be seen in **figure 1.12**, a change in one of the nucleotides in the DNA chain does not necessarily cause a change in the corresponding amino acid. If a single base coded for a single amino acid, then any mutation of the DNA would always change the corresponding amino acid.

There is, however, rarely a direct comparison between the primary sequence of the protein and the order of nucleotide bases in the gene expressed by the protein. In

First position	Second position				Third position
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Figure 1.12: The genetic code, illustrating which triplets of nucleotide bases code for which amino acid. The code is degenerate, with more than one codon occurring for each amino acid. The codon AUG is the initiation code, and also codes for internal Met.

most eucaryote organisms, genes contain areas of nucleotide sequence which are expressed (exons) interspersed with areas which are not (introns). Introns are read by the mRNA, but are removed or 'spliced out' during the polymerisation of the peptide chain. This has implications for evolution to be discussed below.

The genetic code has several other features that are important in molecular evolution. The DNA template is not read at random along the chain; it has a definite starting point, or start codon, which is defined in the same way as an amino acid. The way in which the series of bases is read is also important, it is possible that codons (blocks of three amino acids) may be read separately, or they may overlap. Amino acid sequences of mutated Tobacco Mosaic Virus and abnormal haemoglobins showed that substitutions of a single nucleotide base usually altered only one amino acid; thus if the third base on the DNA molecule is altered, only the first amino acid is changed, rather than the first three (**figure 1.13**). However, if a base is deleted or added to the DNA molecule, then all subsequent codons will be affected.

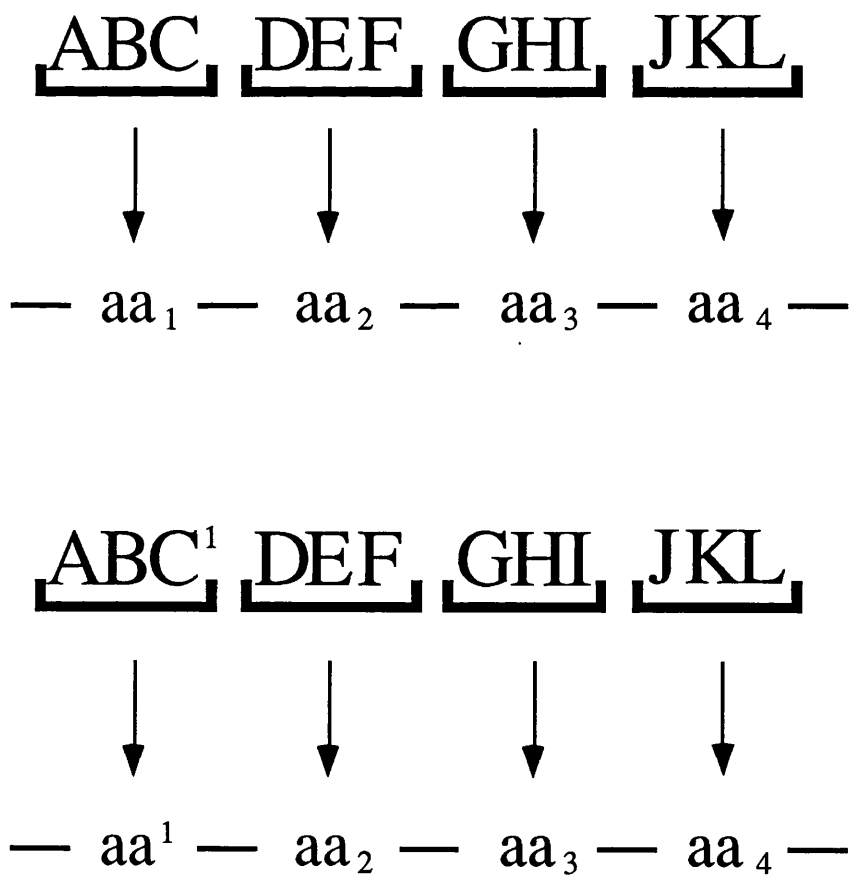


Figure 1.13: The reading frame utilised during translation of genetic information from mRNA to proteins. The frame reads each codon successively in blocks of three rather than overlapping. If one of the nucleotide bases has mutated, then the corresponding amino acid does not always change (**figure 1.12**).

1.4.2 Formation of proteins

The genetic information in the genome (contained in DNA) needs to be transferred from the cell nucleus, to areas of protein synthesis in the cell, the cytoplasm, on ribosomes attached to the endoplasmic reticulum of the cell. In the ribosomes, amino acids undergo condensation reactions to become linked together by peptide bonds into chains. As the newly formed peptide moves away from the area of active bond formation, the chain passes through the membrane of the endoplasmic reticulum. As this occurs, the chemical properties of the amino acid side chains begin to cause the chain to fold. This action takes place spontaneously, so that hydrophobic side chains become buried within the protein mass and the polar, charged and hydrophilic side chains are exposed. Other interactions in the peptide will also influence protein architecture, such as the formation of disulphide bridges between Cys residues. The amino acid sequence therefore specifies the three dimensional structure of the protein, and this tertiary structure may be predicted from the primary sequence, although X-ray crystallography is necessary to confirm the three dimensional structure of proteins.

The mechanism by which the genetic information contained within DNA is passed on to proteins was discovered in stages. The discovery of ribonucleic acid (RNA) as a messenger (Brenner *et al.*, 1961) within the cell between DNA and the site of protein formation elucidated the path of protein formation. RNA is similar to DNA in that it consists of four bases (here uracil (U) replaces thymine), sugars (ribose, rather than deoxyribose) and phosphates. The molecule is a long, unbranched polymer, normally single stranded, which is able to interact with DNA. There are several types of RNA within the cell, all of which probably play a crucial role in protein synthesis, and all of which are able to hybridise with DNA, indicating that they were formed from DNA templates. Messenger RNA (mRNA) is the actual template for protein synthesis; by interacting with DNA, mRNA forms a complementary molecule for each gene or group of genes which is to be expressed in protein synthesis, with the ends being denoted by appropriate start and stop codons. The nucleotide bases on the mRNA are the complementary bonding pair to the bases on the DNA. Transfer RNA (tRNA) carries amino acids, activated ready for protein synthesis, to the ribosome for peptide bond formation, where the amino acids are bonded together in an order specified by the mRNA template. Within the ribosomes, there is a third type of RNA, ribosomal RNA (rRNA), which has an as yet unknown role in protein synthesis.

Proteins are constructed from free amino acids present in the cell cytoplasm by a process which follows a template directly from the DNA of the organism. Most organisms cannot construct a number of amino acids; these essential amino acids must be imported from an external source (the food of the organism). Thus, the analysis of the stable isotopes of amino acids from proteins can also reveal something of the environment in which an amino acid formed.

1.4.3 Mutation of the DNA molecule

Mutation, the change by substitution, deletion or addition of nucleotide bases, can occur at any time during the replication of the DNA molecule. However, in order for any mutated molecule to survive and be replicated, it must cause no disadvantage to the living system. If any disadvantage is created, then the molecule is less likely to be replicated, and the mutation will be lost.

The non-direct correlation between genes and their expressed amino acid sequences due to the introns and exons provides a possible scenario for the mechanism of relatively rapid mutations. If an intron were misread as an exon, a 'silent' area of DNA could be expressed, and a new protein result. If this is advantageous to the organism, the trait will be passed on.

For the mutation of a DNA molecule to be passed from parent to progeny, the mutation must take place within the reproductive cells (gametes) of the parent. During reproduction, the information carried in the sperm and egg combine, and replication of the DNA and cell division begins. A mutation in either gamete will give rise to different proteins in the progeny by the expression of the DNA within the genes. If the mutation causes a significant disadvantage, then the organism is unlikely to survive to maturity and be able to pass the mutation on to a second generation. However, if the mutation gives rise to some kind of advantage to the progeny, then it is more likely that the mutation will be passed on to further generations. This is the implication in the Darwinian theory of evolution i.e. "the survival of the fittest", formulated without knowledge of the genetic basis of reproduction.

The difference in the sequence of nucleotide bases does not always give rise to a new species. Genetic fingerprinting (for an introduction, see Connor, 1988) utilises the difference in the nucleotide bases of individuals of the same species. Differences at the specific level and higher must be observed through the differences in the sequence of the protein, expressed by differences in the DNA. This molecular speciation is still subjective (see papers in Patterson, 1987) although constraints are slowly being added to improve the accuracy of the technique.

1.5 Previous Work

The presence of amino acids in fossils was first demonstrated by Abelson (1954). By a combination of ion exchange and paper chromatography, these were isolated and identified from fossils up to 360 Ma. As a guide to understanding organic remains found within the shell, a Recent clam was successively etched by controlled amounts of dilute acid, and tested for protein. Although not uniformly distributed, proteins were found throughout the shell. In every fossil analysed, amino acids were present in the micromole per gram range.

To check for "adventitious contaminants", i.e. molecules adhering to the shell

sourced from percolating groundwaters, $CaCO_3$ was precipitated in the presence of amino acids. The principal amino acids found in fossils had little or no affinity for $CaCO_3$, and hence it is unlikely that the molecules would be adsorbed from groundwater. Fossils from the same horizon were also analysed and contained quantitiatively and qualitatively different concentrations of amino acids, showing that there was no homogenisation of the molecules throughout the sediment.

In 1955, Abelson analysed shells of *Mercenaria mercenaria* from the Recent, Pleistocene and Miocene. He found that by the Pleistocene, the total amount of amino acid present within the shell was only 18% of that present in the Recent. By the Miocene, this had dropped to 2% (Abelson, 1955). The possibility of dating samples by their amino acid content (later developed as amino acid racemization dating), and palaeo-environmental analysis was also suggested. The thermal stabilities of the amino acids were further investigated, to provide insight into their survival potential over geological time. Ala, for example, is thermally very stable over long periods of time, whereas Ser is not (**Chapter Seven**).

Since these pioneering papers, the amino acid composition of an enormous range of fossils has been analysed, and it is clear that amino acids are present in both invertebrate and vertebrate fossils throughout the Phanerozoic (e.g. Akiyima, 1971; Wyckoff, 1972; Armstrong *et al.*, 1983; Jope, 1980; Curry, 1988). From the data produced in these further studies, the difference of both the concentration and the range of amino acids is indicative of the majority of amino acids being indigenous to the fossil.

Briggs (1961), using paper chromatography, completed the first analysis of amino acids and peptides from New Zealand fossils and rocks, and concluded that amino acids were only present in some strata, and that preserved peptides were only present within the fossil samples. This investigation did not include the area of the present study.

Investigation of brachiopod shell molecules was first carried out by Jope (e.g. 1967a, 1967b; 1977; 1979; 1980), who investigated the relationships between living and fossil articulate and inarticulate brachiopods back to the Silurian by the analysis of the proteins and amino acids contained within the intercrystalline organic matrix. These studies utilised both amino acid analysis and disc electrophoresis to differentiate between the proteins present in the genera under study. It was concluded that the ratios between amino acids preserved within the samples were likely to be taxonomic.

Phylogenetic information from fossils was firstly derived from an indirect method of protein analysis. de Jong *et al.* (1974) discovered antigenic properties in molecules extracted from the shells of both Recent and 70 Ma fossil cephalopods. The antibody produced in response to the organic extract of *Belemnitella* (Maastrichtian) completely reacted with shells of Recent *Sepia officinalis*, indicating that fossil, as well as Recent, calcified tissues contain organic material which retains antigenic properties. This cross-reactivity also indicates that identical structures (antigenic determinants) may have

been retained within the shell macromolecules over 70 Ma. Evolutionary trends within the cephalopods were also indicated by the degree of reactivity of antibodies and target molecules.

Degens *et al.* (1967) outlined a phylogenetic tree for the molluscs based on shell tissue variation in living specimens. It was identified that complex relationships existed between amino acids in the sample, and principal component analysis was used to investigate the multivariate distribution. Variation in these samples was contained within eight or nine factors, which were derived from the loading of the amino acids within the original protein. The work also identified the possibility of intraspecific variation of the amino acid content, which was attributed to environmental conditions. Taxonomic data for Recent and fossil planktonic foraminifera was obtained by King and Hare (1972), who reinforced the morphometric classification of the samples by Q-mode factor analysis of the samples.

Weiner *et al.* (1976), extracted fossil glycoproteins of discrete molecular weight from the soluble organic matrix of the 80 Ma mollusc *Scabrotrigonia thoracica.* These were compared to the extant *Neotrigonia morgaritacea* and a repeating amino acid subunit found in the organic extract of this material was present in the fossil glycoprotein.

Towe (1980) analysed trilobite cuticle and graptolite periderm and found only marginal quantities of peptide bound amino acids. This work demonstrated that excellent physical preservation of fossils is independent of biochemical preservation of intact proteins. Towe (1980) suggested that the best preserved biochemical samples would be found inside crystals, rather than around them, which continued the work of Towe and Thompson (1972), who showed that organic material trapped during the growth of the mineral phase would be protected from decay, by being isolated from the environment. Hydrolysis of the peptide bonds of proteins trapped in the shell will therefore only be possible by water which was trapped in the crystal with the organic molecule during biomineralisation or by that formed by the breakdown of other amino acids (e.g. hydroxy amino acids, Bada *et al.*, 1978).

There was, initially, some reluctance in believing how a significant amount of organic matter could be incorporated into the crystal. However, Nickl and Henisch (1969), had already demonstrated that calcite crystals grown on silica gels could incorporate some of the gel during growth, and that the entrapped gel retained the shape of the crystal after decalcification. Both water (Hudson, 1967) and organic matrix (Watabe, 1963) have been reported to be trapped within the calcium carbonate of the shell. Towe and Thompson (1972) showed that the trapped organic matrices and water within the shell are generally discontinuous. Towe (1980) therefore concluded by suggesting that the best chance of good biochemical preservation of proteins in fossils would be as intracrystalline molecules, i.e. molecules trapped in the shell during calcification, and remaining isolated since that time. Following this suggestion of Towe (1980), Collins *et al.* (1988) extracted

intracrystalline molecules from the secondary fibres of brachiopod shells and used immunological techniques to investigate the relationships of Recent articulate brachiopods. Antisera raised against three extant genera, and cross reacted with molecules from other genera, resulted in a new interpretation of taxonomic relationships within the order Terebratulida, reinforcing the traditional division of the order into long- and short-looped genera, but suggesting new assignments for genera which were traditionally difficult to assign.

Berman *et al.* (1988) grew calcite crystal crystals *in vitro* in the presence of both acidic glycoproteins extracted from within the mineralised hard parts of sea urchins and proteins from mollusc shell. Sea urchin proteins were selectively adsorbed onto the surface of the crystals and incorporated into the crystal with further growth. These synthetic crystals had conchoidal fracture analagous to the fracture of the calcite crystals within living sea urchins, rather than that of the normal fracture pattern for inorganic calcite crystals. From this, it was deduced that the material properties of sea urchin calcite was caused by the presence of intracrystalline proteins.

Sucov *et al.* (1987) made a DNA clone selected using an antibody produced in response to sea urchin embryo spicule matrix protein. The clone represented the mRNA code for the 50 kDa protein which forms part of the embryo spicule. From the sequence of nucleotide bases on the mRNA molecule, the primary sequence of the corresponding protein was derived. This, however, was an indirect sequence; the direct sequencing of intracrystalline proteins from sea urchins has yet to be published. Addadi *et al.* (1991) reported the partial amino acid compositions of intracrystalline proteins from the inorganic phase of sea urchins and molluscs.

Curry *et al.* (1991a) extracted and purified to homogeneity intracrystalline molecules from the shells of the Recent articulate brachiopod *Neothyris lenticularis*. From two of the purified proteins, partial sequence data and amino acid composition were acquired. This was the first sequence data to be sourced directly from intracrystalline shell proteins, and neither matched any previously known protein sequence. This initial study was followed by a more detailed investigation (Cusack *et al.*, 1992) of a small (6.5 kDa) intracrystalline protein from several genera of brachiopods. *N*- terminal sequencing of the molecule showed that, although there were some changes in the residues between genera, the protein was homologous, and there was no need to insert gaps into the sequence. Cusack *et al.* (1992) also found a prosthetic carotenoid group attached to this protein, which gives rise to the red colouration of the shell, indicating a function for at least one of these intracrystalline proteins.

Geochemical information available from the study of the amino acid racemisation reaction (Schroeder and Bada, 1976; Bada and Man, 1980) was investigated by Hare and Mitterer (1967), who carried out the first investigation into the racemisation of L-isoleucine in Recent to Miocene shells. Following on from this, Hare and Abelson

(1968) reported that eight of the amino acids found in fossils reach equilibrium (D- form ≈ L-form) by the Miocene. The use of the racemisation reaction in the dating of Plio-Pleistocene fossils and sediments has been well documented, although the use of the reaction in the assessment of organic contamination, is a more recent innovation.

1.6 Statistical techniques used to manipulate biochemical data

1.6.1 Numerical Taxonomy

Numerical taxonomy was described by Sneath and Sokal (1973, page 4) as *"the grouping by numerical methods of taxonomic units into taxa on the basis of their character states"*. Numerical methods can be interpreted as any kind of mathematical data used in the investigation of samples, ranging from simple morphometric measurements (length, width, height) through to derived variables from multivariate statistical analyses. The term "numerical taxonomy" was used in preference to "quantitative taxonomy", as this implied the use of analytical chemistry and/or biology, such as serology and chromatography (Sneath and Sokal, 1973). The name "numerical taxonomy" has been retained in this study of molecular palaeontology, as it represents an extension to morphometric numerical analysis, rather than a new class of technique. Collins *et al.* (1988) utilised cluster analysis of serological techniques to investigate brachiopod taxonomy in the Terebratulides. Curry *et al.* (1991b), following Walton and Curry (1991), used graphical plots of principal component analysis (PCA) of biochemical data to compare the amino acid compositions of fossil and Recent brachiopods.

Among the advantages outlined by Sneath and Sokal (1973) for numerical taxonomy were the notions of repeatability and objectivity. Use of amino acid concentrations and distributions in fossils provides reproducability in analyses and taxonomic groupings. Objectivity is closely linked with repeatability, and is defined (Oxford English Dictionary) as *"impartial judgement; the ability to be free from personal prejudice"*. In many types of morphometric analysis this is rarely fully implemented. However, the order of amino acids in proteins present within the shell is derived directly from a template sourced from the DNA, and is therefore independent of the morphology of the organism and free from prejudice from interpretations of ecophenotypic variation; i.e. the amino acid data comes from the genotype of the organism, not from the phenotype (the expression of the genotype).

Delimitation of taxonomic groups is therefore an objective process, although the definition of boundaries between the taxa will be arbitrary, unless some pre-arranged system for this is in use. Boundary definition must therefore *"be based on comparable criteria in all regions of the taxonomic space under consideration"* (Sneath and Sokal, 1973, page 7).

Data interpretation of the amino acids extracted from fossil samples becomes increasingly complex as the number of samples increases. Several studies of amino acids

from fossils have utilised molecular percentages (Jope, 1967b) and ratios between amino acids (Kolesnikov and Prosorovskaya, 1986) to establish taxonomic relationships. Statistical techniques were first applied to amino acid data by Degens *et al.* (1967), who used PCA in a phylogenetic study of the molecules from living mollusc shells. However, data was expressed only as factor scores, making interpretation of the results more difficult. Graphic presentation of Q-mode factor analysis was used by King and Hare, (1972) to establish taxonomic similarities between the shell molecules of Recent and fossil planktonic foraminifera.

Following the suggestion of Rohlf (1968), data has been analysed by both Principal Component Analysis and Cluster Analysis to extract as much information on the relationships as possible. Both of these types of multivariate analysis are able to convert the amino acid data into a graphical display of differences. Multivariate analysis is a relatively simple process when comparing small numbers with bivariate distribution. However, as both the sample and variable size increases, the calculations required become more and more involved. It is for this reason that the statistical package DATADESK™ for the Macintosh micro-computer was used. Brief descriptions of both methods are outlined below, but further details may be gained from Koch and Link (1971), Davis (1973), and Sneath and Sokal (1973).

1.6.2 Principal Component Analysis

This method utilises the variance of the dataset, within the samples, to recalculate the values and summarise the information. This recalculation is necessary as it is not possible to represent, by conventional means, a set of variables with more than three characteristics.

Data on the distribution of amino acids through samples can be tabulated into a matrix, which represent co-ordinates of points which may be plotted in *n*-dimensional space; hence a two by two matrix (four numbers) can be represented in two dimensions (i.e. on a sheet of paper). The co-ordinates also represent points on the boundary of an ellipse which includes all the points, with its centre at the origin of the co-ordinate system. The seventeen amino acids would normally have to be plotted in seventeen dimensional space, but by using PCA, the values may be recalculated so that a high percentage of the total variance between the samples can be expressed in three dimensions (i.e. X-Y-Z space). In some cases, this may be extended further, so that data can be shown on a scatterplot (two dimensions). The principal components for each sample are determined by the variation of amino acid concentrations within that sample, and hence are independent of the other samples. Principal components can therefore be extracted from one data set and compared to those of others, in order to ascertain those amino acids which influence distribution.

From moles per gram measurements of the amino acid content of a sample, a matrix of the variance/covariance relationships between the samples and the variables may

be calculated. From the variance, the determinants of the matrix are calculated and placed in polynomial simultaneous equations. The solutions to these equations are **eigenvalues**, which represent the amount of variance along a corresponding axis, calculated by substituting back into their equations to extract the vector of their coefficients, **eigenvectors** (or principal vectors), associated with the eigenvalues. There are as many eigenvectors as there are eigenvalues. The eigenvector may be thought of as the slope of an axis of the ellipse which encloses the data, whose length is defined by the corresponding eigenvalue. The term "principal components" simply refers to these eigenvectors. The first principal axis accounts for the greatest amount of variance in the sample, and the remaining axes contain decreasing amounts of the total variation.

Sneath and Sokal (1973, p.246) note that it is usual *"to extract only enough eigenvectors to remove the majority, say 75%, of the total variance of the data matrix"*. From computer calculations, it can be seen that the majority of the variance within the samples can be defined by the first three eigenvectors, and hence the data may be expressed on a rotating plot produced on the computer. This representation of the amino acids in PCA form in three dimensional space is a useful method of comparing multivariate distributions of large sample size (**Chapter Eight**).

1.6.3 Cluster Analysis

Cluster analysis results in the two-dimensional expression of multi-dimensional characters, and is demonstrated by the use of dendrograms. Clustering of the data results in the partitioning of data on the basis of similarities and differences between characters, in this case the molecular percentage of amino acids within a sample. The analysis produces clusters of groups of similar nature.

As in PCA, the grouping of data is derived from the expression of the data in matrix format. The variables are assessed for similarities, and grouped accordingly. In this study, single linkage cluster analysis was used, which is equivalent to the nearest neighbour analysis of Lance and Williams (1968) and also to the minimum method of Johnson (1967).

The method of clustering is based on the relationships of the amino acids to each other. The first grouping in the cluster is based on the two closest data sets. As clustering proceeds, the sample characteristics are compared to the closest single sample within the cluster, and the similarity of the sample to the cluster is the same as that between the sample and its closest member within that cluster. This also holds true for relationships between clusters, whereby the similarity between the two clusters is the same as that between the two closest samples within the clusters.

Chapter Two: Geological introduction

Geological Introduction

2.1 Introduction

In order to assess the effect of time on the state of preservation of molecules contained within fossils, samples need to be collected from a relatively young stratigraphic series with well dated horizons which contain fossil species with extant representatives. The existence of such extant samples is necessary in order to give a starting point or time zero for the molecular data, the inference being that the fossil species began with a similar molecular composition as the Recent sample, and that the majority of changes in the amino acid composition of the fossil of the same species will be due to the breakdown of the original molecules. This also means that comparisons will be possible within and between horizons and also over specific time periods.

The South Wanganui Basin in North Island, New Zealand (**figure 2.1**) con-

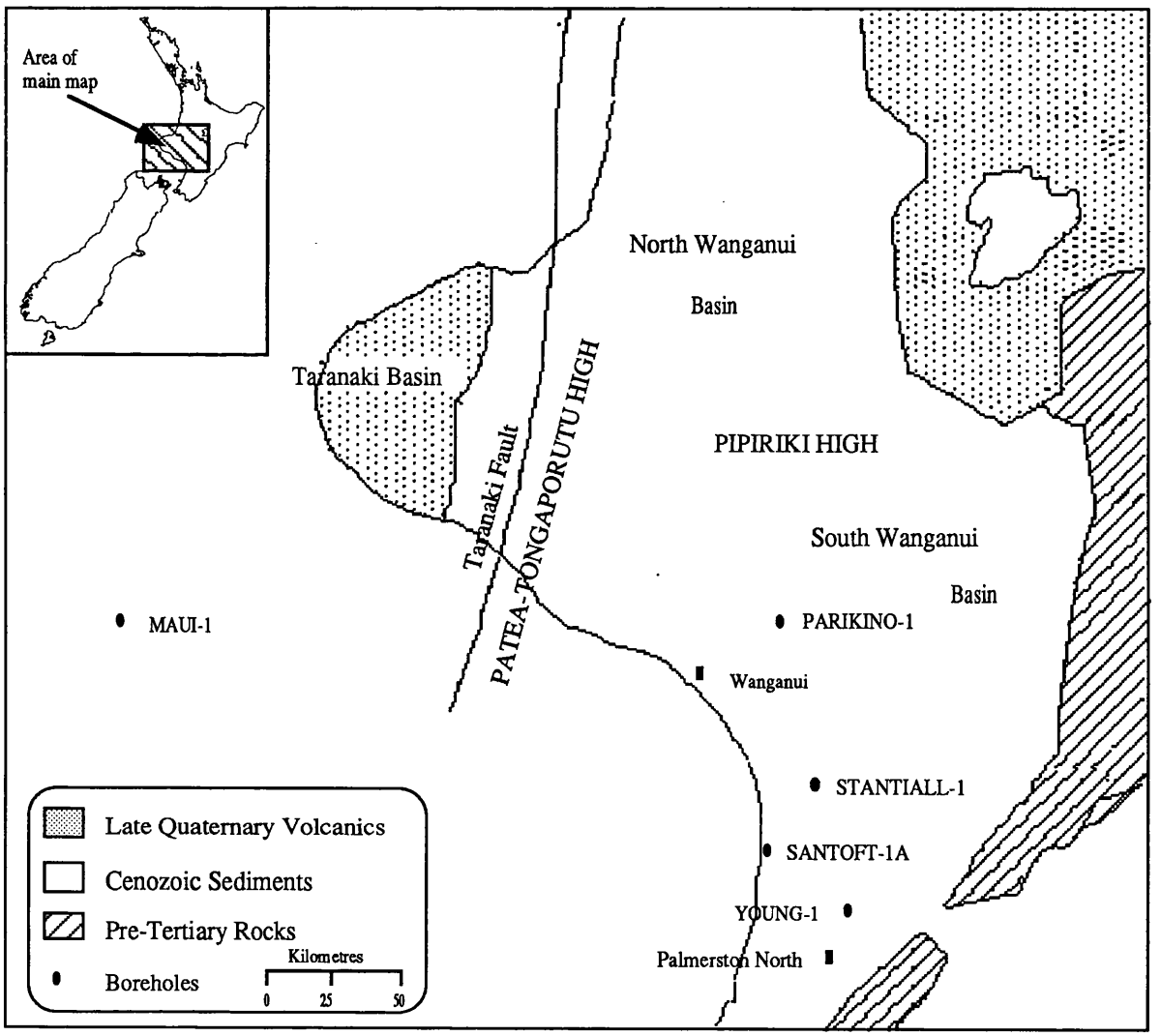


Figure 2.1: Location of the South Wanganui Basin, showing the site of the boreholes referred to in the stratigraphic description.

forms to the above criteria (**section 2.3**). Samples collected from the sediments of the Basin have extant representatives which are easily collected elsewhere in New Zealand. Two periods of fieldwork were undertaken by the author in the early parts of 1990 and 1991, sampling the beds indicated in **figures 2.2** and **2.3**. Collection of extant representatives was from sites throughout New Zealand (**figure 2.4**). This chapter outlines the geological and biochemical significance of the Brachiopoda, and describes in detail the beds from which samples were collected. The geological history of the samples is also considered.

2.2 Geological and biochemical significance of the Brachiopoda

The Phylum Brachiopoda comprises two classes of invertebrate, marine, sessile filter feeders, which are diverse and abundant throughout the Phanerozoic, and is composed of some 2000 genera and 30 000 species. The distinction between the classes is based partially on the shell mineralogy (the Inarticulata have chitino-phosphatic shells, whereas those of the Articulata are calcitic) and partially on morphological characters (e.g. the hinge and lophophore support; Williams and Rowell, 1965b). The superfamily Craniacea, whilst being included taxonomically in the Inarticulata, has a calcitic shell, and the shell proteins have characteristics of both classes (Jope, 1977).

The Phylum has a geological record which spans the Phanerozoic. The inarticulate species *Lingula* has remained morphometrically unchanged since the Silurian (?Ordovician) and is one of the prime examples of evolutionary stasis. By contrast, the Articulata are a diverse class which at several times during the Phanerozoic have been the dominant constituent of marine faunas. Although the phylum is declining from its acme, mainly due to the extinction event at the Permian-Triassic boundary (the Cretaceous-Tertiary event had comparatively little effect), there remains a diverse extant population of 69 genera (Rudwick, 1970) which is overwhelmed by the dominant Phylum Mollusca. The phylum inhabits ecological niches from intertidal to 6000 m below sea level and from the tropics to the poles. They therefore represent a group with enormous potential for present day ecological study which may be extrapolated over geological time via their fossil record.

Brachiopods have long been used as biostratigraphic tools, as the rapid evolution of some genera coupled with mass extinction enables correlation and dating of sedimentary strata vital for the reconstruction of the globe in the past. The stability of the brachiopod shell has also long been recognised. Isotopic variations of strontium in Permo-Carboniferous brachiopods has confirmed the general shape of the strontium isotope age curve for this time period (Popp *et al.*, 1986).

The shell of the brachiopod is similar to that of other 'shellfish' in that it consists of organic layers secreted by the outer epithelial layer of the body, strengthened

Grid references of sample localities

(1)	Rapanui Fm, ?Castlecliff Beach	N137/457910
(2)	Rapanui Fm, Verry's Farm, Waitotara	N137/262 031
(3)	Rapanui Fm, Waipipi Beach	N137/168 993
(4)	Landguard Sand, Landguard Bluff	N138/558 836
(5)	Upper Castlecliff Shellbed, Castlecliff Beach	N137/492 884
(6)	Tainui Shellbed, Castlecliff Beach	N137/485 888
(7)	Pinnacle sand, Castlecliff Beach	N137/479 895
(8)	Lower Castlecliff Shellbed, Castlecliff Beach	N137/470 902
(9)	Lower Castlecliff Shellbed, Waipuka Road	N138/705 871
(10)	Kupe Fm, Castlecliff Beach	N137/459 908
(11)	Kaimatira Pumice Sand, Kai-iwi Beach Road	N137/442 929
(12)	Okehu Shell Grit, Okehu Stream mouth	N137/418 932
(13)	Tewkesbury Fm, Turakina Valley	N138/911 922
(14)	Waipuru Shellbed, Turakina Valley	N138/919 933
(15)	Shellbed 4243, Wanganui River Road	N138/661 991
(16)	Nukumaru Brown Sand, Great Northwestern Road	N137/382 997
(17)	Shellbed 4357, Parapara Road	N138/803 029
(18)	Hautawa Shellbed, Parapara Road	N138/723 033
(19)	Te Rama Shellbed, Rangitatu West Road	N137/416 069
(20)	Upper Waipipi Shellbed, Waipipi Beach	N137/170 992
(21)	Middle Waipipi Shellbed, Waipipi Beach	N137/165 993

Map references refer to the maps accompanying Fleming (1953)
Localities (13) and (14) not shown on map opposite

Correlation	Biostratigraphic classification			Lithological classification			Collection data	Age
	Series	Stage	Substage	Group	Formation	Member		
Boundary Uncertain	Recent			Recent	Recent beach sand Patea Dune-sand Swamp deposits Volcanic ash beds Undifferentiated recent alluvium	Egmont Ash	1 2 3 4 5 6 7 8	
Pleistocene	Hawera			Ponaki	Papaiti Alluvium St Johns Alluvium Undifferentiated Alluvium Rapanui Formation	Rapanui Dune-sand and Kaiwhara Alluvium Rapanui Lignite Waipuna Delta Conglomerate and Rapanui Marine Sand	1 1 1 2 3 4 5 7 8	0.2 Ma
					Brunswick Formation	Brunswick Dune-sand Fordell Ash Brunswick Alluvium Brunswick Beach Sand Brunswick Pebbly Sand		
					Kaiatea Formation	Dune-sand, ash and alluvial members		
	Wanganui	Castlecliffian	Putikian	Shakespeare	Landguard Formation Undifferentiated formations Putiki Shellbed Mosstown Sand Karaka Siltstone Upper Castlecliff Shellbed Shakespeare Cliff Sand Shakespeare Cliff Siltstone Tainui Shellbed	Landguard Sand Basal Conglomerate *Tawera* lenticles Basal shell conglomerate 	1 2 7 8 1 2 3 1 1 2 3 4 5 7 8	0.35 Ma 0.4 Ma
					Pinnacle Sand Lower Castlecliff Shellbed Seafield Sand	 Toms Conglomerate	1 2 3 1 2 3 4 7	
				Kai-iwi	Upper Kai-iwi Siltstone Kupe Formation	 Pecten Layer Pelecypod Shellbed Member *Mactra tristis* layer Gastropod shellbed member Cross-bedded sand member	1 1 2 3 4 5 7 8	 0.5 Ma
			Okehuan		Upper Westmere Siltstone Lower Westmere Siltstone Omapu Shellbed Lower Kai-iwi Siltstone Kaimatira Pumice Sand	Kaikokopu Shell Grit Basal sand member (local members)	1 4 5 7 1 1 1 1 3 4 7	0.55 Ma 0.65 Ma 0.73 Ma
				Okehu	Upper Okehu Siltstone Okehu Shell Grit Lower Okehu Siltstone Butlers Shell Conglomerate Makirikiri Tuff	Basal shell conglomerate Basal shell conglomerate Ototoka Siltstone Tongue Conglomerate members	1 1 4 1 1	 0.8 Ma 0.85 Ma 0.95 Ma
Boundary Uncertain				Maxwell	Upper Maxwell Formation Mangahou Siltstone Middle Maxwell Formation Pukekiwi Shell Sand Lower Maxwell Formation Tewkesbury Formation	Lignite members Lignite members *Lignite members* Shell sand and conglomerate lenticles	 1 1 1 3 7	1 Ma 1.26 ± 0.17 Ma 1.65 Ma 1.67 Ma
Pliocene		Nukumaruan	Marahauan	Nukumaru	Waipuru Shellbed Undifferentiated Shellbeds Nukumaru Brown Sand Mangamako Shellbed Nukumaru Formation Ohingaiti Sand	 Nukumaru Limestone	1 2 3 7 8 1 2 3 7 1 2 3 1 3	 1.85 Ma
			Hautawan	Okiwa	Undifferentiated Formations Kuranui Limestone Hautawa Shellbed		1 2 3 5 7 8 1 2 3 5	2.15 Ma 2.2 Ma
		Mangapanian			Te Rama Shellbed Parihauhau Shellbed Undifferentiated Formations Te Rimu Sand		1 2	
				Paparangi	Wilkies Shellbed Makokako Sand Mangaweka Mudstone Paparangi Sandstone	 Mangamahu Concretionary member Mangapani Shell Conglomerate	1 1	 2.4 Ma
		Waipipian		Whenuakura	Waverley Formation Waipipi Formation Rangikura Sandstone Pepper Shell Sand	 Upper Waipipi Shellbed Middle Waipipi Shellbed Lower Waipipi Shellbed Snapper Point Shellbed Snapper Point Conglomerate Bremer Shell Grit Tuke Shell Sand Basal sand Member	1 1 2 4 7 8 1 2 7 8 1 1	 3.3 Ma

Key:-
1 = sediment sample
2 = *Neothyris sp.*
3 = *Waltonia sp.*
4 = *Terebratella sanguinea*
5 = *Notosaria nigricans*
6 = *Liothyrella sp.*
7 = Turatellid
8 = *Pecten sp.*

Figure 2.3: Stratigraphic nomenclature for the South Wanganui Basin, giving radiometic/fission track dates (e.g. Beu and Edwards, 1984; Seward, 1974) and details of the samples collected from each horizon (after Fleming, 1953).

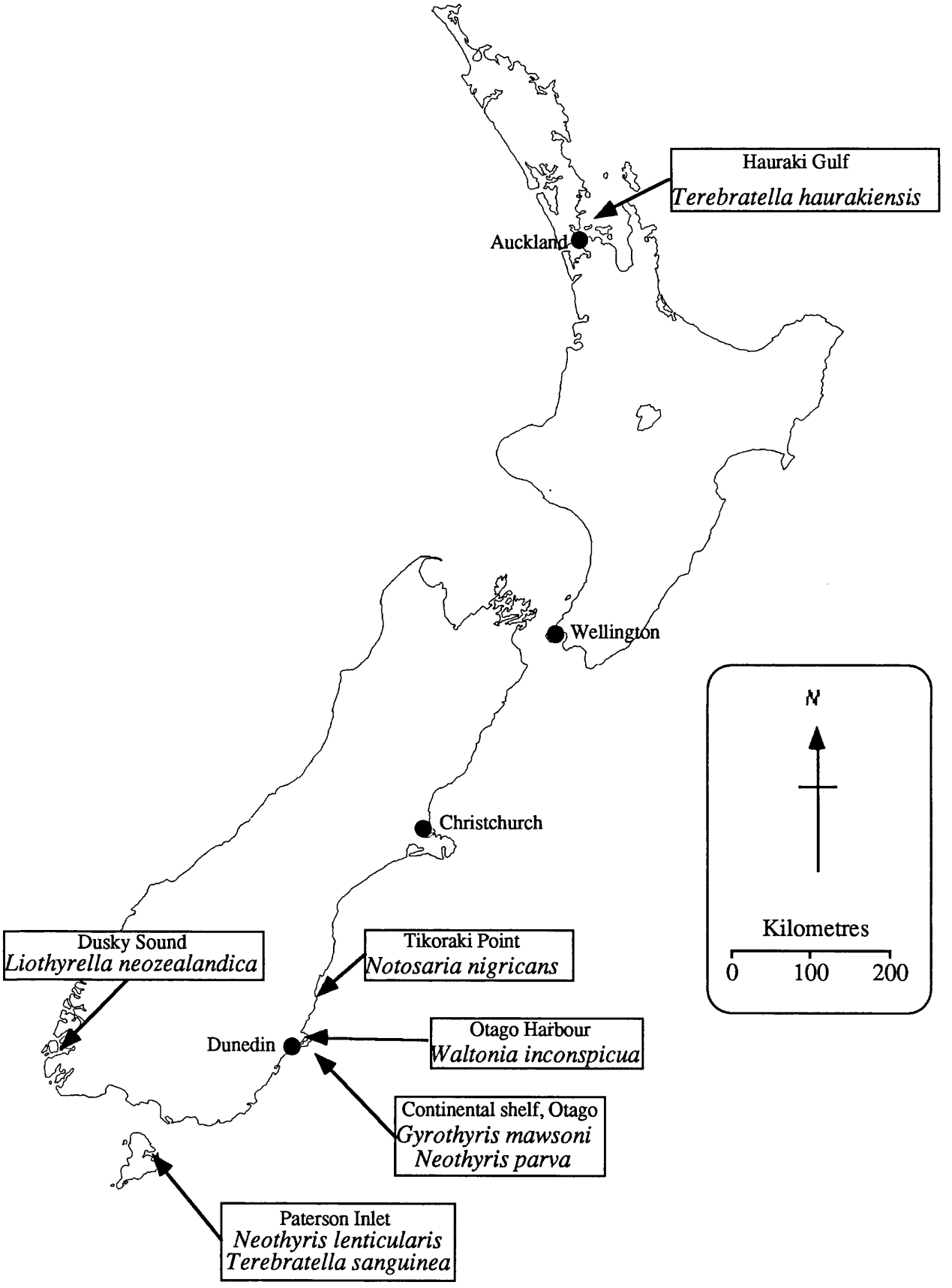


Figure 2.4: Location of collection sites for extant representatives used in the analyses.

by inorganic components. The organic matter within the shell is mainly composed of protein (Jope, 1965), although lipids (Curry *et al.*, 1991b) and carbohydrates (Collins *et al.*, 1991a) are also present in small quantities.

In the Inarticulata, the shell is composed of calcium phosphate, together with 25 - 52% organic matter, mainly in the form of chitin (an amino-sugar polymer) and protein (Jope, 1965). The Articulata, by contrast, have shells composed of calcium carbonate together with 0.93 - 4.7% organic matter, mainly composed of protein (Jope, 1965). The composition of the shells of the Articulata, in contrast to the meta-stable aragonitic shells of the Mollusca, provides a stable environment for the molecules over millions of years. The protein may be present in two localities within the shell (**figure 2.5**). The molecules from the proteinaceous sheath present **between** shell fibres are referred to as *inter*crystalline molecules, whereas those **within** the inorganic phase are known as *intra*crystalline

Figure 2.5: Scanning electron microscope photograph of the secondary layer fibres of brachiopod shells. Each fibre is c. 1 μm wide. Surrounding each of the fibres is the intercrystalline protein matrix, and within each of the crystallites there is also some intracrystalline organic matter. In the investigation, the intercrystalline molecules were routinely destroyed by incubation with sodium hypochlorite. Photo reproduced by kind permission of Sir Alwyn Williams.

molecules.

Organic molecules become incorporated in the shell during biomineralisation, the process by which organisms precipitate inorganic phases (i.e. by which they form minerals; Lowenstam and Weiner, 1989). The process can either take place in a situation whereby the mechanisms of biomineralisation are **specifically** setup for that task ("biologically controlled mineralization"; Mann, 1983), or whereby the process is not designed for that task, but by which minerals may be formed ("biologically induced mineralization"; Lowenstam, 1981). The actual process is, as yet, unconstrained, although a fairly limited number of different basic processes are thought to be involved. These are reviewed in Lowenstam and Weiner (1989) and are not considered further here. The brachiopod shell is secreted by the outer epithelial cells of the body wall of the organism. These cells act as components of a biologically controlled mineralisation system, the onset of which, in both the Inarticulata and the Articulata, is characterised by the secretion of the periostracum (which is not preserved in fossils) as a continuous, exclusively protein, relatively impermeable outer cover to the shell (Jope, 1967a).

In the Articulata, the periostracum is underlain by the calcium carbonate of the shell. The cells of the outer epithelium firstly deposit a layer of compacted acicular crystallites (primary layer), but as the shell thickens, the nature of the cell changes (MacKinnon and Williams, 1974), and protein and calcite is secreted simultaneously, to form a layer of alternating rows of calcite fibres with a proteinaceous matrix between them (secondary layer; **figure 2.5**). Although the fibres are not crystallographic prisms, they are deposited by one cell as it migrates away from the outer layer of the shell (Williams, 1968), and therefore acts as a single crystal. In some genera, the secondary fibres become interspersed with lenses of prismatic calcite which marks the transition between the secondary and tertiary layers of the shell. The boundary is marked by the cessation of the deposition of protein. The tertiary layer consists of large prisms of calcite.

In the Inarticulata, the periostracum overlies alternating laminae of chitinous protein and calcium phosphatic layers. These laminae are subparallel to the outer layer of the shells, with the phosphatic bands often being relatively thick in areas of the shell covering the body cavity, but which thin out laterally and anteriorally. The outer epithelium of the Inarticulata is divided into areas roughly parallel to the margins of the valves. Areas which deposit calcium phosphate are interspaced between those depositing the organic layer (Williams and Rowell, 1965b).

Recent articulate brachiopods contain up to 50% of their tissue mass within the shell (Curry and Ansell, 1986; Curry *et al.*, 1989), a figure which is higher than that reported for molluscs. Although much of this shell protein is derived from the caeca, and hence will not be present within fossil shells, a significant proportion will be located in the secondary layers, between the shell fibres . It is this intercrystalline source that provided the proteins analysed by Jope (1967a; 1967b; 1969; 1977; 1980) and Kolesnikov and

Prosvoskya (1986).

Brachiopods also contain intracrystalline protein, trapped within the calcium carbonate during biomineralisation. Following the suggestion of Towe (1980), Collins *et al.* (1988) confirmed the existence of intracrystalline proteins in secondary fibres from Recent articulate brachiopods using immunology (**figure 2.6**). Recently, Curry *et al.* (1991a; **Appendix 2.1**) extracted several intracrystalline proteins from the shells of Recent *Neothyris lenticularis* (Deshayes), purified them to homogeneity using gel electrophoresis

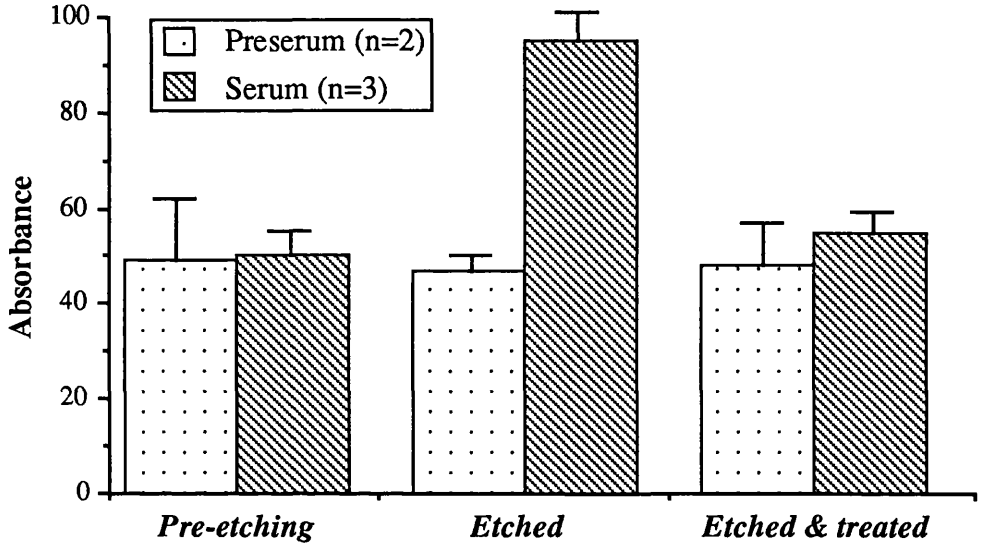


Figure 2.6: Confirmation of the presence of intracrystalline molecules with an immune response. Results for the pre-serum represent the background levels of absorbance of the analysis system. High levels of absorbance in the etched sample indicates the presence of immunologically active compounds within the shell. Treatment with bleach destroys the molecules, resulting in the lowering of response (after Collins *et al.*, 1988).

and obtained primary sequence data which did not correspond to any known sequence.

These intracrystalline molecules are important in molecular palaeontology, as the inorganic phase seals the molecule in a confined microenvironment, ensuring that no leaching of molecules can take place, and that all degradation products remain within the crystal, in a position from where they may be released and characterised. In a preliminary study of fossil brachiopods from this New Zealand sequence, Curry *et al.* (1991b; **Appendix 2.2**) found that taxonomic information is preserved in fossils of 0.5 Ma, and confirmed that the breakdown of intracrystalline proteins takes place *in situ.*

2.3 Stratigraphy

Samples were collected from the suite of Plio-Pleistocene sediments deposited in the South Wanganui Basin, which outcrops partly onshore in the South West of North Island, New Zealand(**figure 2.1**).

All of the strata in the following description vary in thickness throughout their outcrop, with all sediments becoming thicker towards the eastern part of the basin. Thicknesses for each of the groups will be given as either that on the coast, or that slightly inland, unless there are major lateral variations in thickness. Names and stratigraphic relationships can be seen from **figures 2.2 and 2.3**. The lithological descriptions are modified from Fleming (1953).

2.3.1 Wanganui Series

Throughout the Wanganui Series, the rocks were deposited mostly under marine conditions of varying depths. The fluctuations of the sea level are due to glacio-eustatic changes discussed in the following section.

(a) Whenuakura group

This group is composed of thick featureless sediments, composed of marine sands and silts, interspersed with shellbeds, some of which are upwards of one metre thick, shelly lenses and concretionary horizons. The total thickness of the sediments is approximately 140 m.

The conditions of deposition for the Whenuakura group are as follows. The basal unit, the Pepper Shell Sand, was deposited in shallow marine conditions, below low tide, but includes one intertidal species of mollusc. This is followed by shallowing events, leading to estuarine conditions of the Rangikura sandstone, and subaerial exposure, leading to the deposition of beach sand and local shell conglomerates (Bremer Shell Grit). A renewed marine transgression submerged a second beach deposit (Snapper Point Conglomerate) to below low tide level, and the accumulated sands and silts of the Waipipi Formation were deposited. Scouring of the deposits in this area gave rise to the first in a series of rich fossiliferous shellbeds sampled in this study. These are found throughout the Wanganuian, and probably offshore in the Haweran and Recent series.

The formation of most of these shellbeds are thought to follow a similar pattern. The dead shells would accumulate offshore at a depth of several metres below the low tide level, and probably below the storm wave base. Currents in this area lead to the winnowing out and removal of the fine sediment, allowing the development of shelly mound such as those found offshore in Taranaki (Norris and Grant-Taylor, 1989). It is likely that the accumulation of shells in this manner would not take place rapidly, but would accumulate over a period of several thousands of years. Such shellbeds therefore represent a hiatus in the deposition of sediment.

Four main periods of shellbed formation (Snapper Point and Lower, Middle and Upper Waipipi Shellbeds) took place during the deposition of the Whenuakura, with the intervening deposits being fine sands and silts. Towards the end of this deposition, a regressive event began, resulting in the deposition of the Waverley Formation, a series of

shallow marine and intertidal sands, which includes some shell lenses close to the top.

(b) Paparangi Group

This group is characterised by the development of thick sandstone units, containing rare shellbed and conglomeratic members. Total thickness of the group is in the region of 300 m on the coast, and 600 m to the east of the basin.

The basal unit, the Mangapani Shell Conglomerate, was deposited on a beach after an erosion interval wherein underlying sediments were eroded in the intertidal zone. This erosion event was followed by a marine transgression associated with rapid subsidence, which gave rise to the Paparangi Sandstone, a thick (250 m+) barren sandstone, which grades into better sorted, coarser, looser sand near the top, including commoner shellbed members (Makokako Sand), indicating a smaller sediment supply. The topmost unit of the Paparangi Group consists of the Wilkies Shellbed, deposited in shallower water, in the same manner as those of the Waipipi Formation, and containing fossils characterisitic of the intertidal zone. The sand units of the Paparangi Sandstone grade into muds in the east of the Wanganui Basin, and the laterally equivalent Mangaweka Mudstone (offshore) was deposited.

(c) Okiwa Group

The Okiwa Group consists of sands, muddy sandstones and shellbed members split into two unequal parts by the Kuranui Limestone and its lateral equivalent in the east, the Upokonui Sand. The Group contains large thicknesses of sediments which are undifferentiated, and interbedded within named formations. Total thickness of the group is approximately 350 m near to the coast.

The basal unit of the Okiwa Group is the Te Rimu Sand, which is a porous, moderately well cemented sediment, deposited in offshore, shallow marine conditions. Conformably above the Te Rimu Sand lie the Undifferentiated Formations of the Lower Okiwa Group, which represent a poorly defined sequence of interbedded estuarine and shallow marine sediments, but which contain the Te Rama Shellbed in the west and the Parihauhau shellbed in the east. In the east, these Formations are followed by the Hautawa Shellbed, a laterally extensive and consistent horizon, cropping out over some 30+ km. Further west, the shellbed is harder to define, and may be truncated beneath formations of the Upper Okiwa Group. The shellbed was deposited in shallow offshore conditions, at a depth great enough not to include any littoral organisms in the fossiliferous beds.

The Kuranui Limestone is a coarse, pebbly, shell limestone of variable thickness deposited as a littoral/sublittoral sand formed by a renewed marine transgression, after a period of exposure had resulted in the erosion of part of the Hautawa Shellbed. The Kuranui Limestone grades into the Upokonui Sand to the east. The Tuha Sand, a poorly cemented, fine grained, micaceous sand, lies above the Upokonui Sand in the east of the basin, but not in the area covered by this study. The named formations of the Upper

Okiwa, including the Hautawa Shellbed are interbedded with Undifferentiated Formations, which were deposited under shallow marine conditions.

(d) Nukumaru Group

The uppermost muddy sandstones of the Okiwa Group are succeeded by a series of sandstones and lenticular limestones of the Nukumaru Group. The total thickness of the deposits is around 250 m, although there is some confusion about the exact thickness as the stratigraphic relationships are obscured by the Nukumaru Fault Zone, which cuts through the deposits of the Nukumaru Group.

The basal Ohingaiti Sand was deposited under shallow marine conditions, the free sand deposit marking a change of facies from the Upper Okiwa Group. The sand deposit passes upwards into a series of shell limestones, interbedded with sand units, representing the Nukumaru Formation (Nukumaru Limestone). There are five limestone members within the formation, some of which contain cross stratification, pebbles and intertidal fossils, indicating that some of the deposit must have occurred in the nearshore/ intertidal area. The limestones are composed of the broken shells of marine and semi-estuarine organisms, concentrated by tidal scour. As they are interbedded with cross stratified sands, it is likely the limestone lenticles formed as beach and intertidal shellbeds. The limestones grade into the rusty coloured sands of the Nukumaru Brown Sand, a deposit representing a mixture of subtidal conditions.

The upper part of the Nukumaru Group in the east of the basin is characterised by thicknesses of sandstones and mudstones which have not been differentiated. However, interspersed within these deposits are several shellbed members from which collections were made, and which are thought to have originated in shallow marine conditions in the same way as other shellbeds. The Waipuru shellbed is the topmost unit of the Nukumaru group, and is a muddy sand deposit, which is richly fossiliferous and which extends from the east across most of the basin.

(e) Maxwell Group

The Maxwell Group consists of a series of dominantly non-marine silts and sands, with some lignite members, and has an approximate thickness of c. 200 m.

The basal unit of the Maxwell Group is the Tewkesbury Formation, a deposit of alternating fine sands and muds, representing estuarine conditions. The shellbeds in the formation are derived from deposits on beaches and in tidal scour channels, but which are close enough to the shore to allow the incorporation of Moa bones and wood. The Tewkesbury Formation passes upwards into the Lower Maxwell Formation, a lagoonal and river delta deposit which includes lignite members, indicating progradation, probably during a regressive event. The Pukekiwi Shell Sand, which lies conformably on the underlying beds, is a muddy sandstone containing marine molluscs characteristic of the subtidal regime, but is thought to represent only a local marine advance.

The Middle Maxwell Formation, and also the Upper Maxwell Formation, represents a return to similar conditions of deposition as in the Lower Maxwell. The Mangahou Siltstone lies between the Middle and Upper Maxwell Formations is a fine grained, well bedded, pumaceous and fossiliferous siltstone, thought to have been deposited on a coastal mudflat.

(f) Okehu Group

The Okehu Group is a series of silts and shell grits representing the return to marine conditions, after the predominantly non-marine conditions of the Maxwell Group. The Group has an overall thickness of around 40 m.

The basal formation, the Butlers Shell Conglomerate, is a shallow water shell conglomerate including ash layers and quartzite pebbles. In the east of the basin, the shell conglomerate is completely replaced by volcanogenic sediments of the Makirikiri Tuff, a white siltstone composed of rhyolitic glass shards. The deposition of these formations followed the development of an unconformity across most of the basin. The Upper and Lower Okehu Siltstone are sands and clays deposited offshore in relatively deep water, although both have prominent basal conglomerates, caused by a regressive event, prior to submergence by a subsequent sea level rise. The Okehu Shell Grit lies between the Okehu Siltstones, and occupies lenticular depressions scoured in the Lower Okehu Siltstone and consists of cross stratified shell grit together with muddy sand. The deposition of this formation followed a period of emergence of the Lower Okehu Siltstone into the tidal zone.

(g) Kai-iwi Group

The base of the Kai-iwi Group is marked by the reappearance of large amounts of volcanogenic material, absent from the Upper Okehu Group. The deposits are pumaceous sands and silts, entirely marine, and all formations are fossiliferous. The total thickness of the Group is c. 75 m on the coast and 550 m in the east of the basin, where formations are undifferentiated.

The basal unit of the Group, the Kaimatira Pumice Sand, was deposited in shallow water, in or near the tidal zone, probably on a marine/estuarine delta. Locally, underlying beds are scoured, representing the elevation of the deposits and erosion prior to deposition of the Kaimatira Pumice Sand. Rapid subsidence submerged the Kaimatira Pumice Sand and the Lower Kai-iwi Siltstone deposited conformably, as a sequence of fine silts offshore. The Omapu Shellbed represents the accumulation of shells in one of the typical associations of the Wanganuian. The water depth increased, and the Lower Westmere Siltstone, a siltstone with some free sand, deposited. The contact between the two formations is gradational. The Upper Westmere Siltstone overlies the Lower Westmere Siltstone, and has a basal member, the Kaikokopu Shell Grit, deposited following the development of an erosion surface on the elevated surface of the underlying sediment. The shell grit was buried by the silts and sands of the remainder of the formation

as the marine transgression continued, and water depth increased.

The Kupe Formation follows the erosion of the Upper Westmere Siltstone, and consists of richly fossiliferous silts and sands, deposited in variable, but shallow, water depths. The Upper Kai-iwi Siltstone conformably overlies the Kupe Formation, and represents the return to the deposition of silts and fine sands in offshore conditions. The Siltstone is for the most part barren, although fossils are concentrated into two bands, possibly representing shallowing conditions.

(h) Shakespeare Group

The Shakespeare Group consists of a series of marine sandstones and siltstones, often richly fossiliferous and has a thickness of c. 80 m.

The Upper Kai-iwi Siltstone was elevated and subaerially eroded, before a renewed marine transgression took place in the area, depositing the silt and sand units of the Seafield Sand. The basal member of the formation is represented by pebble conglomerates (Toms Conglomerate), especially well developed in the east of the basin, deposited on a beach in advance of the transgression, which later deposited the silt and sand of the remainder of the formation. The sediment passes up into a layer of shell accumulation, the Lower Castlecliff Shellbed, which is uniform in character and which formed in the same way as described earlier. The succeeding sediment, the Pinnacle Sand, was deposited in deeper water, and consists of laminated sands and minor silts. Currents were still active during the deposition, as shown by the concentration of the fossils on the bedding planes.

The Tainui Shellbed lies conformably on the Pinnacle Sand, and was deposited under similar conditions. The marked difference between the two is in the accumulation of large numbers of shells in the former, although the matrix is similar. This probably represents slight shallowing of the sea. The Shakespeare Cliff Siltstone follows the Tainui Shellbed conformably, and is a massive, fine, blue-grey siltstone which becomes sandier towards the top of the formation, and represents a deeper water deposit. The Shakespeare Cliff Sand is markedly different to the underlying formations, as the quiet offshore conditions are replaced by shallow water with vigorous wave action, which led to the scouring of the underlying Siltstone and the deposition of the basal shellbed member. The majority of the formation is represented by shallow water sand deposited at and/or below tide level.

The Upper Castlecliff Shellbed is conformable on the Shakespeare Cliff Sand, and is an accumulation of shells in shallow water caused by the scouring of the material by currents. The Karaka Siltstone represents an offshore, deeper water deposit which resembles the conditions which gave rise to the Westmere Siltstones. Following the deposition of the Upper Castlecliff Shellbed, the sediments were uplifted and erosion took place before intertidal and beach sands, the Mosstown Sand, covered the wave cut

platform. The Putiki Shellbed, overlying the Mosstown Sand, is a thin shallow water shellbed. The top of the Group is marked by the Landguard Formation, which was deposited after the Mosstown Sand, but intervening beds in the area are missing. The Landguard Formation consists of a basal yellow, quartzose, muddy sand (the Landguard Sand) with abundant fossils, which passes up into pebble bands and looser sands. It was deposited in offshore conditions at a shallow depth (but below wave base). Locally, the top of the Landguard Sand is bored, indicating that some emergence took place during the deposition of the group.

2.3.2 Hawera Series

The Hawera follows a similar pattern of marine and non-marine deposition to that in the Wanganuian. They are separated from the rocks of the Wanganuian by a regional unconformity.

(a) Pouakai Group

The sediments of the Pouakai Group represent covering beds on marine and fluviatile benches, and reaches a maximum thickness of 70 m.

The Kaiatea Formation is the basal unit of the Group and consists of alluvial sands and muds containing pebble bands deposited entirely by the action of streams. The formation is deposited on the remnants of a peneplaned Wanganuian land surface, which has been elevated and dissected. The relationship between the Kaiatea Formation and the underlying Shakespeare group sediments is not clear. The deposition of the overlying Brunswick Formation followed the development of an erosion surface. The sands and pebbles were deposited under very shallow marine conditions. Locally, fluviatile, deltaic and lagoonal deposits developed, and emergence followed the deposition of the marine beds. The transition between the marine and terrestrial deposits is marked throughout the basin by the Fordell Ash, which inter-fingers with both types of deposit. Dune sand forms the uppermost deposit of this Formation.

The Rapanui Formation is the youngest of the formations considered in this study, and consists of a veneer of sediment unconformably overlying rocks of Wanganui age. The lowest member of the Formation is the Rapanui Marine Sand, the only member of the Formation to contain marine fossils, and which is conglomeratic at the base. The Waipuna Delta Conglomerate replaces the Rapanui Marine Sand to the east of the basin. The Formation is thought to have been deposited on a wave cut platform developed soon after the formation of the Brunswick Terrace, and the Marine Sand was deposited at the low tide mark. Elevation of the Rapanui Marine Sand led to the deposition of the terrestrial deposits of the Rapanui Lignite. Ample sediment supply in the depression that followed the deposition of the lignite led to progradation of dunes and the deposition of the Rapanui Dunesand.

The overlying beds are difficult to distinguish from the Recent and consequently are not considered further.

2.4 Geological History of the South Wanganui Basin

The South Wanganui Basin (**figure 2.1**) is a broad half-graben structure, trending north-north east, developed in Pre-Tertiary basement greywackes and schists (Anderton, 1981). This basin is bounded to the west by a zone of basement highs and major faults separating it from the Cretaceous to Pleistocene sediments of the Taranaki basin. Greywackes of the Ruahine and Tararua Ranges, part of the Axial Range of North Island, form the Eastern boundary, and are separated by faults caused by syn-sedimentary block fault movements. In the south, the sediments onlap to the basement rocks of the Marlborough Sounds, whilst in the north east, the sediments unconformably lie on basement greywackes. To the north, lie Oligocene to late Miocene sediments of the North Wanganui Basin, which is distinct from the South Wanganui Basin, although the relationship between them is unclear (Anderton, 1981).

Lower Tertiary rocks are found in the oil prospecting wells drilled in the Taranaki Basin (**figure 2.1**), where an almost complete Cenozoic stratigraphic record is found (Maui-1). These rocks may have once covered the entire area, but are not found in the centre of the basin in bore holes which penetrate to basement. The oldest sediments found in the present day South Wanganui Basin are faulted outliers of Oligocene age contained within basement. However, these are thought to represent remnants of older strata which predates the formation of the South Wanganui Basin. In Taranaki, Upper Miocene sediments lie conformably beneath rocks of Wanganui age, although once again, these rocks are absent from wells drilled close to the centre of the basin (Young-1, Stantiall-1 and Santoff-1A), which all penetrate to basement (**figure 2.1**).

Seismic data and well records indicate that the faults bounding the basin to the west were active between the late Cretaceous and Miocene (Pilaar and Wakefield, 1978). These faults lead to early Tertiary sedimentation, but movement continued until the late Miocene when the area was uplifted and an unconformity developed.

Sediments overlying the unconformity form a shallow water (5-60 m) sequence ranging in age from the latest Miocene to late Pleistocene, interrupted by periods of exposure. Late Miocene to early Pliocene sediments of the Waiouru Sandstone and Taihape Mudstone (not collected) are overlain conformably by sandstones, mudstones and shellbeds of the Waipipian. This period represents a marine transgression from the north (Fleming, 1953), with a palaeoshoreline c. 15 km to the south of Wanganui (Anderton, 1981) trending east-south east/west-south west. Basement rocks still had relief, and the Lower Pliocene sediments onlap these. The depocentre lay to the north west of Wanganui and extended to the Taranaki basin (Anderton, 1981). The sediment source for these

deposits was the older Tertiary rocks of North Taranaki, and an area of granites forming a possible extension of the Tasman Range, South Island (Fleming, 1953).

In the Paparangi (Lower Mangapanian), sediments in the west were uplifted into the tidal zone, giving rise to a basal conglomerate. In the centre of the basin, deeper water sediments were deposited, and to the east, shallowing sequences are indicated by local conglomerates (Fleming, 1953). The marine transgression continued southwards during the late Pliocene-early Pleistocene, and the depocentre migrated to the south east of Wanganui. The majority of the sedimentation occurred in the Pliocene (Mobil Oil Corporation, 1971 in Anderton, 1981). Several warm water genera, present in the Waipipian, became extinct during the Mangapanian, indicating that climatic deterioration had begun.

In the late Pliocene to early Pleistocene, basement block movements, and faults on the eastern margin controlled sedimentation in the basin. As the depocentre migrated south eastwards, offlap and emergence occurred in the north.

By the early Pleistocene, the South Wanganui Basin had become separated from the Taranaki Basin. From the Lower Nukumaruan onwards, glacio-eustatic sea level changes are superimposed on the tectonic events in the basin, complicating interpretation. Beu and Edwards (1984) have shown, by correlating the oxygen isotope signature of the sediments from the South Wanganui Basin with those of core V-28-239 from the Soloman Plateau, Pacific Ocean (Shackleton and Updyke, 1973), that up to twenty glacio-eustatic cycles have occurred during the last 2 Ma. Pillans (1983) concluded, from work on the marine terraces of the Taranaki-Wanganui coastal plain, that twelve high-sea level events have occurred during the last 0.68 Ma.

Palaeontological evidence shows that during the Hautawan (Lower Nukumaruan), cool water species migrated north into the Wanganui area, indicating cooling at this time, corresponding to the onset of glaciation in the South Island (Gage, 1945). The glaciation began a series of disconformity bounded cycles in the South Wanganui Basin. During the Pleistocene, sea level changes were of the order of 100 m (Williams *et al.*, 1981), and as the sedimentary sequence was deposited in water depths of between 5 and 60 m (Beu and Edwards, 1984), it is likely that the sediments emerged several times during their geological history. Fleming (1953) recognised several disconformable horizons, which were interpreted as being due to emergence during glacial periods. Beu and Edwards (1984) therefore conclude that most of the sediments deposited in the Nukumaruan and Castlecliffian were deposited during interglacials.

The top of the Nukumaruan is marked by an angular unconformity, which represents the erosion of seven of the oxygen isotope cycles prior to the deposition of the basal beds of the Castlecliffian. The unconformity at this horizon is interpretted as being caused by further migration of the depocentre in the mid-late Pleistocene (Beu and Edwards, 1984) to 50 km south of Wanganui, where 2000 m of sediment accumulated, and

of the northern margin of the basin to just north of Wanganui (Anderton, 1981). This migration caused further offlap and emergence to the north.

Deposition during the Okehuan (Lower Castlecliffian) incorporated layers of ash and banded rhyolite (Fleming, 1953), corresponding to eruptions in the Taupo Volcanic Zone, which had ceased by the Putikian (Fleming, 1953). Fission track dating has been completed on glass from these formations (e.g. Seward, 1974).

Post-Putikian, a period of emergence led to development of a regional unconformity prior to deposition of the Hawera Series, represented by deposition on marine terraces. Twelve terraces have been identified (Pillans, 1983) and each terrace consists of a sub horizontal wave cut platform, which is covered by up to 12 m of marine sands, often with a conglomeratic base, and rare shell horizons. The inner margin of each terrace is marked by a fossil sea cliff. The marine sediments are overlain by a variety of terrestrial deposits, including lignites and tephras which are important in correlation and dating. The total thickness of sedimentary cover exceeds 40 m in some cases. The terraces are formed by high sea level events during interglacials during the last 0.6 to 0.7 Ma (Fleming, 1953; Pillans, 1983). After the Wanganui Series ended, glacio-eustatic control was the main sedimentological mechanism in the South Wanganui Basin. The youngest beds sampled (Rapanui Marine Sand) occur on the third oldest terrace. Two younger terraces lie offshore at a palaeosealevel of -20 and -11 m at ages of 60 ka and 90 ka respectively (Pillans, 1983). Other late Quaternary shellbeds have been identified in cores from off the Taranaki Coast (Norris and Grant-Taylor, 1989) and have a mean date of 16 ka. These offshore shellbeds, if further studied, may yield more detailed information on the formation of the fossil shellbeds in the South Wanganui Basin.

Chapter Three: Methods and materials

Methods and Materials

3.1 Introduction

The following chapter is a detailed account of the experimental procedures used during the study, and which are summarised in the relevant sections. Where appropriate, reference has been made to a standard text for that process. The sources of materials are listed in **section 3.9**, and the addresses are given in **Appendix 3**.

3.2 Extraction of intracrystalline amino acids and proteins

3.2.1 Selection of material

Samples of the best possible physical preservation were selected for further study, although this is not necessarily a sign of good biochemical preservation (Towe, 1980). Wherever possible, samples which were excessively bored or fractured were omitted from further study. Recrystallisation of original shell material is also of major concern in the study of fossil molecules. During the recrystallisation of calcium carbonate from meta-stable aragonite to stable (low magnesium) calcite, it is likely that some, if not all, of the molecules included within the shell could be lost, or other non-indigenous molecules incorporated. The majority of the samples under study belong to the Phylum Brachiopoda, which have shells composed of primary low magnesium calcite (**Chapter Two**), and hence do not suffer from this problem.

3.2.2 Cleaning of material

All sediment was scrubbed from the surface of the sample and any encrusting epifauna removed by scraping. Articulated shells were disarticulated, and the inside surfaces cleaned. Any remaining body tissue was removed, and the shells soaked in an aqueous solution of bleach (10% v/v) for 2 hours at room temperature to digest adhering particles of organic matter, washed extensively with Milli RO™ water and left to air dry. Samples were ground using a ceramic pestle and mortar, before being incubated in an aqueous solution of bleach (10% v/v) under constant motion for 24 hours at room temperature. The sodium hypochlorite in the bleach destroys the intercrystalline molecules by oxidation. After bleaching, the samples were washed by repeated agitation with MilliQ™ water and centrifugation, until no bleach could be detected (typically ten washes). Powders were then frozen (-20°C) and lyophilised. Once this process is complete, samples may be stored at room temperature in airtight tubes almost indefinitely, until ready for further preparation.

3.2.3 Demineralisation

To release the incarcerated molecules from within the calcium carbonate, it is necessary to dissolve the inorganic phase. Various methods have been employed to complete this, as discussed below:

(a) Ethylene diamino tetra acetic acid, disodium salt (Na_2-EDTA)

An aqueous solution of Na_2-EDTA (20% w/v) at a ratio of 23 mL/g shell was used to dissolve the fossil shell powder by chelation of the calcium ions. Several days under constant motion at 4°C are required for complete decalcification. Once demineralisation is complete, samples were centrifuged (20 g.h.) to remove any remaining insoluble particles. The resulting solution of EDTA and protein was too dilute for amino acid analysis, and the EDTA also interferes with PITC amino acid analysis, by coeluting with the amino acids Glu, Ser, Gly and Pro, thus distorting the pattern of amino acid composition of the molecules (**figure 3.1**).

Removal of the EDTA/calcium complexes and concentration of the sample is therefore necessary prior to further analysis, and either ultrafiltration or dialysis is normally used. EDTA is notoriously difficult to remove from a solution by dialysis (Worms and Weiner, 1986) and also by ultrafiltration (e.g. Amicon; M. Cusack pers. comm.). Attempts to use these methods resulted in contamination of the sample by coelution of peaks (**figure 3.1**), and by the distortion of the protein bands on gels. A new method of preparation, which filters the sample using the Minitan™ tangential flow system (Millipore), with 10 kDa cutoff filters, rapidly removed all of the EDTA/calcium complexes from the solution whilst concentrating the peptides (Cusack *et al.*, 1992). Following removal of EDTA and concentration of the sample on the Minitan™, the sample was further concentrated using a Minicon™ static mixer (Amicon), also with a 10 kDa cutoff. Once the extract has concentrated, it is washed with MilliQ™ and centrifuged. The supernatent was used in further study.

Due to the pore size of the filter used with the system, all small peptides and free amino acids present within the sample are lost from the filtrate, hence this preparation technique requires the dissolution of large amounts of shell material in order to recover significant concentrations of preserved peptides. Although information on fossil peptides will be valuable, such large quantities of shell material are unlikely to be available for all samples which undergo molecular analysis. Rare samples, such as those stored in museums, could not be analysed using these techniques. Information is more likely to be gained using a technique requiring less shell powder, as it is more likely that a comprehensive survey of samples can be undertaken.

Similar results have been discussed for the recovery of DNA from bone extracts (discussion following Horai *et al.*, 1991); EDTA decalcification cannot used as complete removal of EDTA is difficult and necessary.

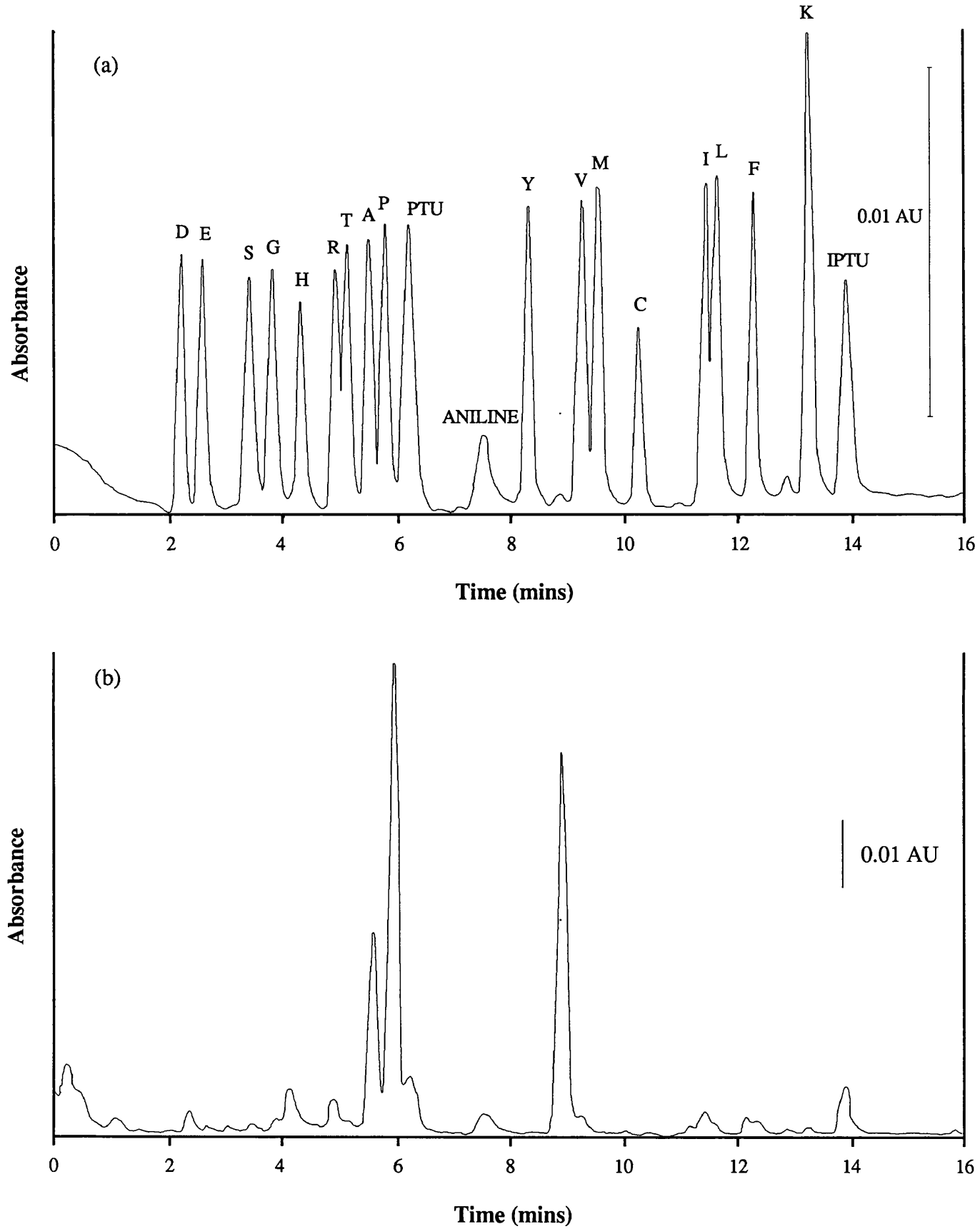


Figure 3.1: (a) Typical chromatogram produced by the amino acid analyser, showing the relative positions of each of the amino acid peaks. (b) Chromatogram produced by contamination with an aqueous solution of EDTA (20% v/v). Note false levels of amino acids, especially Ala and Pro.

(b) Mineral acids

Several mineral acids were tested to find a decalcification solution which did not interfere with the analysis system, did not damage the protein to any great extent and which could be used at a level which would require no further concentration. Aqueous solutions (20% w/v) of acetic acid, formic acid and hydrochloric acid were added to shell samples at the same ratio as for Na_2-EDTA. However, several days after the Na_2-EDTA sample had completely demineralised, the mineral acids had not reacted to completion. A series of reactions between different molarities of acid at different ratios to shell powder was set up to attempt to find a balance between strength of acid which would not damage any remaining protein, but which would leave the remaining molecules in a solution which would not require any further concentration prior to analysis, and hence retain the ability to quantify free amino acids and peptides.

Eventually, 2 normal (2N) hydrochloric acid was selected for decalcification. This fulfiled the criteria for decalcification of fossils in the following ways:

(i) Test decalcifications with Recent samples indicated low concentrations of free amino acids, when analysed without prior hydrolysis, indicating that the damage to the proteins/peptides is slight, i.e. the hydrolysis of peptide bonds is not significant by the exposure to 2N HCl at room temperature (**Chapter Five**).

(ii) Hydrochloric acid, unlike Na_2-EDTA, does not produce spurious peaks on the chromatogram, nor does it seriously affect the recovery of amino acids. The affect of HCl was tested by adding 20 µL 2N HCl to known concentrations of standard amino acids (Pierce Standard H) on the analysis frit. The elution of Asp, Glu, Ser and Gly from the hplc column was delayed due to the decrease in the initial pH of the buffer system. This effect was easily corrected for, by changing the system calibration file. No other effects were recorded for the amino acids.

(iii) Hydrochloric acid could be obtained at a very high purity, from Applied Biosystems for use in protein hydrolysis.

(iv) For complete demineralisation of shell powder, hydrochloric acid is required at a ratio of 11 mL/g shell, less than that required for Na_2-EDTA. As there are no spurious peaks, there is no requirement for filtration, hence small peptides and free amino acids are retained in the fraction of the sample to be analysed.

3.2.4 Removal of insoluble compounds

When the shell matrix of the fossil is demineralised, an insoluble residue remains (**section 3.4**). This residue is thought to be derived from the reaction between sugars and amino compounds (Hoering, 1973). The insoluble residue, known as melanoidin and similar in structure to humic acids (Hoering, 1973), cannot be quantified

by the analyser, and needs to be removed prior to loading, as the melanoidin can block the analysis system. This is completed by centrifugation (2000 g.h.). The formation and amino acid composition of these samples is discussed in **Chapter Nine**.

3.2.5 Direct confirmation of the presence of intracrystalline amino acids

Brachiopods contain organic molecules with antigenic properties within their shells (Collins *et al.*, 1988; **figure 2.5**). To confirm the presence of these molecules by direct analysis, and to confirm that the bleaching process effectively removed the intercrystalline molecules, the following experiment was completed.

A sample of shell powder (0.70 g) of the Recent brachiopod *Terebratella sanguinea* was directly weighed into newly pyrolysed (500°C/4 hours) glass universal bottles, 10 mL of pure water (MilliQ™) added and the sample incubated at 110°C for 24 hours. On removal, the sample was allowed to cool, and an aliquot (4.5 mL) removed and concentrated on a rotary evaporator (Howe Gyrovap) to 25 μL. An aliquot (20 μL) was added to the sample frit, and the standard hydrolysis and derivatization cycles run. Results indicated that for every milligram (10^{-6} gram, mg) of powder analysed in this way, there are 0.7 nanograms (10^{-9} gram, ng) of amino acid. The shell powder was frozen and lyophilised again, and the inorganic phase demineralised by 2N HCl as above. The concentration of amino acid yielded by this process, was 169.4 ng/mg, slightly lower than the average for Recent *T. sanguinea* (180 ng/mg), but within the range of experimental error. The proportion of amino acid which may be attributed to the intercrystalline fraction which remains undestroyed by the bleaching procedure is c. 0.4%. The amino acid quantified upon demineralisation must therefore have been enclosed within the shell of the brachiopod i.e. intracrystalline.

3.3 Separation of soluble intracrystalline molecules

The analysis of fossil peptides requires purification to homogeneity. Separation of fossil molecules was attempted, and brief details of the two techniques used in this study are given below. More extensive coverage of these techniques and information of the others are to be found in Harris and Angal (1989).

3.3.1 Liquid chromatography

Chromatography is a technique which allows the separation of molecules on the basis of their chemical and physical properties. Organic molecules such as amino acids and proteins differ in these properties (**Chapter One**), thus enabling separation via liquid chromatography. The techniques employed within this study were gel filtration (a type of fast protein liquid chromatography, fplc) and reverse phase (a type of high pressure liquid chromatography, hplc).

Both hplc and fplc use interactions between the sample with the solid phase of the beads which are used to pack the chromatography column. The chemical and physical characteristics of the beads determines the way in which separation is completed, although the principles behind the methods are the same in all types of chromatography:-

(i) At a constant flow rate of the mobile phase, the liquid sample is injected and binds onto the column.

(ii) Buffers are pumped through the column and an appropriate gradient applied. Components of the gradient compete in the system, and at specific points in the gradient the protein or peptide will elute from the column.

(iii) The molecules travel through the detector where absorption of UV light is correlated with the concentration of the molecules, and are collected by a fraction collector.

(a) Gel filtration

This method, also known as size exclusion chromatography, separates molecules solely in terms of their size. Small molecules pass into the cross linked acrylamide beads of the column, and thus take longer to pass through the column. Larger molecules are excluded from the beads and hence travel faster through the column. Gel filtration has the advantage that standard proteins of known molecular weight may be passed through the column, and thus the molecular weight of the unknown proteins estimated by comparison with the standards. Gel filtration of samples was completed using a Superose 6 column (Pharmacia) and MilliQ™ water as the mobile phase. The buffer was pumped through at a rate of 1 mL/minute for 60 minutes, and fractions collected every minute. Eluates were monitored at wavelengths of 280 nm for aromatic structures and 214 nm for the peptide backbone.

(b) Reverse phase

This method uses a column packed with beads which have apolar surface properties, which bind to hydrophobic regions of the protein. Molecules are eluted from the column using a gradient which slowly lowers the polarity of the buffers until the interaction between the bead and a particular molecule diminishes. Molecules are therefore separated in terms of hydrophobicity. Samples were applied to an Aquapore™ column (Applied Biosystems) in an aqueous solution of trifluoroacetic acid (TFA, 0.1% v/v) at a flow rate of 0.1 mL/minute. After 5 minutes, a 40 minute linear gradient of 0 to 70% (v/v) acetonitrile in 0.1% (v/v) TFA was applied to separate the molecules. The eluate was monitored at 280 nm and 214 nm, and fractions collected every minute.

3.3.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Proteins are charged molecules at any pH other than their iso-electric point.

This characteristic allows proteins to be separated in terms of their charge when an electric current is passed through a mixture, in a process known as electrophoresis. To nullify effects such as diffusion and convection, this process takes place in a support medium. In this study, polyacrylamide gels were used as the support medium, prepared to the method of Schägger and von Jagow (1987), formed from a mixture of acrylamide and N, N'-methylene bisacrylamide (Bis), with polymerisation being initiated by ammonium persulphate and N, N, N', N', -tetramethylethylenediamine (TEMED). The size of the pores within the gel may be altered by changing the proportion of acrylamide present within the buffer; in this case, 15% polyacrylamide gels were used with the following components:

Resolving Gel: 2 mL acrylamide solution (48 g acrylamide, 1.5 mL Bis in 100 mL MilliQ™); 3.3 mL gel buffer (3M Tris, 0.3% (w/v) SDS, pH 8.45); 2 mL glycerol; 2.6 mL MilliQ™; 100 µL APS; 10 µL TEMED.

Stacking Gel: 0.35 mL acrylamide solution; 1.03 mL gel buffer; 2.75 mL MilliQ™, 50 µL APS; 10 µL TEMED.

In order to swamp the charge of the proteins and utilise the polyacrylamide gel as a molecular sieve, the anionic detergent sodium dodecyl sulphate (SDS) was used. An aqueous solution of the protein was mixed with an equal volume of buffer solution containing final volumes of 0.15 M Tris/HCl, pH 6.8, 0.2 M 2-mercaptoethanol, 0.1% (w/v) SDS, 30% (w/v) glycerol and 0.0002% (w/v) of the tracking dye, bromophenol blue, and incubated at 100°C for 4 minutes. The heat, mercaptoethanol and SDS cause the molecules to denature, and all intra-molecular bonds are broken, allowing the SDS to bind uniformly. The majority of proteins bind to SDS in a constant ratio of 1:1.4, masking their intrinsic charge and allowing separation of proteins and peptides solely in terms of the molecular radius, which approximates to the molecular size. Mixtures of protein and buffer are loaded into the wells of the stacking gel (**figure 3.2**), and a constant voltage of 100 V applied to the gel between the cathode buffer (0.1 M Tris, 0.1 M Tricine, 0.1% (w/v) SDS) and the anode buffer (0.2 M Tris, pH 8.9). Molecules migrate according to their molecular radius, with small molecules being less restricted by the polyacrylamide than the larger molecules, and hence travelling faster. The gel therefore acts as a molecular sieve. Proteins of known molecular weight are electrophoresed alongside samples, and the molecular weight of the proteins in the extract can therefore be determined by comparison with a standard curve.

Following electrophoresis of the samples, molecules were visualised by one of two stains with differing sensitivities:

(i) **Coomassie Brilliant Blue R-250** A solution of CBB-250 (227 mL methanol, 250 mL acetic acid, 1.25 g CBB-250, 23 mL MilliQ™) was used to fix and to reveal protein bands. Destain (750 mL methanol, 250 mL acetic acid) reduces background staining on the gel. This staining technique can

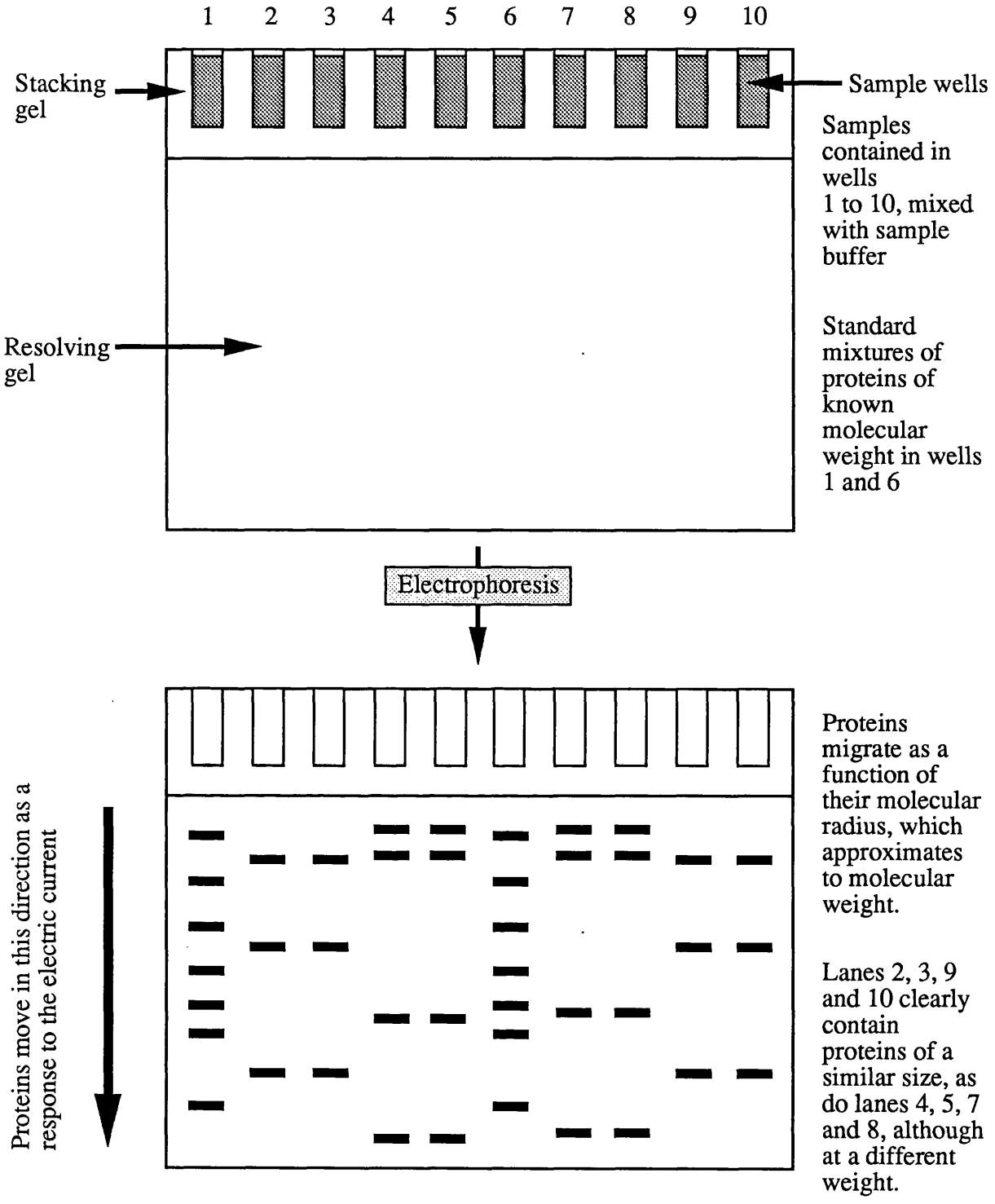


Figure 3.2: Generalised diagram representing the process of electrophoresis. Samples are loaded into the wells at the top of the gel, before being exposed to an electrical current which acts to separate the molecules in terms of their molecular radius (see text), smaller molecules travelling faster than the larger ones.

detect approximately 1 μg of protein in a band.

(ii) **Silver nitrate** staining has a sensitivity to 1 ng of sample, and is a more complex method. The protein bands in the gel are incubated with fixer (50 mL methanol, 12 g TCA, 50 mL MilliQ™), the gel reduced by DTE solution (5 mg DTE per litre MilliQ™) and stained by a solution of silver nitrate (0.2 g $AgNO_3$ in 100 mL MilliQ™). The bands are visualised in developer (20.22 g $Na_2CO_3.10H_2O$, 0.125 mL formaldehyde (35 % (w/v)), made up to 250 mL with MilliQ™), and the staining stopped by 1% (v/v) acetic acid.

3.4 Extraction of acid insoluble intracrystalline molecules

To ensure that the results of this survey were consistent with those gained from protein and amino acid quantification from the same horizons, the methods of HCl decalcification (above) were followed as closely as possible. In summary, shells were cleaned, powdered and bleached to remove intercrystalline material, before being washed, frozen and lyophilised. Shell powders were weighed directly into microcentrifuge tubes, and incarcerated insoluble compounds were released from within the calcium carbonate of the shell by the addition of 2M HCl in the ratio 11 μL acid/mg shell. Upon completion of the reaction, the samples were placed on a microcentrifuge shaker (Eppendorf) for 10 minutes before being centrifuged (2000 g.h.), and the acid soluble portion removed by pasteur pipette. Samples were then repeatedly washed with pure water (MilliQ™) by mixing and centrifugation to remove any remaining HCl, frozen and lyophilised. The gain in weight of the microcentrifuge tube represented the mass of insolubles present in the shell.

The insoluble compounds were transferred from the microcentrifuge tubes to newly pyrolysed glass hydrolysis tubes by suspension in 0.5 mL MilliQ™, followed by drying down in a rotary evaporator (Howe Gyrovap). Samples were manually hydrolysed and amino acids recovered using an aqueous solution (0.025% w/v) K_3-EDTA added to the hydrolysis tube. Some samples of Na_2-EDTA soluble fossil protein were also hydrolysed manually. Samples were dried down in the tube using rotary evaporation prior to hydrolysis and reconstitution in the same way as above.

3.5 Preparation of sediments for analyses

3.5.1 Collection

Where possible, the samples were collected from a few centimetres within the exposure, before being sealed in plastic bags until arrival in the laboratory. The fractions for analysis were taken from the inside of the collected sediment wherever possible, to minimise the possibility of introducing non-indigenous amino acids from fingertips,

which can be a major source of contamination (Walton and Curry, 1991; **Chapter Four**).

3.5.2 Extraction of amino acids

Amino acids were extracted from sediments using a modification of the method used by Cronin *et al.* (1980) for meteorites. Sediment was directly weighed into glass bottles, pure water (MilliQ™) added, the bottles sealed, and samples incubated at 110°C for 24 hours. When cool, the bottles were shaken and an aliquot (1 mL) of each sample was centrifuged to remove suspended sediment and an aliquot (700 μL) of the supernatent concentrated to ≈ 150 μL. All samples were diluted to 200 μL with MilliQ™.

3.6 Sampling of amino acids from fingertips

Free amino acids from finger-tips were collected from the fingers of members of the Department of Geology and Applied Geology, University of Glasgow. Sample selection was random with respect to the time of day, sex and age of the individuals. An aliquot (500 μL) of citrate buffer (100 mM, pH 2.2) was added to a sterile Cel-cult tissue culture plate well, and mixed on a Luckman (Model 802) suspension mixer to ensure complete coverage of the base. To remove traces of solid particles, the index finger of each individual was wiped with a tissue immediately prior to sampling. Each finger was then pressed into the well for 15 seconds. Nine samples were taken, including some from fingers which had been washed with either Millipore Reverse Osmosis Water (Milli RO™) or soap and tap water immediately prior to the test. One well was left untouched to provide information on the background level of amio acids, and one was touched with the finger in a latex glove that had been used for sample preparation and collection. An aliquot (30 μL) of each of the samples was applied to the sample frit of the analysis system.

3.7 Amino Acid Analysis

3.7.1 Protein and peptide hydrolysis

Hydrolysis of the peptide bonds which link individual amino acids in proteins and peptides is a prerequisite for amino acid analysis. This procedure frees amino acids for further analysis by cleavage of peptide bonds. There are numerous hydrolysis methods, summarised in **table 3.1**. However, in this study, vapour phase hydrolysis, using 6N HCl was completed in two ways.

(a) Manual hydrolysis

Acid insoluble molecules were manually hydrolysed in newly pyrolysed (500°C/12 hours) pyrex glass tubes (6 mm x 50 mm; University of Glasgow, Department of Chemistry). Samples were placed in the base of the tube, and the tubes introduced into glass hydrolysis bottles. An aliquot (500 μL) of constant boiling 6N HCl (Pierce) was

	Method	Conditions	Comments	References
(1)	6M HCl (±phenol)	110°C, 20-70 hours, vacuum	loss of Trp/Cys/Thr/ Ser/Tyr	Moore and Stein (1963)
(2)	6M HCl and Na_2SO_3	110°C, 24 hours, vacuum	improved Cys, Met and Tyr recoveries	Swadesh *et al.* (1984)
(3)	12 M HCl/ Propinoic acid (1:1)	150-160°C, 15 minutes or 130°C, 2 hours, vacuum	as in (1)	Westall and Hesser (1974)
(4)	HCl/TFA (2:1), and 0.005% phenol	166°C, 25 minutes, vacuum	as in (1)	Tsugita and Scheffler (1982)
(5)	HCl/TFA (2:1), plus 5% (v/v) thioglycolic acid	166°C, 25-75 minutes, vacuum	as in (1)	Yokote *et al.* (1986)
(6)	6M HCl, 0.5-0.6% (v/v) thioglycolic acid	110°C, 24-64 hours, vacuum	improved Trp yields	Matsubara and Sasaki (1969)
(7)	3M ρ-toluene-sulfonic acid	110°C, 22-72 hours, vacuum	as in (6)	Liu and Chang (1971)
(8)	3M mercaptoethane-sulfonic acid	110°C, 24-72 hours, vacuum	as in (6)	Penke *et al.* (1974)

Table 3.1: Examples of some of the possible methods available for protein/peptide hydrolysis.

placed in the bottom of each of the bottles (not in the tubes), the bottle charged with argon to minimise sample loss due to oxidation, and then sealed by a screw cap with teflon insert (Pierce; **figure 3.3**). The bottles were incubated in a preheated oven at 165°C for 1 hour. Hot acid vapour rose from the base of the bottle and entered the tubes, hydrolysing the peptide bonds within the sample. After incubation, the acid vapour was released in a fume hood whilst the bottle remained hot, to avoid subsequent condensation of the HCl in the tubes. Once cool, the samples in the tube were dried in a rotary evaporator (Howe Gyrovap) to remove excess acid.

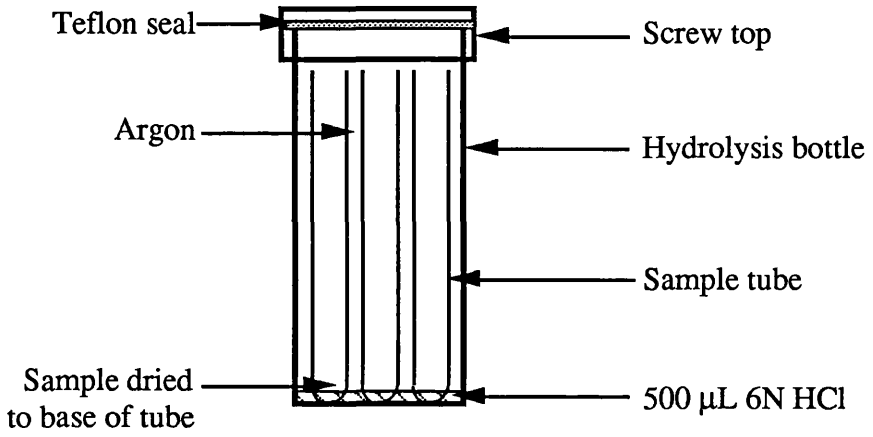


Figure 3.3: Diagramatic representation of a manual hydrolysis bottle.

(b) Automated hydrolysis

The same method of hydrolysis has been adapted for auto-hydrolysis on the Applied Biosystems 420H Amino Acid Analyser. Samples were applied to the frit of the analysis system either in aqueous solution, or immobilised on ProBlott™ PVDF membrane (Applied Biosystems, see **Appendix 2**). The hydrolysis head (**figure 3.4**) clamps over the frit to form a sealed chamber. Argon is delivered to the frit in order to dry the sample and also to provide an inert atmosphere for hydrolysis. Heaters within the head heat up to 200°C, increasing the temperature on the frit to c. 165°C. Acid is delivered towards the frit, but is vapourised within the head, ensuring that only vapour comes into contact with the frit. The temperature is maintained for 75 minutes (90 for PVDF samples), before being reduced slightly, the acid vapour removed and the sample dried with argon.

Automated hydrolysis has advantages over manual hydrolysis in that the process is reproducible and is used in conjunction with automated derivatization, saving a considerable amount of time. Up to 72 samples may be hydrolysed in a single run. However, in both methods, care must be taken to avoid contamination of samples. Standard proteins and peptides were used in every run to ensure that hydrolysis proceeded to completion, and analyses were completed with no sample on the frit to check for background levels of contamination.

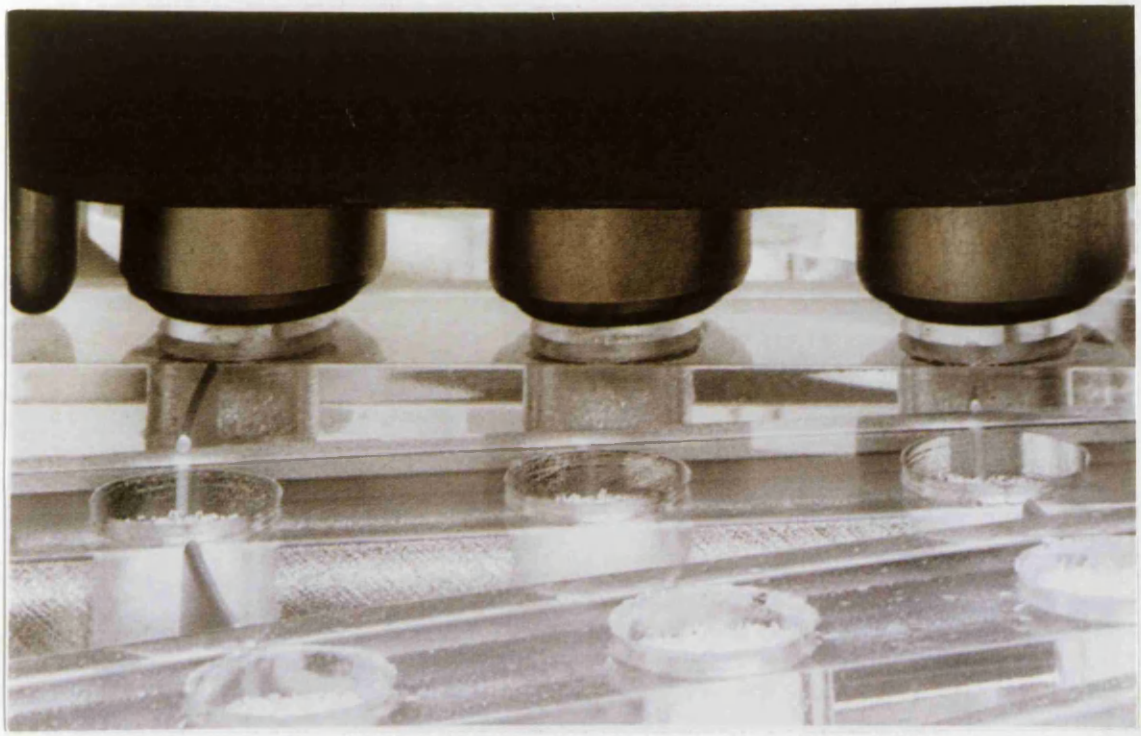


Figure 3.4: The hydrolysis head of the 420H amino acid analyser. The jaws clamp down onto the slide, forming a reaction chamber. All three positions are hydrolysed at the same time. Hydrolysis uses 6N HCl vapour to break the peptide bonds. Photo credit: Applied Biosystems.

3.7.2 Effect of hydrolysis on amino acids

A single hydrolysis will not be able to quantify all of the amino acids present within a sample to 100% accuracy, as vapour phase hydrolysis with 6N HCl causes the modification and destruction of some of the common protein forming amino acids. To quantify these amino acids, it is usually necessary to utilise a 'scavenger', a compound which will enhance the recovery of the amino acid. However, in this study, amino acid scavengers were not used for two reasons. Firstly, amino acids which are altered under hydrolysis conditions will also have been substantially altered through natural hydrolysis over time. Secondly, the use of scavengers is, as yet, almost solely confined to manual methods of hydrolysis, and are only just beginning to be incorporated into the automated hydrolysis procedure of the 420H amino acid analysis system.

Although no scavengers were used to enhance recoveries, any modification or destruction of the molecules will be of a similar nature in all samples, and hence the effect can be negated as all samples will suffer reductions of the same proportion. The effect of the hydrolysis conditions should be noted however, as they may be important in the analysis of Recent molecules.

(a) Modification of amino acids by hydrolysis

(i) **Asparagine and glutamine**: Asparagine and glutamine (**Chapter Two**) are uncharged amino acids. Hydrolysis converts these amino acids to their charged, acidic derivatives, aspartic acid and glutamic acid respectively. Hence, although these residues are seen in protein sequences, they are represented by their acidic derivatives in amino acid analysis.

(ii) **Pyroglutamic acid**: Glutamic acid is one of the most stable amino acids (Vallentyne, 1964) in that it is always recognised in hydrolysates of pyrolysed samples. However, the amino acid readily undergoes lactamisation to form pyroglutamic acid with the loss of water (Wilson and Cannon, 1937). Hydrolysis of this compound reforms the glutamic acid. To test this, commercial pyroglutamic acid (Sigma) was analysed both with and without hydrolysis. The chromatogram of the analysis without hydrolysis only showed systems peaks, whereas that which had been hydrolysed showed a peak corresponding to glutamic acid. This reaction is important in the consideration of fossil amino acids (**Chapter Seven**).

(b) Destruction of amino acids by hydrolysis

(i) **Serine**: This amino acid is subject to dehydration and esterification of the hydroxyl group in the side chain. Losses due to this reaction tend to be within the range 15-20%.

(ii) **Threonine**: Thr is also subject to the alteration of the hydroxyl group, although losses are only usual in the range 10-15%, due to the increased stability from an extra carbon atom present within the side chain.

(iii) **Tyrosine**: The phenolic ring structure present within the side chain of Tyr is subject to attack by chlorine and hypochlorite radicals present in the HCl used in sample hydrolysis. Losses typically amount to 15-20% of the total, although this may well vary depending on the quality of the acid used.

(iv) **Methionine**: The side chain thioester ($-CH_2SCH_3$) present in the side chain of the amino acid is very sensitive to oxidation. Losses are variable and depend on the sample quantity, acid quality, exposure to air and amount of acid used.

(v) **Cysteine/Cystine**: Both the free sulphydryl (-SH) group in Cys and the disulphide group (-S-S-) in peptides are sensitive to a variety of reactions, causing losses of greater then 50% from the total in the sample.

(vi) **Tryptophan**: 100% of this amino acid is lost due to attack on the carbon-carbon double bond (C=C) present within the ring structure of the amino acid.

(c) Coelution of hydrolysis products with amino acids

Both the dehydration product of Ser, dehyroalanine, and the oxidation product

of Met, methionine sulphoxide elute from the hplc column with PTC-arginine.

3.7.3 Amino acid separation and quantification

After hydrolysis has broken the peptide bonds, individual amino acids may be separated from the mixture and quantified. Amino acids are charged molecules, which may be separated by various methods of liquid chromatography. However, most of the protein amino acids do not possess significant UV absorbance and fluorescence (usual methods of detection) at wavelengths suitable for detection. Amino acid molecules are derivatized by chemicals to produce a derivative which has some property which enables detection, usually strong colouration or UV absorbance.

Early methods of derivatization developed by Moore and Stein (1948) and Spackman *et al.* (1958), utilised post-column derivatization with ninhydrin, a compound which binds to the amino acids, allowing detection to a sensitivity of 1 nmol. Improvements in the technique using ninhydrin have reduced the time necessary for analysis, and have increased sensitivity to c. 100 pmol. Ninhydrin, however, has limitations in use because of its chemical characteristics. The compound requires an organic solvent, and is also unstable and therefore requires an inert atmosphere, cooling and darkness. Ninhydrin-amino acid complexes, have absorption maxima at 570 nm and 440 nm, for primary and secondary amino acids respectively, requiring dual wavelength monitoring.

This study utilised pre-column derivatization, which has fewer of the problems found in post-column derivatization, using Phenylisothiocyanate (PITC; Heinrikson and Meredith, 1984) on the ABI 420H (Dupont *et al.*, 1989). Diisopropylethylamine (DIEA) produces a basic atmosphere to neutralise any remaining HCl and to aid derivatization, before PITC is delivered to the sample. PITC reacts with amines (including the amino group of the amino acids) to yield phenylthiocarbamyl amino acids (PTC-amino acids). This derivative absorbs UV light strongly at 254 nm. Unfortunately, some solvents and buffers also absorb strongly at this wavelength, and care is required to avoid sample contamination. PITC derivatization, followed by reverse phase hplc (Heinrikson and Meredith, 1984), can routinely detect 10 pmoles of amino acid. Other methods of derivatization are summarised in Allen (1989), and will not be considered further here.

3.8 Summary of amino acid analysis steps

Aqueous solutions of samples are applied to the glass frit on the analysis system (**figure 3.5**) in the load position. The slide, containing three sample frits, is rotated to the hydrolysis head, and the jaws clamp to the slide (**figure 3.4**). Vapour phase hydrolysis takes place on all three positions at the same time. Once hydrolysis is complete, the turntable rotates until the newly hydrolysed slide is in the derivatization position. The derivatizer head clamps on each frit sequentially (i.e. on one frit at a time), and PITC

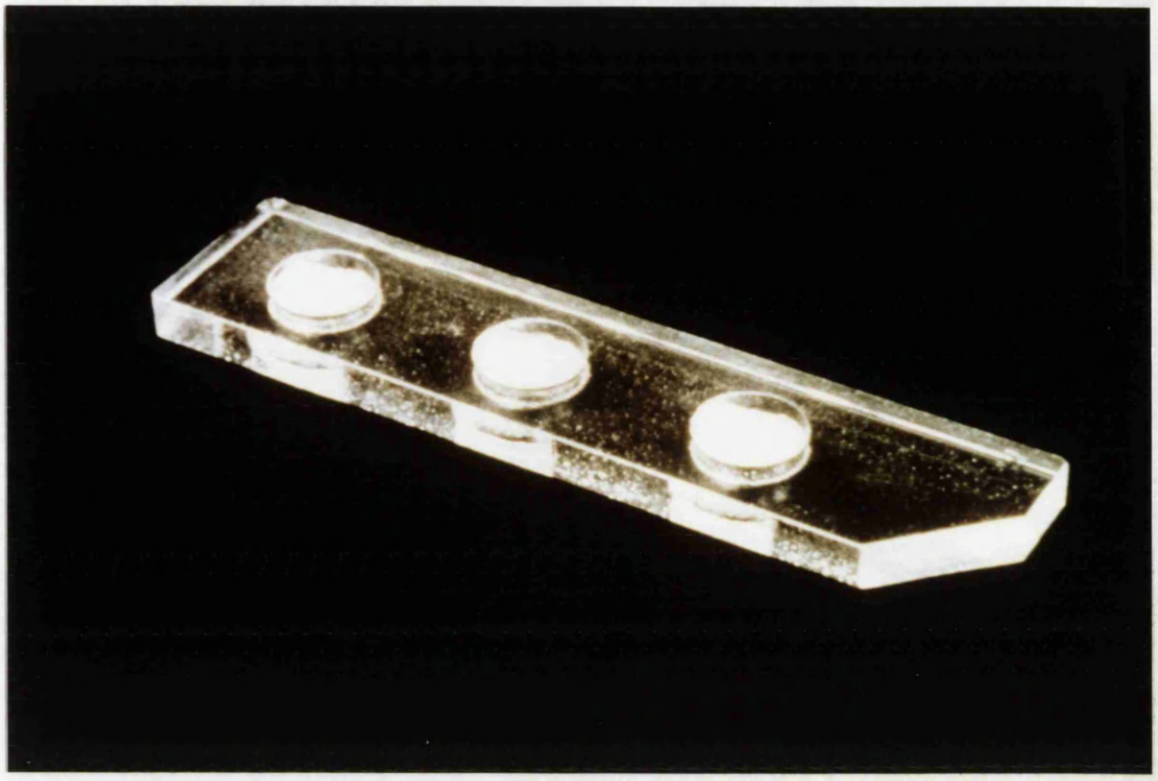


Figure 3.5: The slide from the 420H amino acid analyser. The sample is spotted onto the frit (the discs on the slide) where hydrolysis (if necessary) followed by derivatization takes place. Photo credit: Applied Biosystems.

derivatization takes place (**figure 3.6**). When this is complete, the sample is dissolved in 600 µL 29 mM sodium acetate buffer, pH 5.5 and taken to the transfer flask (**figure 3.7**) where the sample is mixed.

The sample is transferred to a dedicated on-line reverse phase hplc system for separation and quantification, with a sample loop of 200 µL capacity, in 95% (v/v) 50 mM sodium acetate, pH 5.4, 5% (70% v/v) acetonitrile/32 mM sodium acetate, pH 6.2. A gradient is set up to elute the amino acids (**table 3.2**). PTC-amino acids are detected by absorption of UV light at 254 nm, with the amount of absorption being correlated with concentration of the amino acid. A chromatogram is produced by plotting absorption and elution time (**figure 3.1**). Amino acids are quantified by comparison with a calibration file set up by analysis of amino acids of known concentration (Pierce Standard H).

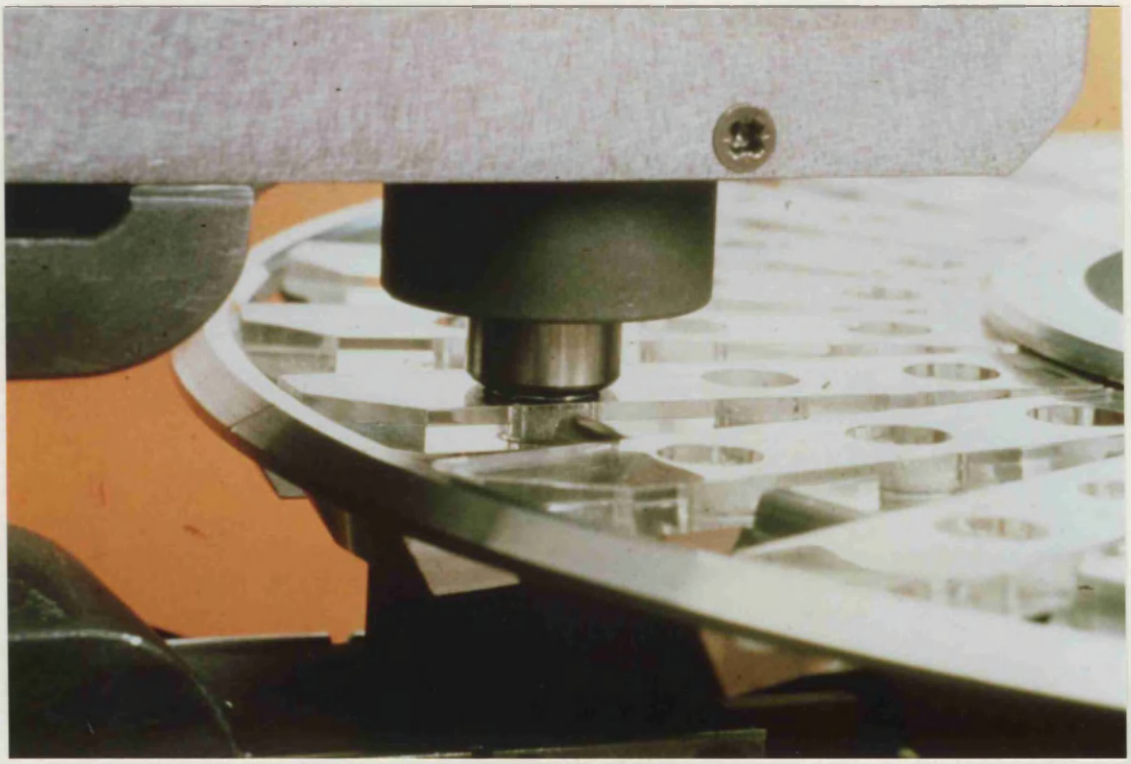


Figure 3.6: The derivatization head of the 420H amino acid analyser completes one analysis at a time, each cycle lasting approximately 45 minutes. The jaws seal to form a reaction chamber for the process, which ends with the sample being dissolved and taken to the transfer flask (**figure 3.7**). Photo credit: Applied Biosystems.

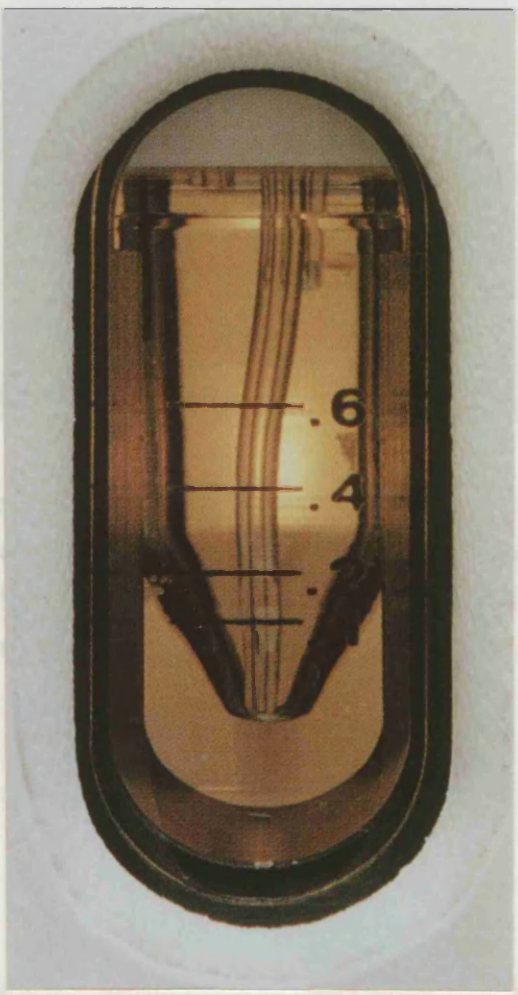


Figure 3.7: The tranfer flask. After derivatization, the sample is mixed in the transfer flask and 200 µL injected onto the analysis column. Photo credit: Applied Biosystems.

Time (minutes)	%B
0.0	5
4.0	15
10.0	32
20.0	63
25.0	100
30.0	100
31.0	5

Flow rate = 300 µL/minute
Column temperature = 37°C
Column type PTC C-18 (Applied Biosystems)

Table 3.2: Gradient used to elute PTC-amino acids from the hplc system of the amino acid analyser.

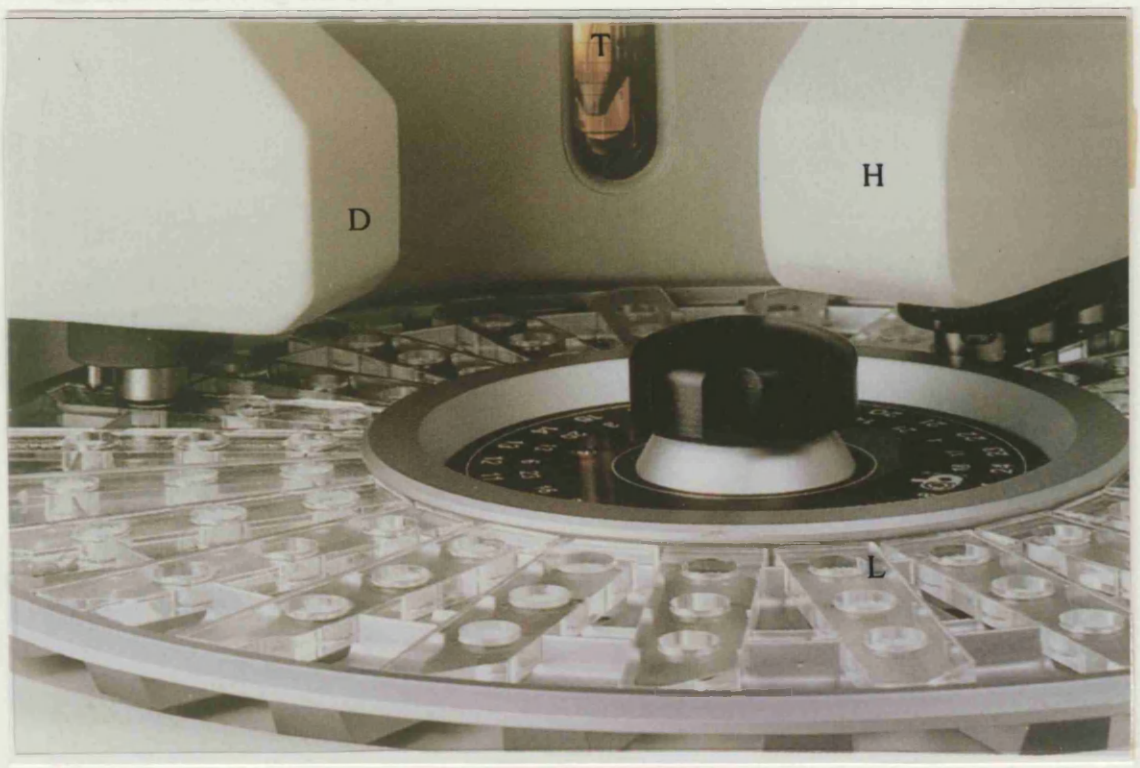


Figure 3.8: View of the analysis system. The sample is loaded at the front (L), taken to the hydrolysis head (H) and then to the derivatization head (D) by rotation of the turntable. The sample is then taken to the transfer flask (T) prior to injection on the on-line hplc system for separation. Photo credit: Applied Biosystems.

3.9 Materials

Item	Supplier
<u>Sample preparation</u>	
Sodium hypochlorite	Boots plc
MilliQ™ Pure water	Millipore
Centrifuge tubes, 50 mL capacity	Falcon
Centrifuge tubes, 500 mL capacity	Heraeus
Suspension mixer	Luckham
Centrifuge, Omnifuge	Heraeus
Centrifuge, Suprafuge	Heraeus
<u>Decalcification</u>	
Ethylene diamino tetra acetic acid disodium salt (Na_2-EDTA)	Boehringer Mannheim
Sodium hydroxide	BDH Chemicals Ltd.
Formic acid	Rhône-Poulenc Ltd.
Acetic acid	Rhône-Poulenc Ltd.
Hydrochloric acid	Applied Biosystems
Centrifuge tube, 1.5 mL capacity	Scotlab
Microcentrifuge shaker	Eppendorf
Centrifuge, Microcentaur	Scotlab
<u>Protein/peptide purification</u>	
PM10 filter	Amicon
Amicon ultrafiltration system	Amicon
Minitan™ tangential flow system	Millipore
Pure water	Millipore MilliQ™
Ammonium acetate	BDH Chemicals Ltd.
Sodium hydroxide	BDH Chemicals Ltd.
Minicon™ concentrator	Amicon
<u>Protein/peptide separation</u>	
Waters 650 hplc/fplc system	Millipore Waters
Pure water	Millipore MilliQ™
Gel filtration column (Superose 6)	Pharmacia
Reverse phase column (Aquapore™)	Applied Biosystems
Trifluoro acetic acid (TFA)	Applied Biosystems
Mini Protean II system	BioRad
Acrylamide	Sigma

Bis-acrylamide	BioRad
N, N, N' ,N' tetramethylethylendiamine (TEMED)	Sigma
(Tris [hydroxymethyl]-aminomethane)	Sigma
Sodium dodecyl sulphate (SDS)	Sigma
Glycerol	Sigma
Mercaptoethanol	Sigma
Tricine	Sigma
Methanol	Rhône-Poulenc Ltd.
Acetic acid	Rhône-Poulenc Ltd.
Coomassie Brilliant Blue R-250 (CBB)	Sigma
Pure water	Millipore MilliQ™
Auto pipettes	Rainin (Scotlab)
Pipette tips; 2.5 mL, 1 mL, 250 µL	Scotlab
Iso-propyl alcohol	Scotlab
Ammonium persulphate (APS)	BioRad

<u>Amino acid analysis</u>

Diisopropylethylamine (DIEA)	Applied Biosystems
Phenylisothiocyanate (PITC)	Applied Biosystems
Methanol	Applied Biosystems
Ethylene diamino tetraaceticacid tripotassium salt (K_3-EDTA)	Applied Biosystems
3M Sodium acetate buffer, pH 3.8	Applied Biosystems
3M Sodium acetate buffer, pH 5.5	Applied Biosystems
6N HCl	Applied Biosystems
Amino acid standard H	Pierce chemicals
Hydrolysis test peptide	Applied Biosystems
Acetonitrile (Methyl cyanate)	Applied Biosystems
420H Amino acid analyser	Applied Biosystems
Hydrolysis bottles	Pierce
Hydrolysis caps	Pierce
Teflon inserts	Pierce tuf bond
ProBlott™ PVDF membrane	Applied Biosystems

<u>Finger tip amino acids</u>

Citric acid	Sigma
25 Well plates	Cel-cult
Pure water	Millipore MilliQ™
Water	Millipore RO™
Suspension mixer	Luckham Model 802

Sediment amino acids

Pure water	Millipore MilliQ™
Glass Universal Bottles, 20 mL capacity	Sterilin
Centrifuge tubes, 1.5 mL capacity	Scotlab
Rotary evaporator (Gyrovap)	Howe

Computer manipulation

Software for the Macintosh microcomputer

DataDesk Professional™ 3.0	Odesta Corporation
Excel™ 1.5.1	Microsoft Corporation

Amino Acids as contaminants

4.1 Introduction

Amino acids are ubiquitous in living organisms as the building blocks of proteins. The possibility of contamination of samples undergoing molecular analysis is an ever present problem in molecular palaeontology, as amino acids may be introduced to samples from sources such as micro-organisms, the sediment which surrounds the fossil sample (through adhesion and percolating groundwaters) and from the fingertips of the analyst, at any stage in the preparation of the samples.

The information present in the amino acids and polypeptides trapped within the shells of fossil organisms represent a wealth of data regarding both the taxonomic relationships between different fossils and also conditions in which the organism lived. However, due to the ubiquity of the molecules in nature, it is necessary to be certain that any information gained from fossils is representative of the sample, and also that non indigenous amino acids have not been introduced to the sample.

The majority of the data in this chapter has been published in a different form by Walton and Curry (1991) in the journal *Palaeontology*. The aim of the study was to investigate the abundance (both absolute and relative) of amino acids present as preserved fossil polypeptides from fossil *Neothyris* brachiopods, as molecules adsorbed onto sedimentary particles, and also as molecules sourced from human fingertips. Once the molecules have been quantified, a means of manipulation of the data is necessary in order to discriminate between the sample types.

The question of contamination by fingertip amino acids was first raised in 1965 (Hamilton, 1965; Oró and Skewes, 1965), but investigation of this phenomenon has not kept pace with the incremental increase in the technology and sensitivity of amino acid analysis. Amino acids were collected from human fingertips in order to assess potential contamination of samples, as a prelude to the detailed investigation of both Recent and fossilised amino acids using the same equipment (**Chapters Five, Six and Seven**). As all subsequent sample preparation would take place using latex laboratory gloves, a sample sourced from the finger of the gloves used in the collection of samples was included. It was hoped that identification and understanding of the characteristics of potential contaminants, when compared with data generated from fossils and sediments using an identical high-sensitivity analysis system, would allow the recognition of such contamination and hence eradicate it as a potential distortion of our analyses.

4.2 Methods

Amino acids were sampled from each source as described earlier (**sections 3.1,**

3.4 and 3.5). Loadings on the analyser were as follows:

Fossil samples	20 µL
Fingertips	30 µL
Sediments	20 µL

Amino acids were analysed with derivatization only for fingertips, but with both hydrolysis and derivatization for fossils and sediments (**section 3.6**).

4.3 Results

4.3.1 Absolute abundance of amino acids

In absolute terms the quantities of amino acids that can be transmitted by finger-tips are considerably greater than those present in fossil shells and in sediments (**table 4.1**; **figure 4.1**). Not surprisingly, there is slight variation in the yields for individual amino acids in some samples, but typically the levels for finger-tips are 10 times higher than in sediments, and 100 times higher than in fossils (e.g. Gly, **table 4.1**). Such levels are probably reasonably realistic assessments of the amounts of free amino acids which would be transferred to fossils or sediments by touching them with wet fingers. As the samples were not hydrolysed, amino acids still attached to one another in proteins from the skin would not have been quantified; small fragments of human skin are constantly being shed, and there is no doubt that the potential levels of amino acids transferred by finger contact would be greater than that measured here. Any contamination occurring at a critical stage of sample preparation will clearly overwhelm the indigenous amino acids, and lead to an obvious, and false, high concentration of amino acids present within a sample.

Amino Acid	***Neothyris* (4)**		**Finger tips (8)**		**Sediments (3)**	
	mean	**S.D.**	**mean**	**S.D.**	**mean**	**S.D.**
Aspartic acid	0.659	0.450	18.347	15.170	0.000	0.000
Glutamic acid	0.161	0.058	16.020	11.470	0.310	0.197
Serine	0.168	0.089	38.772	44.526	1.091	0.303
Glycine	0.505	0.143	56.211	19.670	5.978	0.768
Alanine	0.211	0.164	22.337	7.383	5.221	1.349
Valine	0.105	0.057	9.323	3.800	3.200	0.910

Table 4.1: Results of amino acid analyses (nmol/g). Figures in brackets refer to sample size.

4.3.2 Relative abundance of amino acids

Amino acid data is most often presented in the form of mole percentages which overcomes the difficulty of comparing samples of varying size. **Figure 4.2** shows a re-

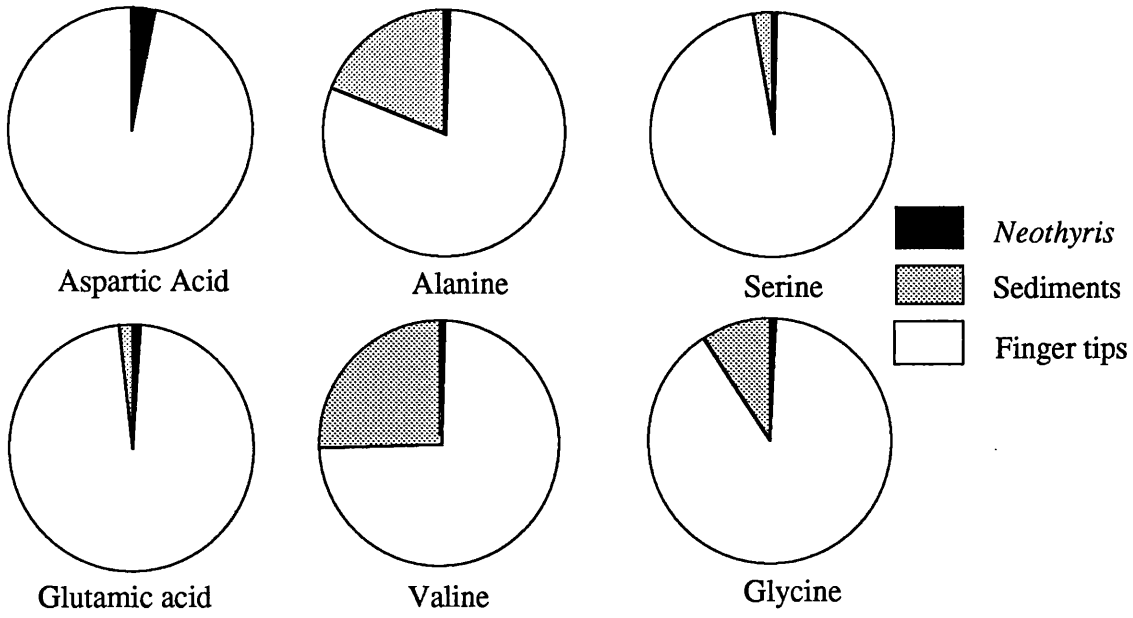


Figure 4.1: Pie charts showing the absolute proportions of selected amino acids (normalised to nmol/g) from the shells of Plio-Pleistocene brachiopods, sediments (from the Plio-Pleistocene South Wanganui Basin of New Zealand), and from finger-tips (1 mL ≈ 1 g).

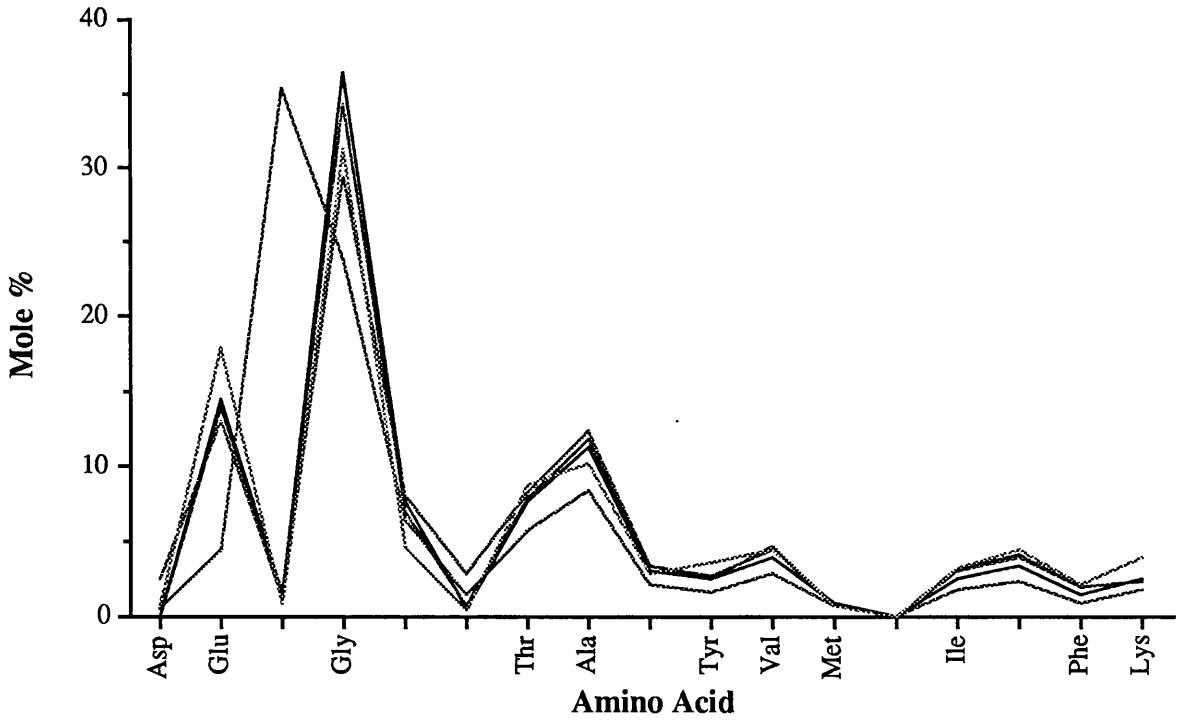


Figure 4.2: Relative proportions (Mole%) of the common amino acids from the finger tips of five individuals from the first run of the experiment.

markably consistent pattern in the relative abundance of finger-tip amino acids, with all individuals being rich in either Ser or Gly. These results show some similarity with those of Oró and Skewes (1965), where Ser is the most common amino acid. In sediments, Ala and Gly are also the most abundant (**figure 4.3**), while Gly is also a major constituent of fossil *Neothyris* shells (**figure 4.4**). The relative proportions of amino acids are quite consistent within all three groups sampled (**figures. 4.2, 4.3 and 4.4**), to the extent that incomplete removal of EDTA from fossil samples becomes readily apparent (**figure 4.4**). Apart from 25019001A, the fossil *Neothyris* all have consistent amino acid profiles, leading to a suggestion of similarity between the preserved peptides.

Superficially there are a number of common features between these three datasets, but such subjective assessments are potentially misleading To produce a more rigorous assessment of the relative similarity between samples, hierarchical cluster analysis using the statistical programme DATADESK™ on a Macintosh micro-computer was applied to the relative abundance data. The resulting cluster diagram resolves the finger-tip, sediment and fossil shell amino acid data into discrete clusters (**figure 4.5**). The citrate buffer and tap water have been included as outgroups. Samples of sediment from the majority of fossil containing horizons were analysed to determine both their amino acid concentration and distribution, to allow comparison with the amino acid distribution from fossils.

Principal component analysis (PCA) also using DATADESK™ was also applied to the datasets (**figure 4.6**), and this confirms the groupings of the cluster plot. A useful attribute of PCA is that it is possible to use the eigen-vector values to determine

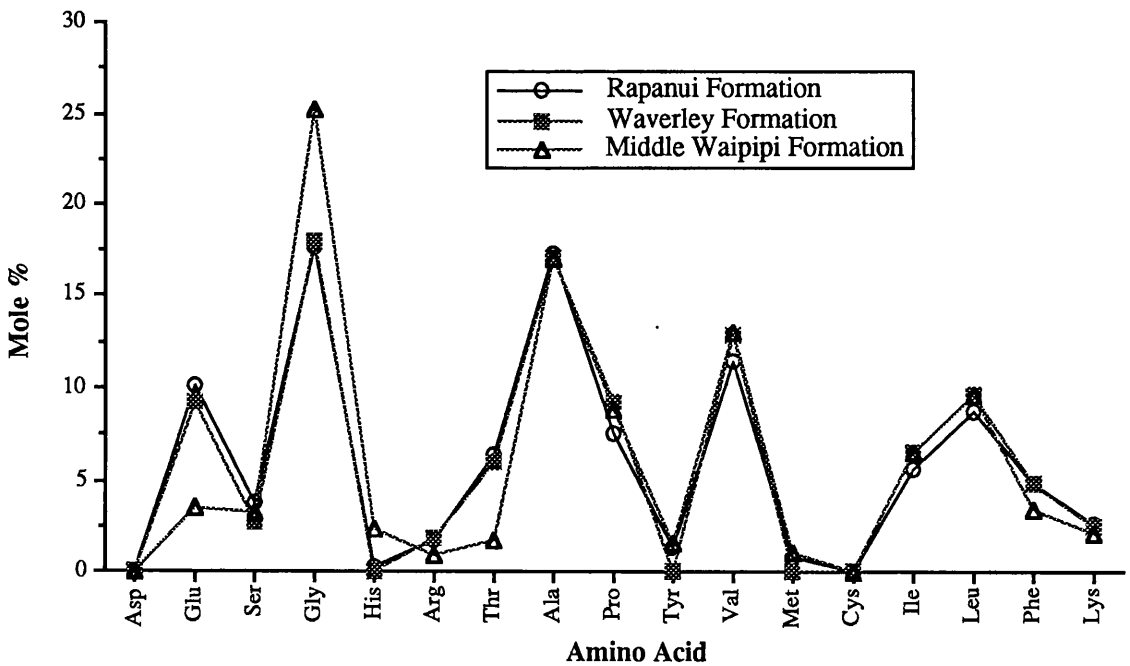


Figure 4.3: Relative proportion (Mole%) of the common amino acids for New Zealand sediments.

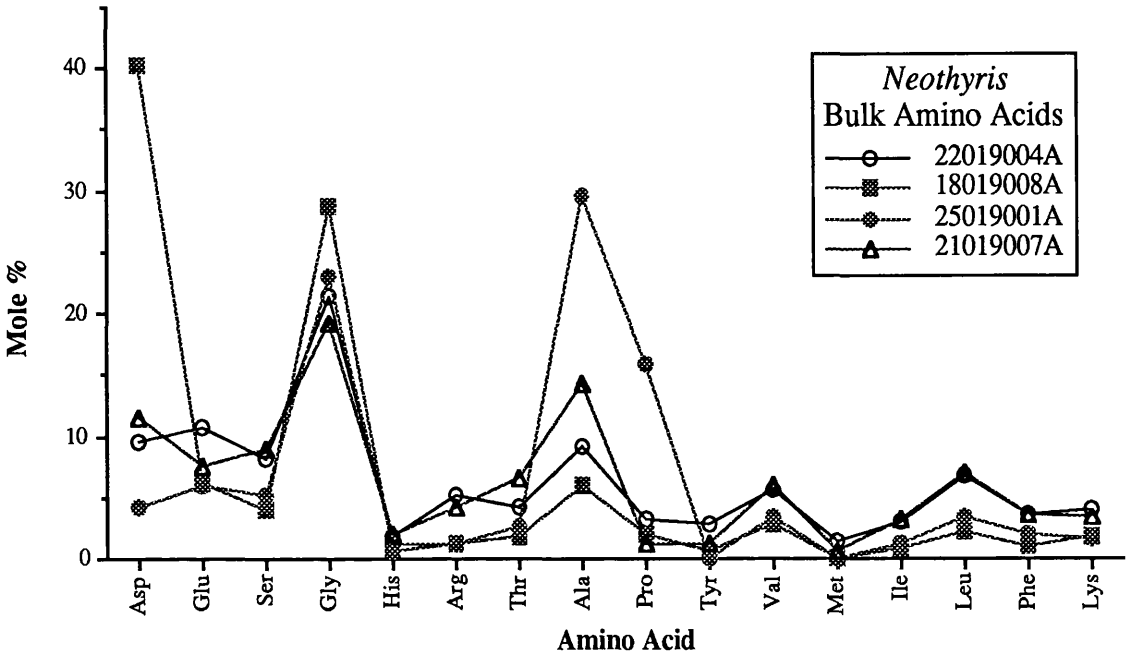


Figure 4.4: Realtive proportions (Mole%) of the amino acids from the shells of fossil *Neothyris*. Sample 25019001A shows the effect of incomplete removal of EDTA, a phenomenum which provides a check on the prepartion purity of the sample.

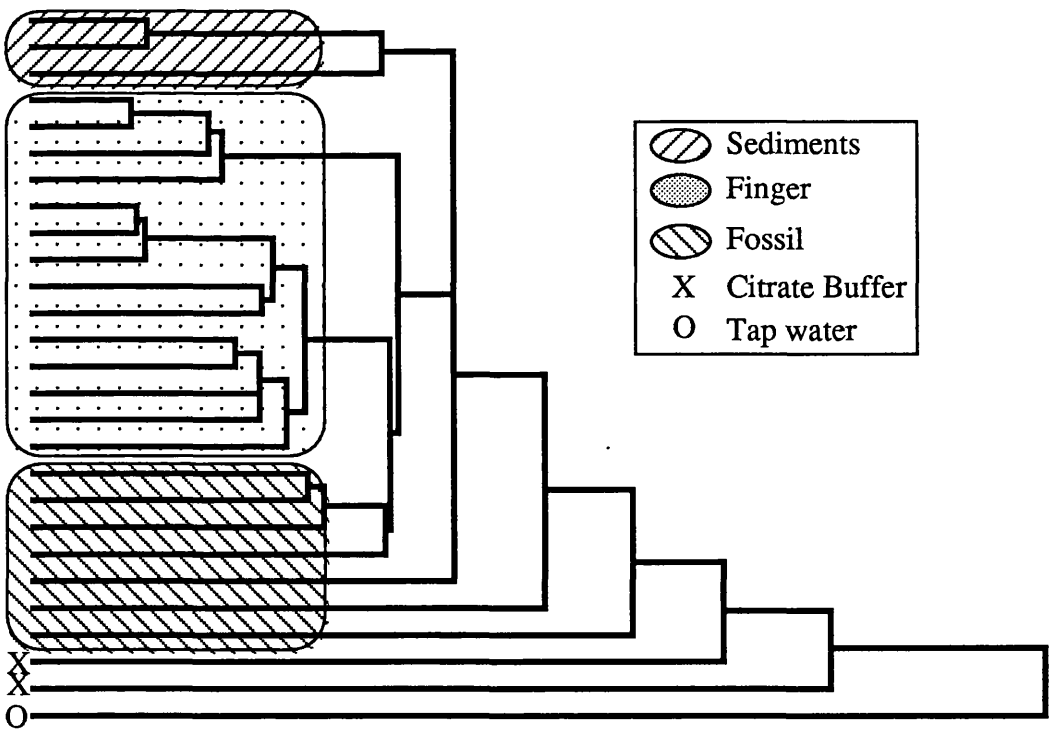


Figure 4.5: Single linkage cluster analysis (representing nearest neighbour groupings) for the relative proportion data (Mole%), generated by DATADESK™.

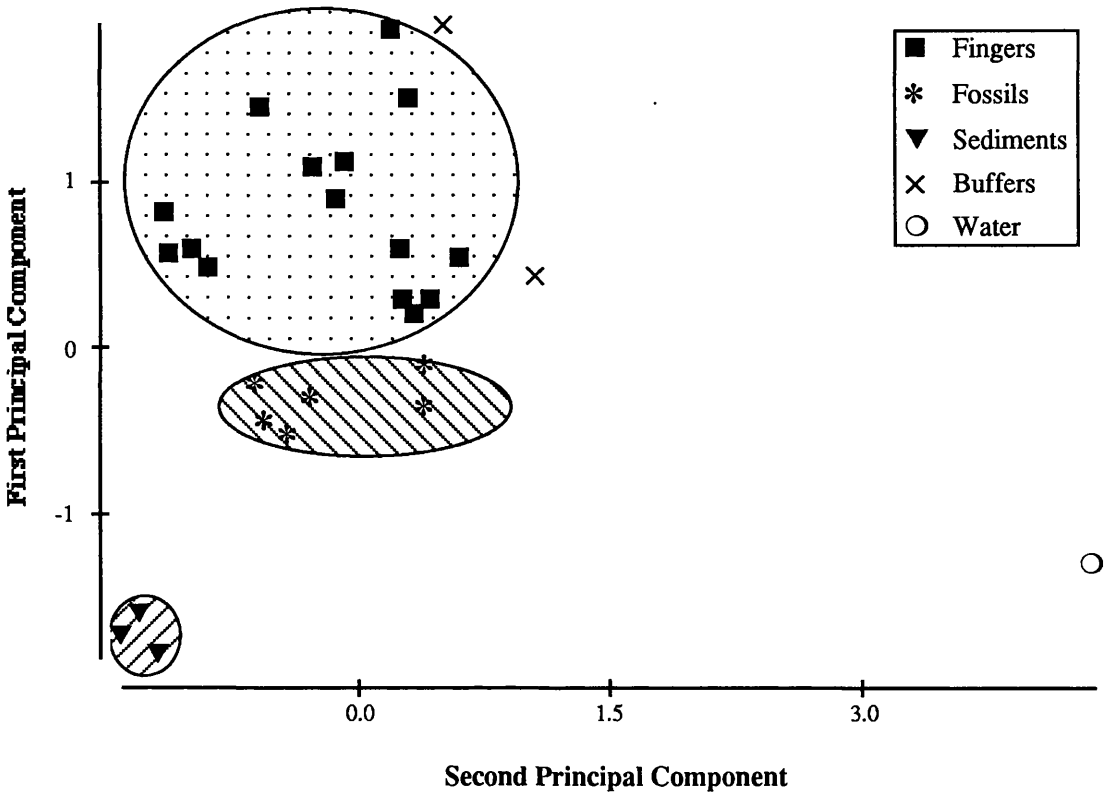


Figure 4.6: Scatter plot of the first two principal components for relative proportion data (Mole%), generated by DATADESK™. These data do not take into account variations in the actual concentration of amino acids.

which elements (in this case individual amino acids) are most important in distinguishing between the different groups. **Figure 4.6** shows that the clear resolution between sediments, fossils, and fingers is achieved along the first principal component axis, which is aligned in the direction of maximum variability and in this case contributes 26.9% of the total variation detected. Examining the eigenvector values from this analysis, it is clear that this differentiation primarily reflects high scores for Val (-0.377), Ile (-0.323), Leu (-0.347) and Phe (-0.345). In all cases, the distribution of amino acids from the sediment was very different to that of the fossil data, illustrated by the Kupe Fm (**figure 4.7**), and hence will not be considered further here.

4.4 Discussion

In one respect it is not surprising to discover a clear differentiation between these three groups of samples. Finger-tip amino acids are predominantly derived from human skin proteins. Amino acids from sediments presumably come from a complex mixture of sources including the breakdown of various life-forms which originally lived in or on the sediment and mobile amino acids carried in by percolating ground waters. The fossil amino acids considered here have been extracted from intracrystalline macromol-

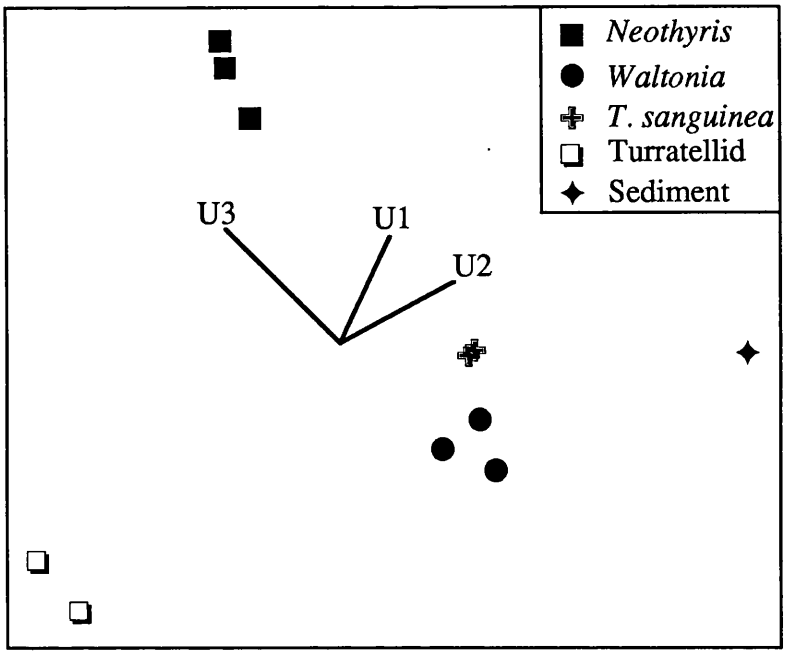


Figure 4.7: Graph of the first three principal components for relative proportion data for fossils and sediments for the Kupe Fm, showing that the fossil samples are not contaminated by the sediment.

ecules; most published data from shells has also included a component of intercrystalline macromolecules (e.g. Jope, 1967a, 1967b; Weiner *et al.*, 1976), which are not protected by shell crystallites, and hence may be easily degraded or contaminated. It is very encouraging for the study of amino acids in these New Zealand successions that the finger-tip data can be so readily distinguished from the fossil and sediment data, both in terms of absolute abundance and mole percentages. Further samples of fossil shells have been analysed, and a similar pattern of differentiation from finger-tip and sediment data is apparent.

In practical terms, the use of clustering and statistical approaches represents an important and rapid method of checking for human contamination in geological samples. As the absolute abundance of finger-tip amino acids is so much greater than that of fossils, any contamination should produce an analysis which clusters with the former rather than the latter, and this is probably also true for sediments although to a lesser extent. The ease of statistical discrimination between the fossil samples and their potential contaminants indicates that:

(i) There has been no homogenisation of the amino acids through the horizon i.e. the amino acids are not evenly distributed throughout the sediment and the fossils.

(ii) There is no significant contamination of the fossil data from sediment adhering to the shell samples.

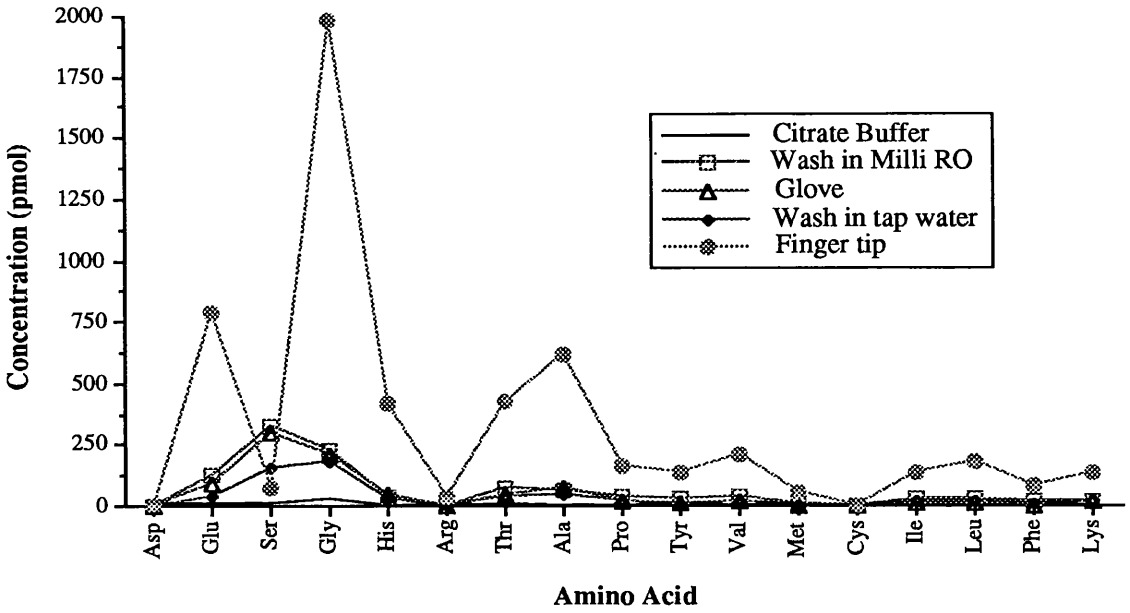


Figure 4.8: Absolute properties (un-normalised, the concentrations (pmol) are for the amino acids in 30 µL of buffer) showing the potential levels of contamination from latex gloves in comparison to finger tip data).

The results of this investigation have also provided important information for sample preparation. High levels of contamination from finger-tip amino acids are to be expected, but these experiments have shown that even hands which have been washed prior to handling samples could be a significant source of contamination. Much more surprising, and potentially worrying, is the discovery that even latex laboratory gloves can be a source of contaminant amino acids (**figure 4.8**). Latex gloves are widely used in laboratories, but obviously they give a false sense of security with modern high-sensitivity amino acid analysers capable of detecting a few picomoles (10^{-12} mole) of each amino acid. For optimum recovery of uncontaminated indigenous amino acids from fossils and sediments, it is necessary to minimize or completely eradicate any handling of the sample or of the interior surfaces of any vessel used in the release of the incarcerated amino acids.

Chapter Five: Amino acids from Recent New Zealand Brachiopods

Amino acids from Recent New Zealand brachiopods

5.1 Introduction

The intracrystalline organic molecules which are present as a mixture within the shells of Recent brachiopods (Collins *et al.*, 1988; **Chapter Three**) are rich sources of information regarding the taxonomy of the organisms (Collins *et al.*, 1988; 1991c; Curry *et al.*, 1991c). The molecules consist mostly of protein, although quantities of lipids (Curry *et al.*, 1991b), alcohols (H. Clegg, pers. comm.) and carbohydrate (Collins *et al.*, 1991a) are also present. Neither the function of these molecules, nor their role, if any, in the process of biomineralisation is clear, although it is beyond doubt that these molecules are of great significance in the study of molecular palaeontology.

Of the molecules present within the shell, neither lipids, alcohols nor carbohydrates contain information of any significant taxonomic value. Proteins, however, contain taxonomic information which is derived from the DNA, and therefore the genes, of the organism. The proteins are composed of relatively stable monomer units, amino acids, which are present in a defined order along the chain of the molecule. It is this information, the primary sequence of the protein, which is conventionally used for biochemical comparison of samples, as differences in the nucleotide bases in the DNA of organisms will be reflected in differences in the order of amino acids (**Chapter One**). The analysis of the **position** of the changes in amino acids or nucleotide bases can only be completed by sequencing either the DNA or the proteins of the organism. The complete sequencing of proteins is a difficult and time consuming process, even using the automated sequencers available with current technology, it is rare to be able to sequence more than 50 amino acids at any one time (counting from the *N*-terminus). Enzymatic and/or chemical cleavage techniques (**figure 6.1**) or DNA probes are necessary in order to proceed further via internal sequencing and to elucidate further sequence information, although there is no guarantee of success in either of these procedures.

Recent successes in the partial sequencing, from the *N*-terminus, of intracrystalline proteins from brachiopods (Curry *et al.*, 1991a; Cusack *et al.*, 1992) have shown that the sequence of proteins do not correspond to that of any other known protein, and also that they are homologous in different species. The differences in the sequence of a small (6.5 kDa) protein sequenced in three New Zealand genera (Cusack *et al.*, 1992) shows few substitutions between the proteins. However, as yet, none of these examples has completed the sequencing of more than a quarter of the total residues contained within the protein.

Differences in the DNA between species will, however, also be reflected by different relative proportions of amino acids within a sample. This will reduce to some

degree the level of information available from the proteins, but this is compensated for by the much shorter length of time required for amino acid analysis, compared to that required for primary sequence data. Amino acid quantification is rapid, and is still able to identify variation, and therefore differences in the composition of the proteins of a sample, although the location of this variation cannot be determined. The relative abundance of each of the amino acids within each sample is a representation of their relative frequency of occurrence within the polypeptides of the sample, and hence differences between the relative abundance will equate to the minimum number of differences between the codons of the DNA of the organisms being compared (some changes in nucleotide base will not result in a changed amino acid residue in the protein, see **figures 1.12, 1.13**).

It was with the inherent difficulties of protein sequencing in mind that the intracrystalline proteins from Recent brachiopods were characterised by their amino acid composition. This information was examined by multivariate statistical techniques in order to extract taxonomic data from the information contained within the molecules. Values obtained by the study of extant species can also be used as a baseline (or time = 0) for studies of molecules contained within shells of the same species in the fossil record (**Chapters Six and Seven**).

Other studies have shown that the number of proteins contained within the shell also varies taxonomically (Curry *et al.*, 1991b). In order to provide a method of secondary confirmation of the taxonomic conclusions of the amino acid technique, several of the samples were extracted by Na_2-EDTA decalcification and then fractionated using SDS PAGE (**Chapter Three**). As this method requires the use of large amounts of shell material, *Terebratella haurakiensis*, *Gyrothyris mawsoni* and *Neothyris parva*, could not be analysed in this way. This method will reveal which species have a similar number and size of proteins present within their shells, which provides a backup in case of spurious relationships derived from the amino acid data.

5.2 Incorporation of molecules within the shell

Intracrystalline molecules were sealed into the shell during mineralisation of the shell carbonate and are therefore in a location inaccessible to micro-organisms and other contaminants, until the shell is decalcified. The molecules may be either active or passive in the process of biomineralisation, acting either as a nucleation site for mineralisation ("biologically induced biomineralization", Lowenstam, 1981), or being passively incorporated into the shell during growth. The cells of the outer epithelium are intimately associated with the shell crystallites (Williams, 1968), and it is likely that the protein incorporated in the shell crystal will be sourced from these cells.

The function of intracrystalline proteins has mainly been studied with regard to those from the spicules of sea urchins (Berman *et al.*, 1990), where it has been suggested

that the organic molecules provide strength for the crystal aggregate. The presence of molecules within sea urchin spicules also alters the material properties of the calcite (Berman *et al.*, 1988). These spicules have a conchoidal fracture, rather than following the regular fracture of an inorganic calcite crystal. This conchoidal fracture in sea urchin spicules is analogous to the fracture of the secondary shell fibres from brachiopod shells (G.B. Curry, pers. comm.), indicating that the molecules may have a similar effect on the calcite of this shell.

As this study is not into the process of biomineralisation, neither the procedure by which the molecules become trapped nor their function have been studied. It is only the taxonomic information which may be gained from these molecules which is considered in the present study. However, further studies of these molecules from extant samples may well reveal aspects important to this process.

5.3 Methods

Samples of living and Recent New Zealand brachiopods were collected from the locations given in **figure 2.3**. Living samples were killed by dehydration. Intracrystalline molecules were extracted from the shells following the method outlined in **Chapter Three**. In summary, the shells were cleaned, crushed and bleached, and the powder then washed, lyophilised and freeze dried. Incarcerated molecules were released by incubation with 2N HCl at a ratio of 11 μL acid per mg shell powder, and loaded onto the analyser for quantification. Samples were analysed both with and without hydrolysis, to determine the concentration of unbound (free) amino acids present within the shell. To assess the proportion of the total shell protein which is due to the intracrystalline fraction, samples of *Neothyris lenticularis*, *Waltonia inconspicua* and *Liothyrella neozealandica* were decalcified without prior bleaching. Samples for SDS PAGE separation were decalcified by incubation with Na_2EDTA at a ratio of 23 mL/g, and the EDTA/calcium complex removed by filtration using the Minitan™ system and the sample concentrated in a Minicon™ static concentrator.

Concentrations of amino acids given by the analyses were converted to weight percentages (wt%) using the EXCEL™ spreadsheet package, before being analysed by the statistical program DATADESK™ on the Macintosh microcomputer. The statistical methods used were cluster analysis and principal component analysis (PCA; **Chapter One**).

The polyacrylamide gels for SDS PAGE were prepared following the methods in **Chapter Three**. Proteins were separated by electrophoresis at 50 V constant, until the tracking dye reached the base of the housing. This lower voltage has the advantage of sharpening the boundaries of the protein band. Proteins were visualised by incubation with CBB-R250 stain. The extract from the red brachiopod shells is deep red (Cusack *et al.*, 1992), a feature which is due to the colour molecule attached to the intracrystalline

protein. From the black shelled *Notosaria nigricans*, however, the solution was colourless once the insoluble compounds were removed, indicating that the colouration of the shell is due to molecules present either in the insoluble fraction, or within the intercrystalline fraction, rather than in the intracrystalline fraction, as is the case for all red brachiopods examined thus far.

5.4 Results

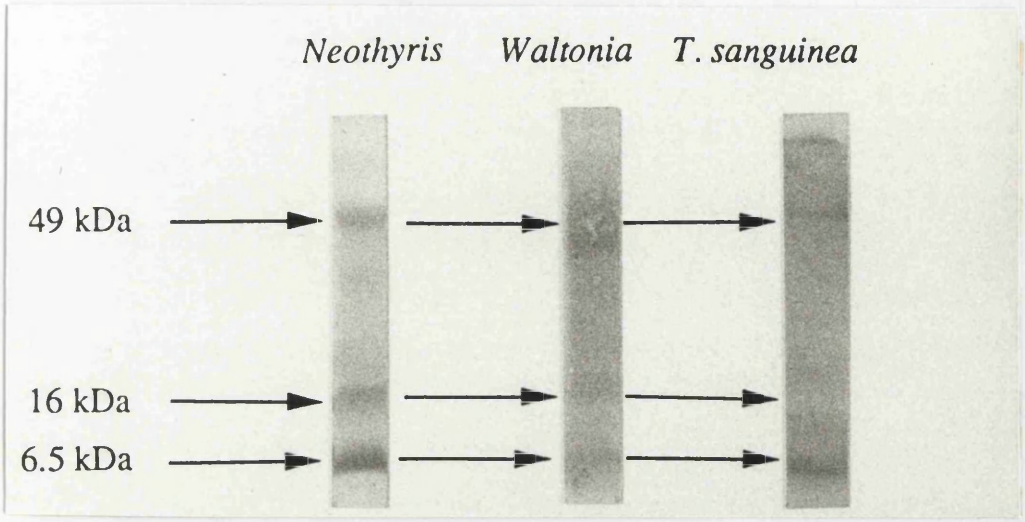
5.4.1 Gel electrophoresis

A composite gel of the samples separated by SDS PAGE is shown in **figure 5.1**. Samples are grouped according to their taxonomic positions given by Williams *et al.* (1965).

Neothyris lenticularis, *W. inconspicua* and *T. sanguinea* (Order Terebratulida, Suborder Terebratellida) all show several major bands at approximate molecular weights of 47 kDa, 16 kDa and 6.5 kDa. *L. neozealandica* (Order Terebratulida, Suborder Terebratulidina) has 4 bands at the approximate molecular weight of 46 kDa, 22 kDa, 16 kDa and 6.5 kDa. *N. nigricans* (Order Rhynchonellida) has 4 bands at approximate molecular weights of 50 kDa, 20 kDa, 17 kDa and 14 kDa. The number and size of the proteins contained within the shell varies taxonomically at the subordinal level, providing a low sensitivity method of taxonomic discrimination.

5.4.2 Absolute abundance of amino acids

Amino acid analysis of the samples that were not bleached show that the molecules extracted from the intracrystalline fraction account in all three cases for between 30 and 40% of the total amino acid present within the shell. Recent brachiopods contain, on average, less than 10% of the total amino acid present within the shell in the form of free amino acids (**table 5.1**), indicating that the vast majority are combined into proteins and peptides. A notable exception to this is Tyr in *N. nigricans*, which is c. 75% free (although this could also be due to the co-elution of another, as yet unidentified, molecule, as this property is not seen in any of the other samples under investigation). As there are very few free amino acids in these Recent samples, this also indicates that the acid decalcification does not hydrolyse many (if any) of the peptide bonds in the protein. The absolute abundance of all amino acids (free and combined) varies taxonomically, ranging between 70 and 800 ng/mg (amino acid/shell), equivalent to 0.007% and 0.08% of the total weight of the shell. *L. neozealandica*, *N. lenticularis* and *W. inconspicua* all contain relatively low concentrations of amino acid, which contrasts with the high concentrations found in *N. nigricans*, a feature comparable to the values found for the total organic matter contained within the shell (Curry *et al.*, 1989), where *N. nigricans* contained twice the amount of organic matter compared to *W. inconspicua*.



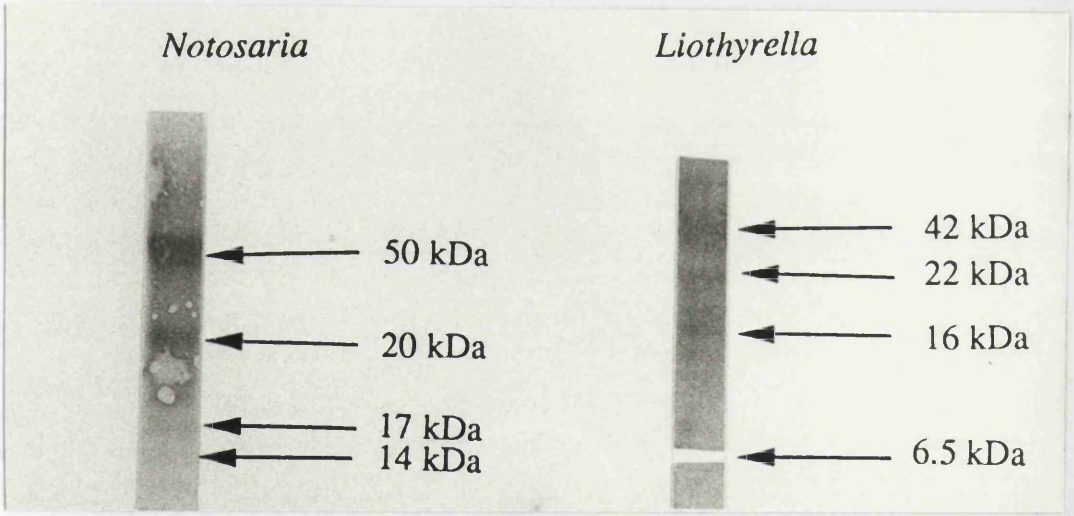


Figure 5.1: Major protein bands from the >10 kDa fraction of the organic extract of Recent brachiopods. Note the differing numbers and sizes of the protein bands in the samples.

Sample		D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Waltonia inconspicua	Absolute	8.23	9.82	6.94	76.76	6.70	3.74	11.66	12.85	2.94	10.74	4.31	3.96	2.22	3.58
	Relative	4.89	5.96	4.17	45.03	3.86	2.27	7.14	7.93	1.72	6.63	3.52	3.42	1.25	2.41
	%Free	0	0	0	1.86	0	0	0	0	0	2.89	0	33.96	0	0
Terebratella sanguinea	Absolute	12.12	13.51	7.66	69.70	6.28	6.58	12.24	17.32	5.42	12.82	7.39	7.69	4.15	7.42
	Relative	6.37	7.10	4.03	36.63	3.30	3.46	6.43	9.10	2.85	6.74	3.88	4.04	2.18	3.90
	%Free	5.45	0	16.13	5.24	0	0	4.25	4.45	0	0	0	11.64	15.18	0
Terebratella haurakiensis	Absolute	19.78	22.06	20.60	132.04	15.79	8.73	30.02	29.70	12.34	24.25	16.26	17.10	14.06	8.82
	Relative	5.32	5.94	5.54	35.54	4.25	2.35	8.08	7.99	3.32	6.53	4.38	4.60	3.78	2.37
	%Free	0	0	0	2.63	0	0	0	0	0	1.77	0	2.57	0	0
Neothyris lenticularis	Absolute	6.76	8.35	3.54	56.06	3.15	2.69	5.18	7.74	1.39	6.16	1.78	2.92	1.70	1.47
	Relative	6.21	7.67	3.25	51.51	2.89	2.47	4.76	7.11	1.28	5.66	1.64	2.68	1.56	1.35
	%Free	9.76	0	0	5.05	0	0	5.21	0	0	0	0	16.10	0	21.77
Neothyris parva	Absolute	4.23	5.34	2.46	34.76	1.63	1.58	2.22	3.47	0.34	3.43	1.41	1.48	0.93	0.43
	Relative	6.66	8.37	3.84	54.76	2.54	2.47	3.46	5.43	0.54	5.37	2.20	2.29	1.44	0.66
	%Free	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Notosaria nigricans	Absolute	214.52	14.35	11.81	218.20	9.25	5.89	16.08	14.96	43.22	7.90	2.81	3.96	13.45	4.39
	Relative	36.94	2.47	2.03	37.57	1.59	1.01	2.77	2.58	7.44	1.36	0.48	0.68	2.32	0.76
	%Free	0.41	0	0	1.04	0	0	2.24	2.74	74.32	3.16	0	9.85	4.01	0
Gyrothyris mawsoni	Absolute	10.29	14.16	6.86	71.15	8.81	4.86	12.37	13.37	3.92	9.66	5.31	5.53	3.10	2.9
	Relative	5.97	8.21	3.98	41.28	5.11	2.82	7.18	7.76	2.27	5.60	3.08	3.21	1.80	1.73
	%Free	0	0	0	2.05	0	0	2.26	0	0	0	0	6.33	0	0
Liothyrella neozealandica	Absolute	7.06	4.91	2.11	27.89	1.36	1.93	3.12	3.62	0.41	4.88	1.37	1.50	0.54	0.65
	Relative	11.51	8.00	3.44	45.46	2.22	3.15	5.09	5.90	0.67	7.95	2.23	2.44	0.88	1.06
	%Free	0	0	0	6.49	0	0	0	0	0	0	0	0	0	56.92

Table 5.1: Amino acid composition of the total organic extract from Recent brachiopods. Absolute concentrations in ng/mg, relative proportion in wt%.

Decalcification of the *Notosaria* shell powder, in contrast to the remaining samples, left a black insoluble residue. This is likely to be caused by the differing nature of the shell protein contained in the shell matrix of this species, which consists partially of an acid insoluble fraction, which is more resistant to the oxidative effect of the exposure to sodium hypochlorite (Collins *et al.*, 1991c). These insoluble compounds were removed by centrifugation prior to amino acid analysis of the soluble fraction.

5.4.3 Relative abundance of amino acids

To provide a basis for the comparison of the amino acids without the discrimination being solely due to the concentration of the molecules, some form of standardisation is necessary. In this study, the concentrations of the molecules were converted to weight percentages. These relative abundances are shown in **table 5.1**, which shows the variation in the amino acid content of the samples. The most striking variations lie in high Asp/Asn in *N. nigricans* (36.94 wt%) and *L. neozealandica* (11.51 wt%). *N. nigricans* also contains high Tyr (7.44 wt%, much of which is present in the free state), but low Leu (0.68 wt%) and Glu/Gln (2.47 wt%) in contrast to the other samples. *N. lenticularis* and *N. parva* contain high Gly (51.51 wt% and 54.76 wt% respectively). These proportions (i.e. high Gly, Ala and Asp/Asn) are comparable to those found in intercrystalline molecules from the same species of brachiopods (Jope, 1977; Kolesnikov and Prosorovskaya, 1986). The actual values are somewhat different, which is not surprising given that the molecules in this study are intracrystalline and are likely to be different to the matrix proteins considered in these other studies.

The low proportions of free amino acids within the samples (**table 5.1**) is important in two ways:

(i) This feature is indicative of intact protein and peptide survival within the biocrystal of Recent shells.

(ii) Low concentrations of free amino acid indicate the absence of contamination by sample handling (Walton and Curry, 1991; **Chapter Four**).

Glycine, the simplest of the amino acids, accounts for the highest proportion of the molecules within the shells, in all cases being higher than 25 wt% and ranging up to more than 50 wt% of the total. Glycine is also the most common of the amino acids found on human fingertips (**Chapter Four**). However, low concentrations of Gly are present in the free state, indicating that it is released by hydrolysis of the proteins and peptides, and not from contamination by sample handling.

Direct comparisons of the relative abundance of amino acids between samples are hard to make, as the overall change of so many variables is difficult to observe. The illustration of the scale and direction of variation of the amino acid content of the samples is not possible, as this would require multi-dimensional space which cannot be used. For this reason, multivariate statistical techniques were applied to the dataset in order to

summarise the variables and produce more meaningful information regarding the variation of the samples. PCA (**section 1.6.2**) is a method of analysing the dataset in a way which summarises it into derived variables, which may then be compared in a graphical form. **Figure 5.2** shows the plot of the first three principal components extracted from the dataset, which represent a summary of the data, so that most of the variation is contained within fewer, derived variables. The new dataset is plotted into a rotating plot, allowing three dimensional observation (**figure 5.2**), and onto a scatterplot (**figure 5.3**), and samples which are closely related plotted close together.

Amino acid distribution in intracrystalline proteins allows separation of the samples solely in terms of the variability in the dataset; i.e. an objective process where the differences between the genomes of the organisms is reflected by different amino acid compositions. Principal component analysis scores each amino acid in terms of how much of the total variation in the dataset is due to that variable (i.e. the size of the eigenvalue). The first principal component contains 47.8% (**table 5.2**) of the variability of the dataset, and this variation is mainly due Ser (-0.325), Ala (-0.349), Pro (-0.351), Ile (-0.361) and Leu (-0.372). The second principal component contains 27.4% of the variability, mainly due to Glu/Gln (0.437), Gly (0.332), Tyr (-0.500), Val (0.362) and Phe (-0.362). The third principal component contains 10.5% of the variation, due mainly to Asp/Asn (-0.304), Gly

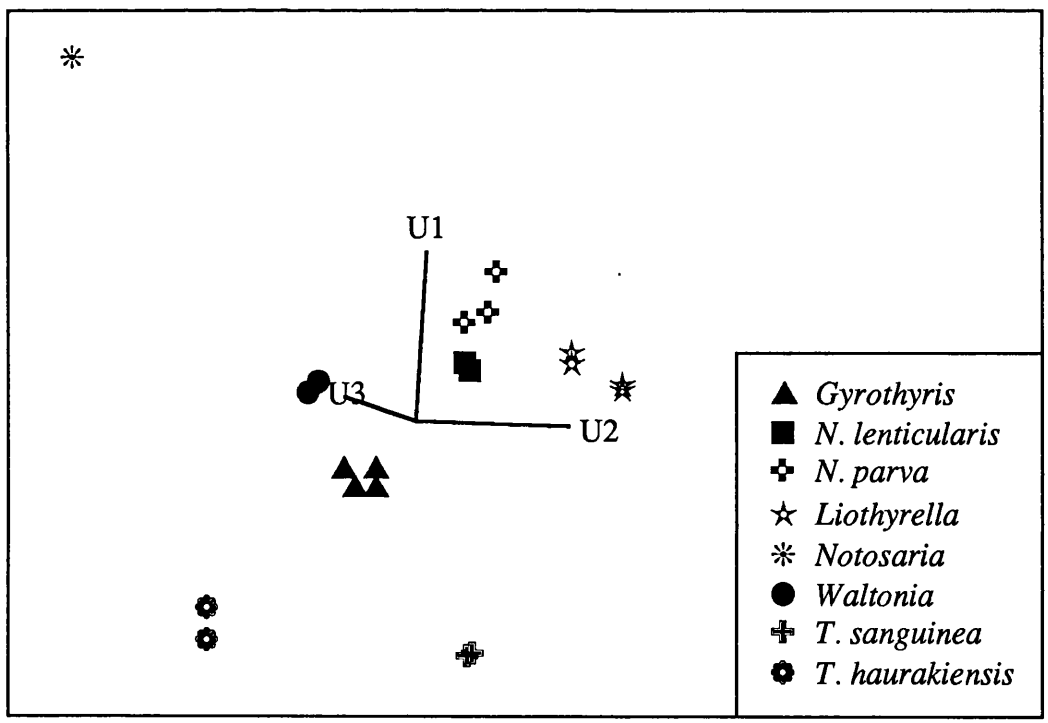


Figure 5.2: Graph of the first three principal components of the relative proportion data for amino acids from Recent brachiopods (complete organic extract). Note close groupings of related species (e.g. members of the *Neothyris* genus) and the distance of more distantly related species (e.g. *Notosaria*).

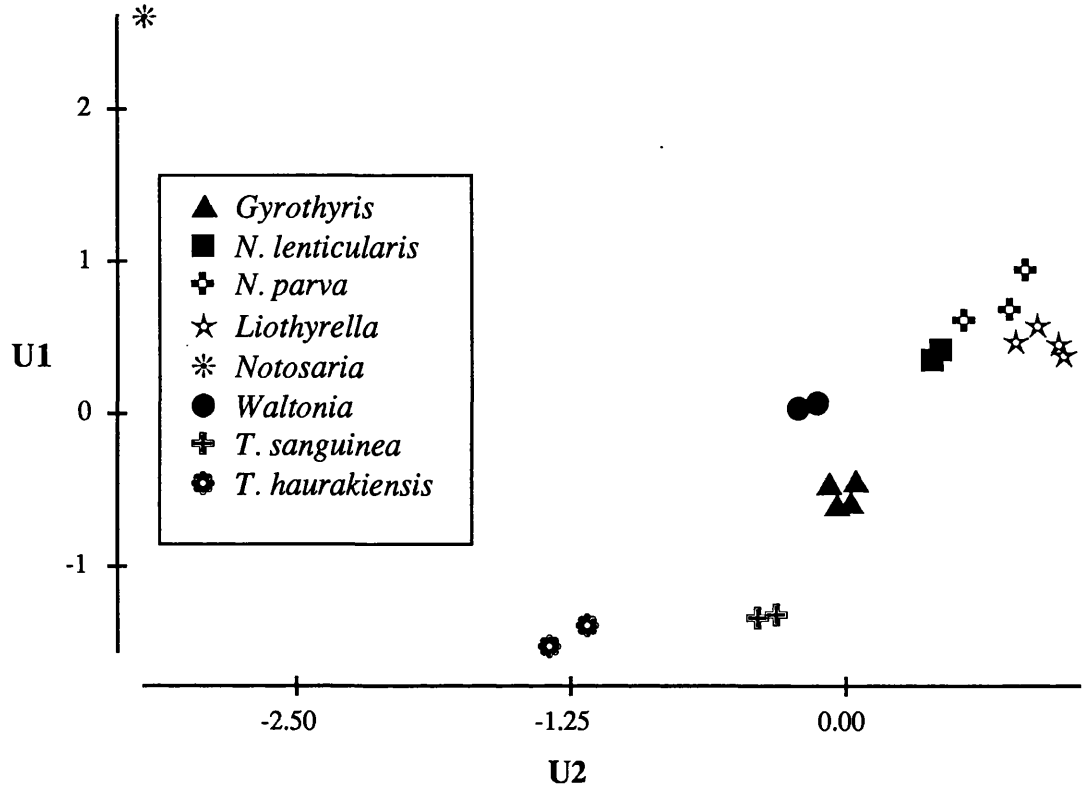


Figure 5.3: Scatterplot of the first two principal components of the relative proportions of amino acids from Recent brachiopods.

(0.350), Arg (0.542), Thr (-0.451) and Lys (-0.375). The signs of the eigenvalues are also important, as these indicate the direction of the sample along the particular eigenvector. Both the first two principal components (75.2%) and the first three principal components (85.7%) contain more than 75% of the total variability, and therefore conclusions reached from this data are valid (Sneath and Sokal, 1973).

Taxonomic information is difficult to derive from the principal component data alone, as this method provides only discrete groupings of samples which are not linked together. The method, at this scale at least, is unable to separate the two species of the genus *Neothyris*. Cluster analysis (**section 1.6.3**) is a method of assaying the degree of "relatedness" of the dataset, and the variables were analysed by single linkage analysis, which is equivalent to the nearest neighbour analysis of Sneath and Sokal (1973). The dendrogram in **figure 5.4** is derived from cluster analysis of the data shown in **table 5.1**. Taxonomic relationships can be seen by the positions of the clusters and in the links between clusters. The genus *Notosaria* plots away from the remaining data, whereas the two species of the genus *Neothyris* form a discrete cluster. *W. inconspicua* and *G. mawsoni* are closely related. The two species of the genus *Terebratella*, however, do not appear to be as closely related as would be expected. *L. neozealandica* appears to be related to the genus *Neothyris*.

EigenValues

	Values	Variance Proportion
e1	6.694	47.8
e2	3.839	27.4
e3	1.470	10.5

EigenVectors

	V1	V2	V3
D/N	0.267	-0.294	-0.304
E/Q	-0.091	0.437	0.019
S	-0.328	-0.032	0.149
G	0.201	0.332	0.350
R	-0.224	-0.081	0.542
T	-0.203	0.289	-0.451
A	-0.349	-0.067	0.176
P	-0.351	0.068	0.029
Y	0.003	-0.500	-0.094
V	-0.180	0.335	-0.263
I	-0.361	-0.058	0.024
L	-0.372	-0.021	-0.085
F	-0.218	-0.362	0.083
K	-0.296	-0.122	-0.375

Table 5.2: Principal component analysis of the relative proportion data, calculated for samples of Recent brachiopods. Only the first three eigenvectors and eigenvalues are shown.

5.5 Discussion

5.5.1 The theory of composition analysis for protein comparison

Although the amino acid composition of proteins does not provide such precise conclusions regarding taxonomy as do sequence comparisons, they are not entirely devoid of such information (Cornish-Bowden, 1983). Amino acid hydrolysates of proteins yield randomly distributed samples of amino acids from the ordered protein, and direct sequence information is lost. The sequence of amino acids in a particular protein is non-random, recording information regarding the genome which formed the RNA template of the protein, and which also defines the structure and function of the protein. It has been argued (Cornish-Bowden, 1979) that differences in the sequence of proteins which carry out similar functions in different species must occur at sites where the requirement for a particular amino acid is weak (i.e. a site where the amino acid does not contribute to the function of the protein). If this were not true, the protein would be either

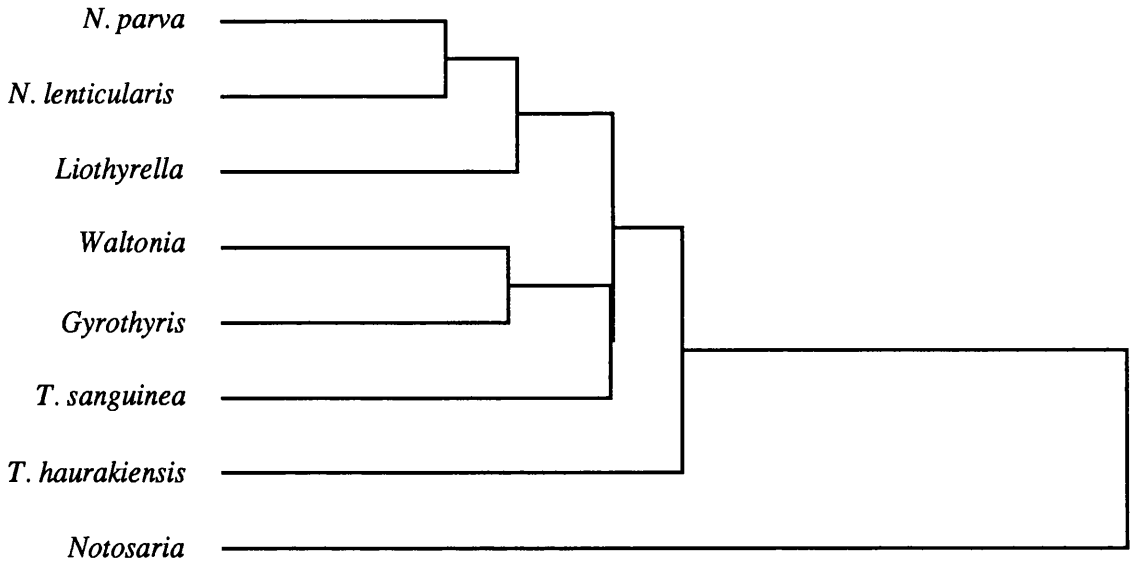


Figure 5.4: Single linkage cluster analysis showing apparent relationships of extant brachiopods from New Zealand. Note the abberrant position of *Liothyrella*, caused by chance similar ratios of amino acids. *Liothyrella* contains a different number and size of proteins (**figure 5.1**), indicating that the extracts should only be compared with caution.

unable to perform its function, or that function would be severely impared and it is likely that that particular sequence would be lost in the course of natural selection. The distribution of these differences may be considered as being randomly distributed in the protein (Cornish-Bowden, 1979; 1983).

Critics of the method of amino acid comparisons have argued that hydrolysis of a bulk sample is like mixing up all of the letters which constitute a book at random, and then trying to read and understand that book (Logan *et al.*, 1991). However, to demonstrate that this process would not lose the taxonomic information, Cornish-Bowden (1979) uses the analogy of the expansion of π, where the sequence of figures are known and are entirely predictable, but these figures would give an insignificant result to any statistical test of the hypothesis that the numbers are distributed multinomially, with a probability of 0.1 of finding any one digit at any one site.

The theory of relating proteins by their amino acid composition requires different approaches for differing degrees of relatedness between proteins (Cornish-Bowden, 1979):

(i) Proteins of a similar size and composition (i.e. proteins which carry out the same function in different organisms) may be directly compared with little or no correction factor.

(ii) Pairs of related proteins in which one is approximately twice the size of the other one is thought to be as a result of gene duplication in the longer sequence. In this case, it is possible to halve the number of residues for the longer chain,

and directly compare it to the shorter chain.

(iii) In cases where the protein size is very different, and the proteins are unrelated, care must be taken to ensure that there is little danger of deducing a relationship which does not exist.

Clearly some knowledge of the size of the proteins in samples is necessary, in order to know whether or not a correction is required regarding the amino acid data, or whether relationships produced by the amino acid data are likely to be misleading.

In the theory of statistical relatedness, Cornish-Bowden (1979) defines an index of compositional difference (denoted by $S\Delta n$), which depends on amino acid analysis of proteins of known length. The amino acid composition of snake venoms of known protein sequence was used to test the theory. Phylogenetic trees of relationships between the venoms were constructed using the UPGMA method of Sneath and Sokal (1973), the first tree being constructed using the number of unpaired residues contained within the longer sequence, and the second with values of $S\Delta n$ (**figure 5.5**). There are differences between the trees, which are probably due to the loss of the "fine tuning" provided by the sequence data, but the overall similarity between the trees is striking. The general relationships within clusters remain correct, although there are differences in some of the positions, and also in the branching positions of the dendrograms. Other pairs of trees were constructed, and although in no case was there perfect agreement between the two, the results *"are similar enough to suggest that compositional comparisons can provide a useful guide to likely evolutionary relationships"* (Cornish-Bowden, 1979, p. 385).

5.5.2 Application to numerical taxonomy

Numerical taxonomy was defined by Sneath and Sokal (1973, page 4) as *"the grouping by numerical methods of taxonomic units into taxa on the basis of their character states"*, and implied repeatability and objectivity. The work of Cornish-Bowden (1979; 1983) provides a theoretical and practical base for the characterisation of proteins by the distribution of their amino acid residues, and is an aid to the use of amino acid differences for taxonomic purposes using the principles of numerical taxonomy, i.e utilising non-subjective criteria.

The theory of the relationship of proteins based on their differences in the composition has important implications for the use of amino acid analysis of intracrystalline molecules as a taxonomic tool. The technique has been much criticised in the literature for not providing enough data on the taxonomic difference between samples, and also for being difficult to interpret (e.g. Prager and Wilson, 1971; Logan *et al.*, 1991; Collins *et al.*, 1991a). However, these arguments have rested almost entirely on the assessment of differences of amino acids on a ratio and relative proportion basis, the limitations of which were demonstrated in **Chapter Four**. In that section, the amino acid compositions of

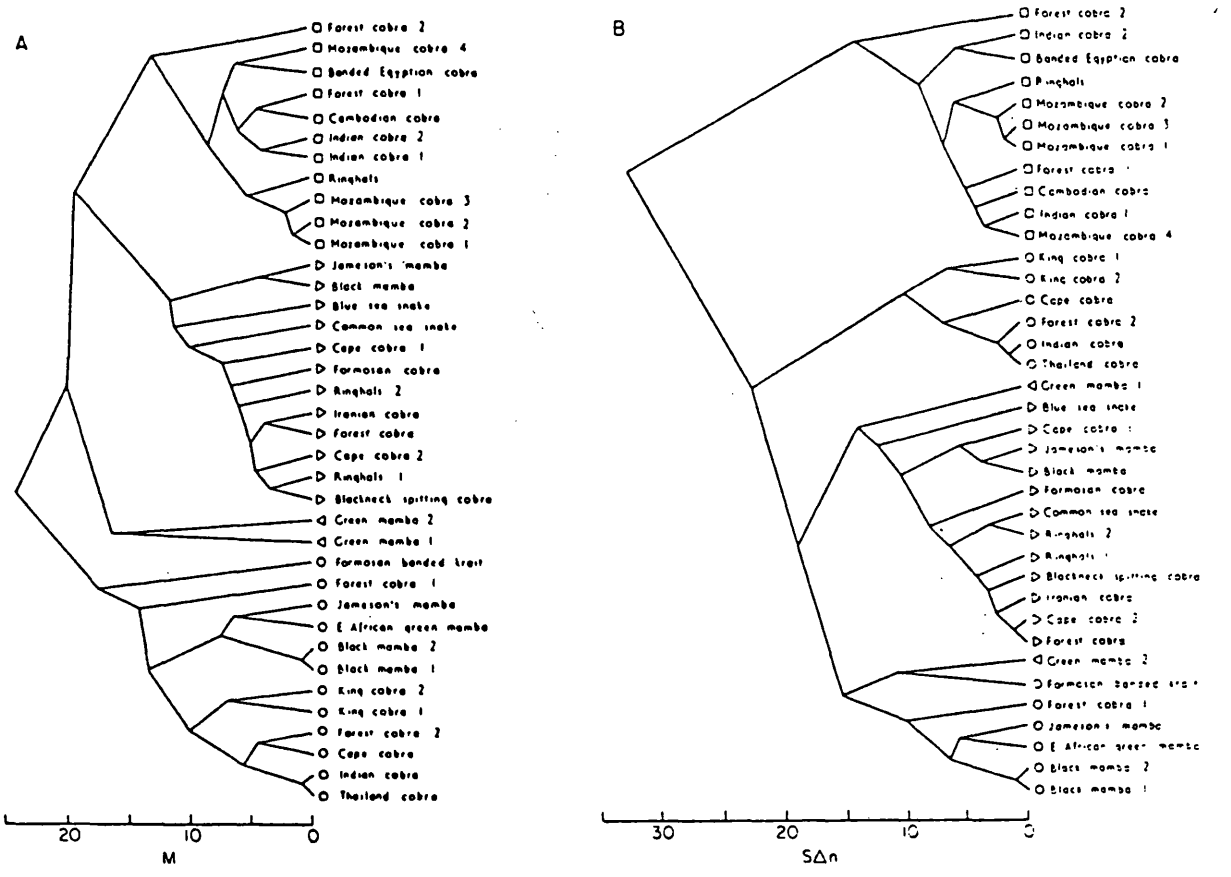


Figure 5.5: Dendrograms produced by Cornish-Bowden (1983) for snake venoms to demonstrate the comparison of proteins by amino acid compositions (see text; after Cornish-Bowden, 1983).

finger tips and sediments showed a similarity with the relative proportions of amino acids present in fossils. Such simple comparisons are misleading, and the use of more powerful multivariate statistical techniques, such as cluster analysis and PCA, can lead to more encouraging patterns of data.

Factor analysis, a multivariate statistical technique, was used by Degens *et al.* (1967) in an analysis of the bulk amino acid composition of both calcified and non-calcified tissues from the mollusca. Degens *et al.* (1967) produced a phylogenetic tree from bulk amino acid analyses of proteins, which is consistent with that from conventional techniques. However, Degens *et al.* (1967) used only numerical descriptions of the factor scores of the data, and made no attempt to describe it graphically, which makes interpretation of the data difficult. Q-mode factor analysis of shell proteins from Recent and fossil planktonic foraminifera was used in a phylogenetic study (King and Hare, 1973), where the data was expressed on a triangular plot, representing the equivalents of the first three principal components. King and Hare (1973) were also the first to conclude that chemotaxonomy of amino acids was viable in extant species.

Both of the above studies utilised analysis of the bulk extract of the shell in

order to examine phylogenies, but neither placed significant limitations on the technique. In order to recognise the full potential, the shortcomings of the method should also be considered.

5.5.3 Application to New Zealand brachiopods

The method of Cornish-Bowden (1979; 1983) may be used as a basis for understanding relationships from the study of amino acids from extant species. The first approach of the method (**section 5.5.1**) deals with differences between proteins which are of similar size. This approach may be applied when the samples of *N. lenticularis*, *N. parva*, *T. sanguinea*, *T. haurakiensis*, *W. inconspicua* and *G. mawsoni* (all members of the Superfamily Terebratellida) are compared, as several of the samples contain a similar number and range of size of proteins in SDS PAGE. Comparison of these samples with the intracrystalline proteins of *N. nigricans* and *L. neozealandica* must follow the rules of the third approach, as there are different numbers of proteins of different sizes contained within the shell. Thus, when amino acid data is considered, comparison of the Terebratellida with other samples may produce a relationship which is not a true reflection of the phylogeny of those samples.

As the sequence of the intracrystalline proteins has only partially been determined (Cusack *et al.*, 1992), it is not possible to be rigourous with the application of the method of Cornish-Bowden (1979). Values of $S\Delta n$ cannot be calculated, as this requires the knowledge of both the number of different residues, and also their locations. However, the general rules of the theory will still apply, even though the data presented here includes the variation of the amino acids in **all** of the polypeptides and proteins within the shell. Within the Terebratellida, samples have a similar number of proteins of a similar size and therefore differences between proteins will be directly reflected by differences in the amino acid composition of the proteins. The information revealed by this process will not be at the same level as that obtained from analyses of proteins separated out from a natural mixture, as changes in one polypeptide chain may be cancelled out by changes in another. From the brachiopod shells, the 6.5 kDa protein occurs with a similar sequence (up to 20 residues) in three of the genera studied here, and Cusack *et al.* (1992) have surmised that they are homologous and that there is no need to insert gaps into the sequence in order to match the sequence, and will therefore be likely to perform the same function in each of the species. This rule is likely to hold true for the other shell proteins extracted from the Superfamily Terebratellida which have not yet been sequenced. Taxonomic interpretations based on this method are therefore likely to be accurate, and the variation in amino acids between the mixture of proteins in each of the samples will indicate taxonomic variability. When comparing samples of *N. nigricans* (different Order) and *L. neozealandica* (different Superfamily) to the Terebratellida, spurious results may be obtained, as the number and size of the proteins are different (**sections 5.4.1 and 5.5.1**) and

care is required in the interpretation of the amino acid data.

The following taxonomic conclusions may be reached for the samples in this study. Samples of the genus *Neothyris* are considered to be members of separate species, in terms of their position in the cluster diagrams (**figure 5.4**). The plot of the first three principal components (**figure 5.2**) is unable to separate the samples, indicating that they are closely related. *G. mawsoni* and *W. inconspicua* are also closely related, although this relationship is obviously not as close as for the *Neothyris* samples. This conclusion is in agreement with the data from the *Treatise* (Williams *et al.*, 1965), which shows that these two species are members of the same subfamily.

The genus *Terebratella* is represented in this study by *T. haurakiensis* and *T. sanguinea.* If these species were members of the same genus, then it would be expected that they would plot together on the PCA plot, and form a discrete cluster, similar to the two species of *Neothyris.* The morphological difference between the two species is in terms of the ribbing pattern of the shell; *T. sanguinea* is heavily ribbed, whereas *T. haurakiensis* is smooth. This could be due to ecophenotypic variation, as seen by the ribbing of the related normally smooth shelled genus *W. inconspicua,* which is "not uncommon" (D.E. Lee, pers. comm.), or it could be due to a more fundamental difference. It seems likely, from the bulk amino acid data, that the classification of this genus may need revising, and further work on the intracrystalline amino acids and proteins from an area where the two species coexist is necessary in order to confirm these findings. Unfortunately, only a limited amount of material was available for the study of *T. haurakiensis*, rendering it impossible to undertake electrophoretic studies, to assess whether changes in the amino acid composition were due to changes in the amino acid composition of similar sized proteins or whether the samples contain different numbers of proteins (c.f. the amino acid composition of *L. neozealandica*).

The genus *L. neozealandica* is problematical. Although belonging to the same order as the Terebratellida, it is a member of a different superfamily. However, in PCA (**figure 5.2**) *L. neozealandica* plots closely to the other data, and in the cluster analysis (**figures 5.4**) it clusters with the genus *Neothyris*. SDS PAGE shows that *L. neozealandica* contains a different number of protein bands compared to the Terebratellida, and at different sizes, indicating that the close relationship of the samples in the analysis is an artifact of the data, and that similar relative concentrations of amino acids are present in different proteins. This is an example where the third approach of Cornish-Bowden (1979) is required, the amino acid composition of proteins of different size has produced apparent relationships which do not exist. This is the major limitation of this technique of using comparisons of bulk amino acid compositions to relate proteins, although, as this study has shown, the use of an additional technique such as SDS PAGE can assist in the recognition of false taxonomic interpretations.

The data from the amino acid analysis of the intracrystalline molecules from

N. nigricans will also require the third approach. Morphologically, this brachiopod differs from the remaining samples at the ordinal level, and also contains shell proteins of different molecular weights to the Terebratellida. The difference in the composition of the shell proteins is shown by the distance between *N. nigricans* and the other species, both in cluster analysis and PCA (**figures 5.2 and 5.4**), which accurately records the differences between the amino acids present within the species. This shows similarity with the immunological investigation of Endo (1992), where antibodies produced against *N. nigricans* showed no reaction with any other brachiopod shell protein, indicating large differences in the composition and sequence of the proteins.

This study has shown that the separation of Recent samples is possible to the specific level in brachiopods using only the proteins and amino acids extracted from within the shell, i.e. using a technique which uses only objective methods. The technique is rapid and the relationships between the samples are easily seen by the graphical presentation of the multivariate statistics.

5.5.4 Anomalies between partial sequence and amino acid analyses

The high proportions of Gly found within the bulk amino acid composition of the samples contrasts with the partial sequence data so far derived for brachiopod intracrystalline proteins (Curry *et al.*, 1991a; Cusack *et al.*, 1992; M. Cusack pers. comm.), which includes only moderate amounts of Gly. This could be due to the different sample preparation methods; the proteins for sequencing undergo concentration by filtration which removes most compounds with a molecular weight of less than 10000 (10 kDa, although this depends on the conformation of the protein). Polypeptides with a molecular weight of less than this may contain larger abundances of Gly. In a study of the relative abundance of amino acids in 207 proteins of known sequence (modified from Klapper, 1977), Gly had an average abundance of 8.51%, indicating again that the Gly appears, in these proteins, to account for a higher proportion of amino acids than average.

5.6 Implications for the study of fossil molecules

The information contained within intracrystalline molecules was sealed by the mineralisation of the shell carbonate; hence both surviving molecules and their degradation products will be trapped in a contained microenvironment. This contrasts with the previously published data for the amino acid composition of the intercrystalline shell matrix (Jope, 1967 a; 1967b; Kolesnikov and Prosorovskaya, 1986). These molecules are in sites where they may be easily degraded and leached by percolating ground fluids or consumed by micro-organisms, which would alter the molecular record. In a study of the Recent articulate brachiopod, *Terebratulina retusa*, Collins (1986) found that the intercrystalline protein between the secondary shell fibres of brachiopods are degraded in less than one year, and that after this time the fibres are able to behave as individual units,

rather than being part of a layer, and may separate from the remainder of the shell. This corresponds to the decay of intercrystalline molecules and possible leaching of the decay products (Abelson, 1955). This indicates that intracrystalline molecules from fossils will be a much more reliable source of historical information. In the study of fossil molecules, however, study of Recent samples is required to indicate a starting point for the samples (**Chapter Seven**). The study of the Recent samples has also shown that it is possible to assess taxonomic relationships by the use of multivariate statistical analysis of amino acid compositions, regardless of the numbers of different proteins which were originally present.

5.7 Conclusions

Taxonomic analysis of the brachiopods using the wt% of the amino acids reinforces the morphological taxonomy. Using the classification of the brachiopods from the Treatise on Invertebrate Palaeontology (Williams *et al.*, 1965), the data is divided into two orders, the Order Rhynchonellida, represented by *Notosaria*, and the Order Terebratulida, which encompasses all of the remaining brachiopods in this study. Within the Order Terebratulida, *Liothyrella* is included within the Suborder Terebratulidina, whereas the remainder are included in Suborder Terebratellida. At the subfamily level, *Neothyris* is included in the Subfamily Neothyridinae, and the remainder in the Subfamily Terebratellinae. From this data, an approximate dendrogram may be produced (**figure 5.6**), to show suggested relationships. It would be expected that *Liothyrella* and both species of *Neothyris* would plot away from the Terebratellinae.

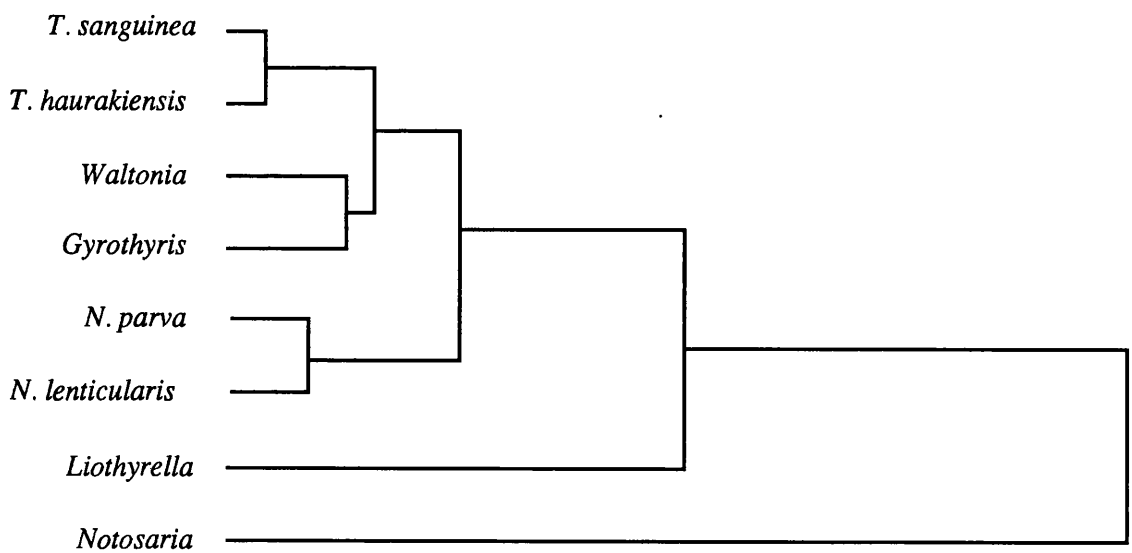


Figure 5.6: Dendrogram showing morphological relationships from the Treatise (Williams *et al.*, 1965).

In **figure 5.4**, however, *Liothyrella* plots closest to *Neothyris*, and the species of the genus *Terebratella* do not form a discrete cluster. As would be expected from the data in the Treatise, *Gyrothyris* and *Waltonia* plot together as members of the same subfamily. It has been shown by gel electrophoresis, that the relationship between *Neothyris* and *Liothyrella* is false, and due to the occurrence of similar ratios of amino acid rather than any true genetic similarity.

The technique also has implications for the study of fossil molecules. The degradation products of the intracrystalline molecules will remain trapped within the shell until the shell material is dissolved, recording taxonomic differences, by molecules, in the fossil record. Changes between Recent and fossil molecules in the study of the intercrystalline matrix may be diagenetic rather than genetic. The effect of time on molecules, and the state of their preservation, as well as taxonomic details are presented in **Chapters Seven and Eight**.

Characterisation of fossil proteins

6.1 Introduction

Intracrystalline proteins from Recent samples exist as a complex mixture of compounds (Curry *et al.*, 1991a; **Chapter Five**). The examination of proteins requires separation of the molecules from the mixture, and purification of the fraction to homogeneity (Curry *et al.*, 1991a; Cusack *et al.*, 1992), to ensure that the information recovered from the shell applies directly to a molecule of known size and charge. As the molecules from Recent samples exist in such a mixture, it is reasonable to assume that the fossil samples of the same species originally had intracrystalline proteins of similar sizes. The examination of molecules from Recent samples can therefore act as a guide to methods which may be used to isolate proteins/peptides from fossil shells. Large quantities of the extant representatives are relatively easily collected, and this allows experimentation with differing analytical techniques, in order to determine the best methods for examination of the molecules, before utilising these methods to attempt to study the lower quantities of available fossil material.

As the molecules from fossil shells are likely to have originally been similar to those of Recent shells, techniques that are successful in the separation of intracrystalline molecules from Recent brachiopods (Curry *et al.*, 1991a; Cusack *et al.*, 1992), were applied to organic extracts from fossils in an attempt to separate and purify fossil peptides. If the molecules have undergone little or no degradation in the fossil record, then the resulting peptides may be easily separated and purified, in the same way as intact proteins/peptides from Recent samples, or samples which have undergone enzymatic or chemical cleavage (**figure 6.1**).

The aims of this part of the study were therefore:

(i) To attempt to extract and purify intracrystalline proteins and peptides from fossils.

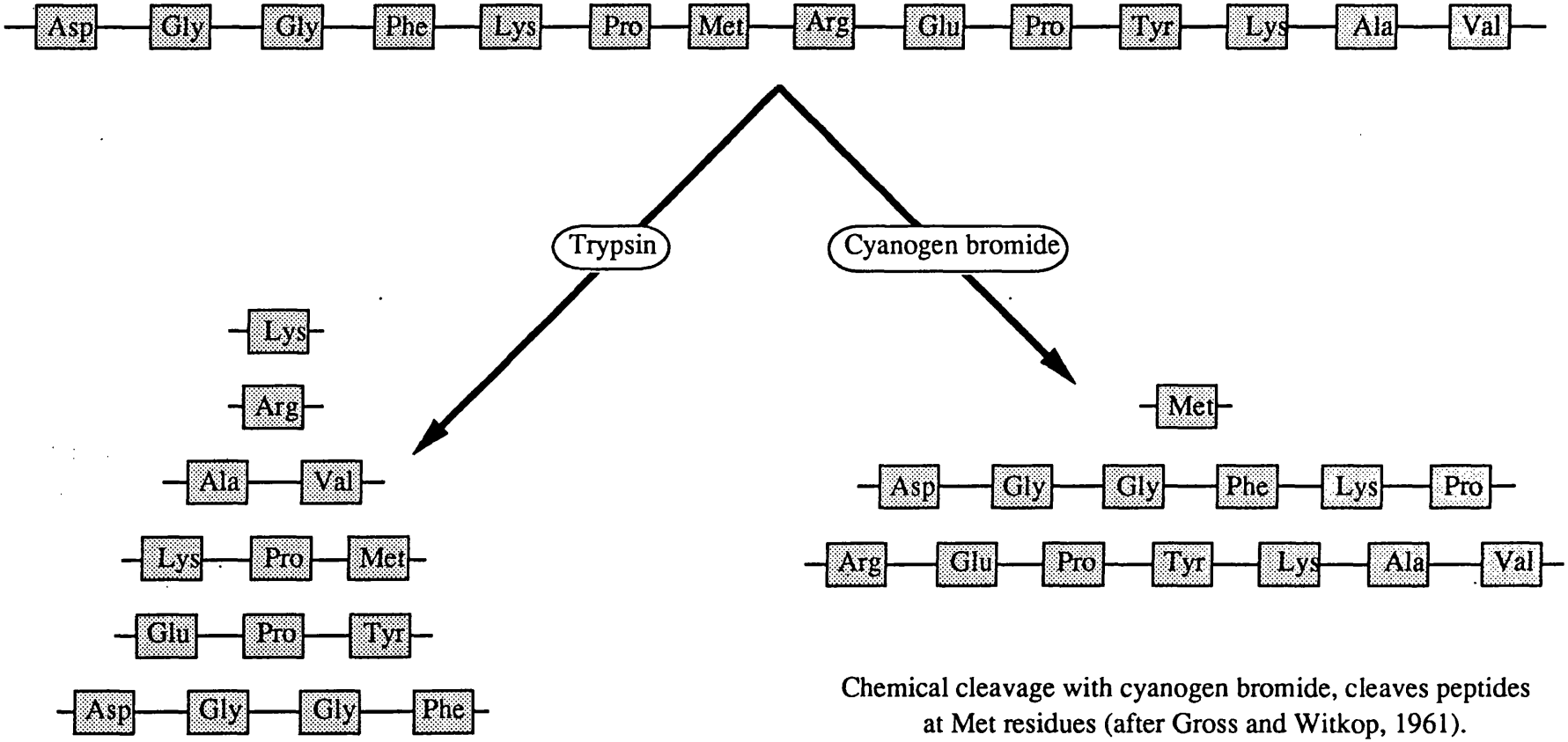
(ii) To characterise these molecules once purified.

(iii) To assess the molecular state of preservation in the fossil shells.

6.2 Previous Work

Prior to further consideration of fossil molecules, it is necessary to assess the current state of knowledge regarding protein degradation and the preservation of macromolecules over time, and to use this knowledge as a basis for further study. The best way to complete this is in a review of the literature specifically dealing with the separation and preservation of proteins and peptides in the fossil record.

The examination of protein degradation has been effectively confined to two



Enzymatic cleavage at Arg and Lys residues, except when the second amino acid is Pro where cleavage is much slower (after Geiger and Fritz, 1983).

Chemical cleavage with cyanogen bromide, cleaves peptides at Met residues (after Gross and Witkop, 1961).

Figure 6.1: Chemical and enzymatic cleavage of a schematic peptide. Trypsin and cyanogen bromide are two methods of cleavage, for other examples see Harris and Angal (1989). Natural hydrolysis of peptide bonds (degradation due to time, temperature and/or pressure) would have a similar effect.

time periods, the examination of old proteins from non-fossil sources and those from relatively old fossil material. The distinction is between old proteins, in the order of 1-1000's of years old, compared to proteins and peptides which have been preserved over many millions of years. The intervening period (100's of ka to less than 5 Ma) has rarely been studied.

Protein decomposition was first noted over long periods of geological time in a series of studies of shell proteins (both inter- and intracrystalline) from various species of fossil by Abelson. In samples of Recent and subfossil (1 ka) *Mya*, Abelson (1955) found that there was no change in the protein content of the two samples. In the mollusc *Mercenaria*, Abelson (1955; 1956) noted that the protein content of the Pleistocene shells was 15.35% of that present in the Recent, and by the Miocene, this had dropped to 2.15%. It was also calculated that between 1 and 5% of peptide bonds from these proteins were broken in the period 10 ka to 100 ka, supported by amino acid analyses. The loss of peptides from the shells was equated with the breaking of the peptide bond, resulting in the successive release of smaller soluble peptides from the shell by the leaching action of percolating groundwaters.

The degradation of shell proteins was examined by Ho (1966) in a stratigraphic study to equate the concentration of protein (expressed by protein-nitrogen values) with time. Variation in the degradation of proteins due to climatic effects was noted, indicating that both water and temperature, as well as time, would have an effect on protein breakdown. The water content of fossils will therefore tend to influence the rate of natural hydrolysis of the protein (i.e. decomposition of the constituent peptide bonds). The first direct measurement of protein degradation through primary sequence analysis (the order of amino acids along the protein chain) was completed by van Kleef *et al.* (1975) who showed that proteins could decay by non-enzymatic processes in a relatively short length of time in the bovine eye lens. The lens is built up like the layers of an onion, with successive layers (representing later growth stages) covering earlier ones, and stopping biological activity in the juvenile lens. The juvenile lens in older animals contains short chain versions of proteins present in the outer layers of the eye. This study indicated that the cleavage of peptide bonds was likely to occur preferentially at sites between particular amino acid residues (in this case, the following bonds are likely to be easily degraded; Ser-Ala, Ser-Ser, Asp-Ala, Asn-Glu and Thr-Ala), and that the first stage in the breakdown of proteins was the natural hydrolysis of peptide bonds by the addition of a molecule of water. This study did not take into account the conformation of the protein; tightly packed globular molecules are likely to be better protected from degradative reactions (Eglinton and Logan, 1991), and hence would take longer to become degraded.

The identification that the bonds between different amino acid pairs in a protein would decay at different rates has been followed by several other studies of peptide bond decomposition within proteins and peptides, e.g. the half life for hydrolysis of the

Phe-Gly bond, when Gly is at the *c*-terminus, is 7 years (Kahne and Still, 1988). However, these measurements of degradation rates can only be applied to proteins of known sequence; if the pairs of residues on either side of a peptide bond are unknown, then it is not possible to estimate the time required for the degradation of a particular bond.

Macromolecules released from materials collected at archaeological sites represent some of the best opportunities to assess the potential for intact protein preservation. The extracts from old barley and other cereals have been examined, showing that some seed storage proteins are not preserved in samples older than 200 years. After this time, it is not possible to separate or visualise them using starch gel electrophoresis (Zeven *et al.*, 1975), or by SDS PAGE of samples up to 8 ka (Shewry *et al.*, 1982). Instead of sharp, narrow, protein bands (as in **figure 5.1**), the stained gels contained lower concentrations of a range of much smaller peptides, indicating the decomposition of the parent proteins. However, amino acid analyses of both old and Recent grains (of a related family) showed similarity, although losses of some amino acids are shown. Preservation of similar amino acid compositions indicate *in situ* decomposition of proteins and restricted leaching of reaction products. Antibodies produced against specific polypeptides in Recent seed storage proteins showed no reaction with samples dated at 3 ka (Shewry *et al.*, 1982), indicating the loss of epitopes and hence degradation or denaturation of the molecules.

Immunology is the most common method utilised in the study of fossil molecules. Variations on the general methods of immunology have been employed, and antigenic properties have been recorded from samples up to 70 Ma (de Jong *et al.*, 1974), although this has not been reproduced. Analyses of the reactivity of one specific protein over time, via immunology, has only been completed in a few studies. Lowenstein (1980) studied the immunological response of reactive collagen from samples up to 1.9 Ma, and found that, although the level of reactivity was much reduced, this protein was still recognizable (**figure 6.2**). This reduction in antigenicity indicates an increase in the

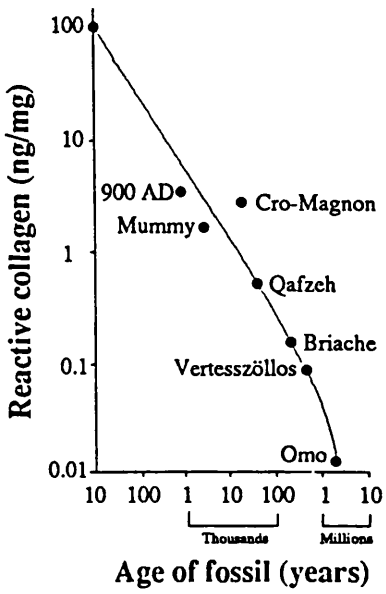


Figure 6.2: Immunological reactivity of the collagen molecule over time, showing a near linear decrease in the concentration of the reactive collagen over time. After Lowenstein (1980).

degradation of the protein to smaller peptides, via the loss of determinants (highly specific areas of proteins, consisting of only a few amino acids, which are used as a target for antibodies). A similar loss of reactivity between antigens from living, Recent (stored in formalin for 6 years), and subfossil (of unknown age) brachiopods, and antibodies produced against proteins from living brachiopods, was noted by Collins *et al.* (1988), although no figures were given. The immune response to fossil brachiopods (0.5 Ma) represents only 0.1-1% of the reactivity of Recent samples (Collins *et al.*, 1991b), indicating large scale decomposition of proteins over this short time period.

The racemisation of amino acid molecules is a reaction which accompanies protein and peptide degradation. After the initial hydrolysis of the peptide bonds, fragments of the protein form diketopiperazides as intermediates, prior to dipeptides and free amino acids, and it is likely that racemisation occurs in this intermediate stage (Steinberg and Bada, 1983). Protein degradation therefore begins prior to racemisation. This is confirmed by the comparison of the strength of the immune reaction with the degree of racemisation, which shows an almost linear relationship (Muyzer *et al.*, 1988; **figure 6.3**), hence samples which have undergone racemisation have also undergone some degree of natural hydrolysis and cleavage of determinants and therefore the proteins are also degraded.

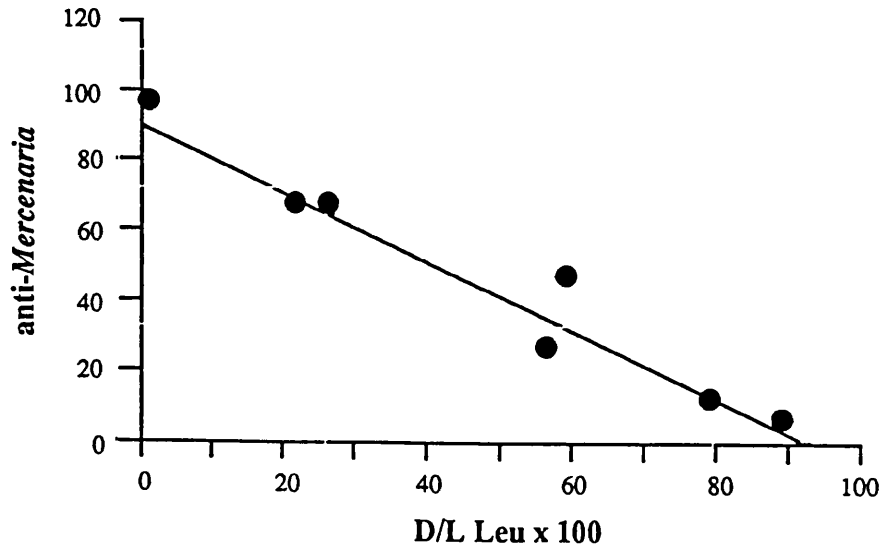


Figure 6.3: Immunological response (absorbance of UV light) of the anti-*Mercenaria* antibody reacting with antigens from *Mercenaria* samples with differing D/L ratios. The increase in the D/L ratio shows a near linear relationship with the decrease in immunological response.

Weiner and Lowenstam (1980) confirmed the observation of Towe (1980) that excellent physical preservation was no indication of good biochemical preservation by the analysis of samples which had only undergone mild inorganic diagenesis. The proteins in these samples were almost completely degraded. Samples need not therefore be subjected to either heating or burial (i.e. increase in pressure) for the molecules to degrade, although these effects will undoubtedly increase the rate of reaction.

Robbins and Brew (1991) used 2-dimensional SDS PAGE to separate out proteins from the mixture contained in the tests of 300 ka foraminifera of mixed species, picked from deep sea cores. These results showed that, although there had been some loss of protein from the sample since death, between 25 and 30% of the equivalent protein from the Recent can survive through to this age, in the relatively unprotected sites of foraminifera tests.

Collins *et al.* (1991a) used liquid chromatography and SDS PAGE on extracts from fossil samples from New Zealand. Although discrete fractions were eluted from the column during the analysis of Recent samples, the fossil counterparts could not be separated in the same way. SDS PAGE of the fossil samples revealed compounds with a range of molecular weights, with no fraction forming a discrete band, whereas in the Recent samples, discrete bands were formed.

In summary, proteins tend to lose their reactivity relatively rapidly, probably due to denaturation of the proteins, and separation of chains (loss of quaternary structure). The first reaction in the degradation of primary structure of the protein is natural hydrolysis, the rate of which is dependant upon temperature and water, and also upon the nature of the residues on either side of a particular peptide bond. Further reaction produces smaller peptides and racemisation reactions of the amino acids whilst in intermediate compounds. This is accompanied by the loss of immunological reactivity over the same time period (reorganisation of the molecular structure must therefore result in the destruction of determinants). Depth of burial and increases in temperature will increase reaction rates, although they appear not to be necessary in order to initiate the degradative reactions. The degradation of different proteins proceeds at different rates, so generalisations regarding time for this process cannot be made. Independent immunological experiments on samples from the same horizons as those analysed for amino acids in this study, show a loss of response of 99.0 - 99.9% over only 0.5 Ma (Collins *et al.*, 1991b).

6.3 Methods and materials

Fossil molecules were to be analysed both by conventional biochemical separation techniques and by amino acid analysis. The samples were prepared as in **section 3.2**, and separation techniques were applied as in **section 3.3**. In summary, samples were cleaned, powdered and bleached to remove encrusting epifauna and

intercrystalline molecules, washed, frozen and lyophilised. The shell powder was decalcified by incubation with Na_2-EDTA, and centrifuged to remove insolubles. The EDTA/calcium complexes were removed using the Minitan™ tangential flow system and the extract concentrated using a Minicon™ static concentrator. Samples were mixed with sample buffer and incubated at 100°C for 4 minutes to denature the protein, and the extract examined by SDS PAGE. A current of 100 V constant was applied to the gel to separate the proteins until the tracking dye reached the base of the housing.

6.4 Results

6.4.1 SDS PAGE

Gels for SDS PAGE were made as described previously (**section 3.3.2**). It was anticipated that any remaining soluble proteins would only be present in low concentrations, so large amounts of shells were prepared (**section 3.2**), in order to produce a concentrated solution of the molecules for analysis by SDS PAGE. During the concentration procedure, free amino acids and small peptides were lost through the membranes employed, which had a cutoff of 10 kDa. This cutoff value is also dependent upon the conformation of the protein, as a small, 6.5 kDa, protein from Recent brachiopod shells does not pass through the membrane (Cusack *et al.*, 1992).

After electrophoresis, gels were initially stained with CBB-R250 to visualise separated proteins. In all cases, this detection technique was insufficient to visualise the concentration of protein present. Extracts of 15 µL from a 100 µL sample representing over 100 g shell material from only 0.5 Ma produced no bands. Staining with silver stain produced results shown in **figure 6.4**. In no case did a single discrete band appear in the gel. Instead of this, a broad band of lower molecular weight material was visualised. This contrasts to the gels obtained for Recent material, which show several protein bands, as revealed with CBB-R250, of various molecular weights (**figure 5.1**). A similar result is obtained for fossils of all species studied here. Proteins from old and Recent barley (Zeven *et al.*, 1975; Shewry *et al.*, 1982), and from fossil brachiopods (Collins *et al.*, 1991a), show similar patterns. This is in contrast to the results of Robbins and Brew (1991).

6.4.2 Liquid chromatography

Results from SDS PAGE separation of the soluble fraction of fossil molecules indicated that molecules with a wide range of small molecular weights are present within the shells of the organisms. Gel filtration (size exclusion) chromatography showed similar results, and reverse phase chromatography was used to attempt to separate the peptide fragments in terms of their hydrophobicity.

The chromatogram in **figure 6.5** show the results of the size exclusion chromatography. The broad peak at 10-30 minutes indicates again the presence of a wide

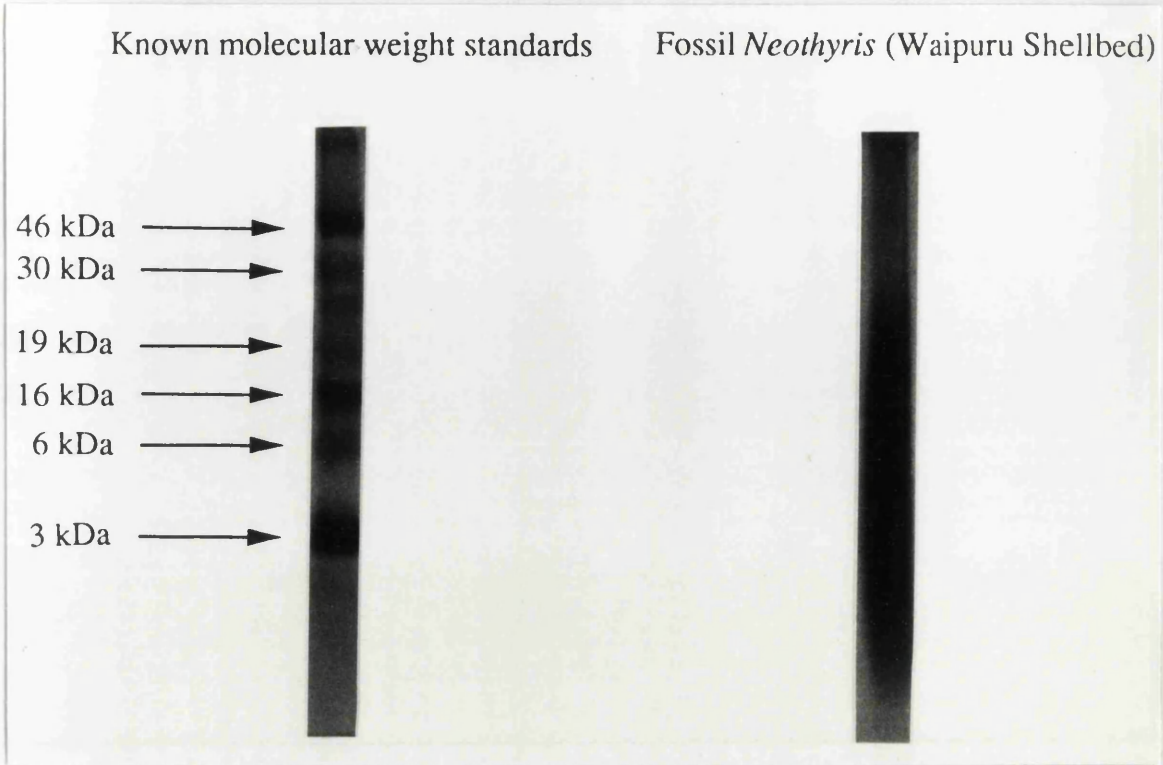


Figure 6.4: SDS PAGE of fossil samples stained with silver nitrate. The fossil sample shows none of the sharp bands found in the standards or in SDS PAGE of the organic extract from Recent brachiopods (**figure 5.1**), indicating the range of molecular weights present in the extract.

range of molecular weight compounds, rather than a sharp peak which would indicate a discrete molecular weight compound. The highest point of this peak corresponds to EDTA, tested by analysing EDTA solution on the same analysis system. Reverse phase chromatography is a much harsher separation technique, and acts to separate the molecules in terms of their chemical reactivity, which is a second possible way to purify single compounds. The chromatogram in **figure 6.6** is again composed of a wide peak, rather than sharp peaks indicative of concentrations of compounds with a particular chemical reactivity (Curry *et al.*, 1991a).

6.5 Discussion

Following the principle of uniformitarianism, it is reasonable to presume that species of fossil organisms will initially have had a similar molecular composition as extant representatives of the same species. Work completed during this study (Curry *et al.*, 1991a; 1991b; Cusack *et al.*, 1992) have shown that intracrystalline proteins from extant brachiopod species occur in a mixture, at relatively low concentrations and with a variety of molecular weights. These molecules may be extracted and purified to

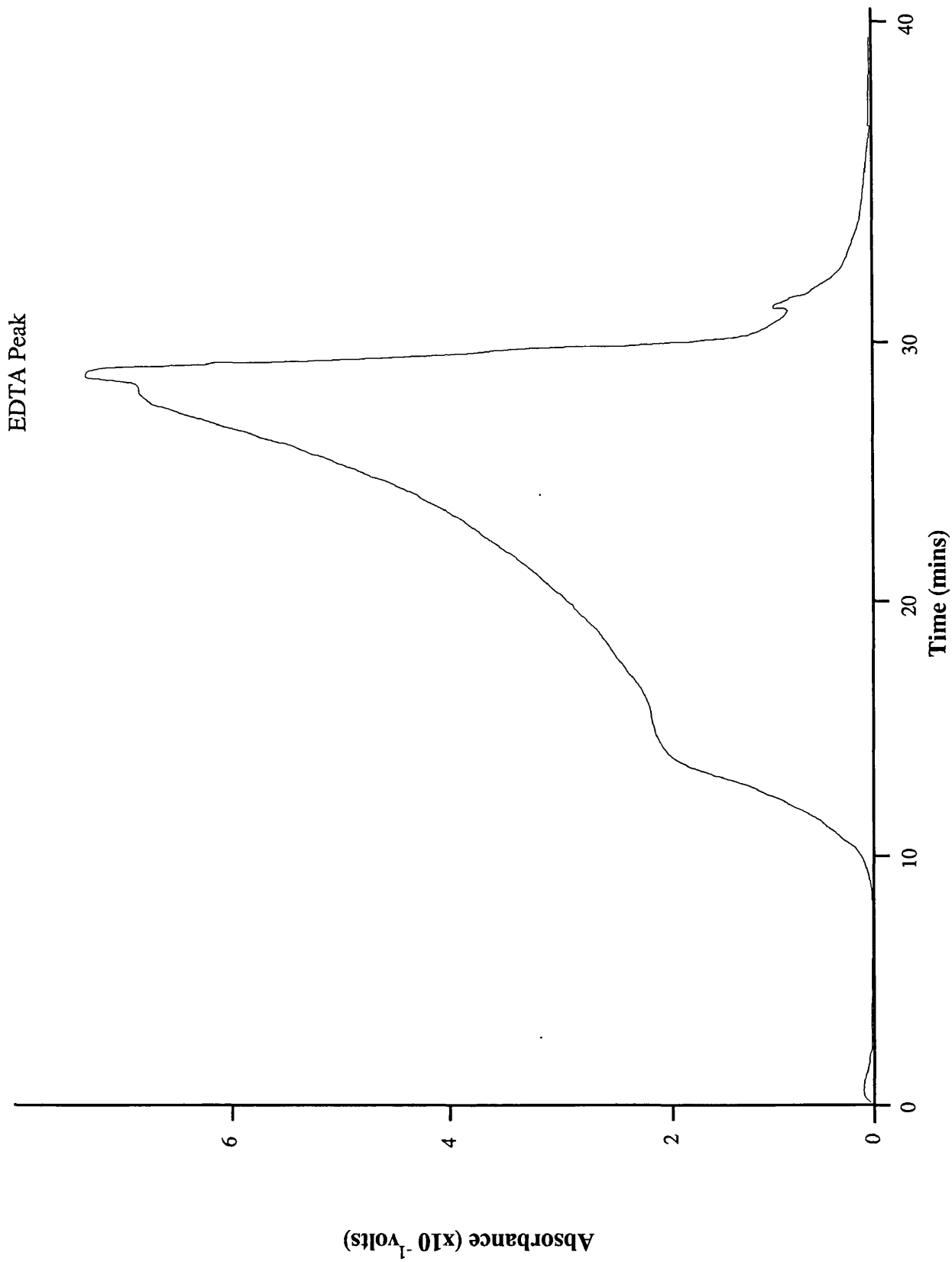


Figure 6.5: Gel filtration (fplc; monitored at 214 nm) of the > 10kDa extract from fossil *Notosaria* (Rapanui Fm). The large peak represents EDTA contamination of the sample. The chromatogram shows no clear peaks which represent discrete size fractions of peptide, indicating the range of molecular weights present in the extract.

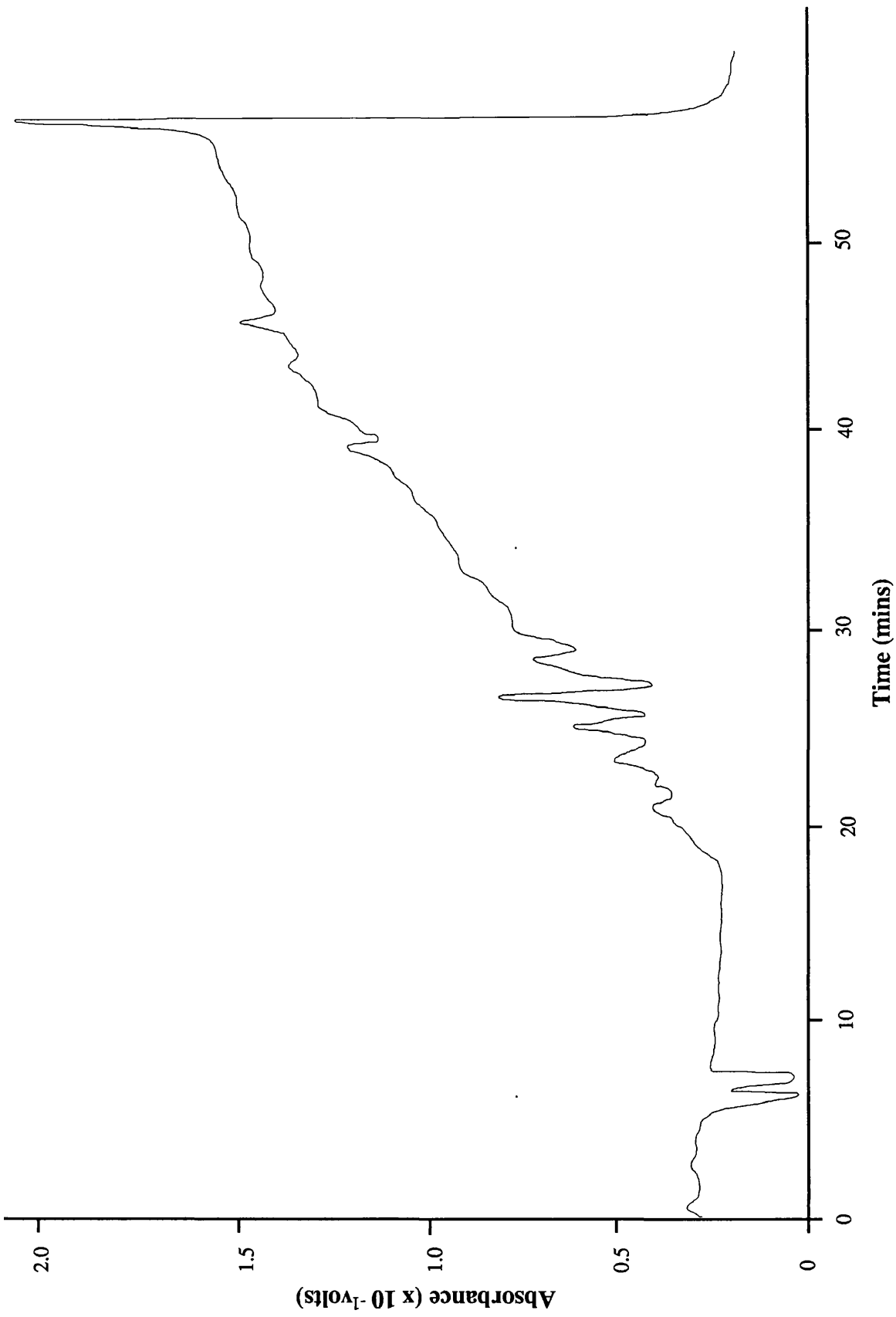


Figure 6.6: Chromatogam produced by reverse phase hplc of the organic extract (> 10kDa) of fossil *Neothyris* (Waipuru Shellbed). The lack of discrete peaks indicates the range of molecular weights present in the extract and hence the degradation of the constituent proteins.

homogeneity using the techniques described above.

Fossil molecules have been successfully separated by SDS PAGE from the entire (i.e. inter- and intracrystalline molecules) tests of foraminifera from up to 0.3 Ma (Robbins and Brew, 1991). This contrasts to the work in this study which utilises only intracrystalline molecules, which represent only c. 35% of the total amino acid present in the Recent shell (**Chapter Five**). This lower concentration may partly explain the lack of protein bands from fossils in this study, as a lower initial starting concentration would rapidly reach levels at which samples cannot be detected by SDS PAGE and liquid chromatography from these samples.

The identification of the existence of a range of molecular weights and hydrophobicity for the molecules present in the fossils indicates that the proteins identified in Recent samples have undergone severe degradation. This conclusion has been independently reached by Collins *et al.* (1991a), who attempted to apply similar methods of separation to similar fossil extracts. If degradation been limited, then it would be expected that the molecules would degrade in a similar fashion, analogous to enzymatic or chemical cleavage of proteins (**figure 6.1**). As described above, peptide bonds between different residues have differing stabilities; a low level of degradation would be expected to only affect the most susceptible of these bonds, and hence break the molecules into large peptides. The effect of this limited breakdown would be the same in all the samples in a particular horizon, where all samples are of a similar age, and which have similar geological histories. These fragments could then be concentrated and purified in the same way as non-degraded proteins. However, from the evidence described above, the protein molecules studied here are too highly degraded, and the proteins broken in many places. This extensive degradation produces a range of molecular weights and chemical activities. As the reactivity of a peptide is determined by the amino acids which it contains, the differences in these residues will also alter the reactivity of the resulting peptides.

Collins *et al.* (1991a) attributed the failure to separate out fossil peptides to condensation reactions between amino acids and sugars within the shell rather than molecular degradation, and also concluded that the protein bands from Recent brachiopods were difficult to stain due to glycosylation of the protein. The problem in staining proteins from Recent extracts has not been reported by others (**Chapter Five**; Curry *et al.*, 1991a; 1991b; Cusack *et al.*, 1992) who used the extraction technique described above (**section 3.2.3**), which differs to that of Collins *et al.* (1991a). As no evidence for glycosylation of the intracrystalline proteins has been found utilising the methodology described here, the assumption that separation is inhibited specifically by reaction with carbohydrates cannot be made in this study. It is possible that impurities remaining from the extraction of the molecules may affect the characteristics of the molecules on separation.

6.6 Conclusions

Separation of discrete molecular weight compounds by SDS PAGE and liquid chromatography was not possible for fossil molecules sampled in this study. There are several potential reasons for this:-

(i) The use of intracrystalline molecules only results in low concentrations of fossil molecules, insufficient to allow separation by the techniques employed.

(ii) Molecules have undergone reactions within the shell over geological time, which affects their molecular weight and reactivity in a way which prevents their identification.

(iii) The molecules have undergone degradative reactions to form a range of lower molecular weight compounds, with different chemical reactivities, and which can only be considered as a broad band of compounds.

It is possible that a combination of the three processes is taking place in the shell, although from the evidence presented thus far, it is likely that the molecules are being highly degraded, in as few as 0.2 to0.5 Ma, producing lower molecular weight compounds. As degradative reactions proceed, the peptides become smaller, and may be lost from the solution during purification and concentration through filters with large pore sizes. No direct evidence for the inhibition of separation by other compounds has been seen, although other compounds including lipids and carbohydrates are present within the shell (Curry *et al.*, 1991b; Collins *et al.*, 1991a). Separation of fossil intracrystalline proteins and peptides is not likely to be possible, unless very large amounts of shell material (probably in excess of several hundred grams) are used, although there is still no guarantee of successful separation. Such large amounts are not usually available, and thus an alternative method for the characterisation of the molecules is necessary. As the methods for concentration of molecules also lose the free amino acids and small peptides, there is a great potential for the loss of information in sample preparation, thus not allowing a true picture of the molecular composition of the sample to develop. Such molecules should be quantified in any method of analysing fossil molecules.

Chapter Seven: Amino acid analysis of fossil molecules

Amino acid analysis of fossil molecules

7.1 Introduction

In the previous chapter, attempts were made to separate and analyse the intracrystalline molecules from fossil brachiopods following the methods used for Recent samples (Cusack *et al.*, 1992). These attempts proved unsuccessful, indicating that the state of preservation of the molecules was poorer than first thought. Amino acid analysis of the fossil extracts analysed by SDS PAGE and liquid chromatography showed that only low concentrations of amino acid, and therefore only low concentrations of fossil peptides were present (by comparison with the concentration of Recent samples). The filtration necessary to remove the EDTA/calcium complexes from the solution had also removed small peptides and free amino acids from the solution. To resolve this, a new method comprising acid decalcification followed by amino acid analysis (**Chapters Three and Five**) was used to analyse the fossil samples. This method allows both remaining original peptides, free amino acids and degradation products to be sampled in concentrations sufficient for analysis, from amounts of shell material usually very much less than one gram (c.f. the greater than 100 g required for samples analysed by SDS PAGE). As there is no filtration/concentration stage whereby small peptides and free amino acids could be lost, this method also has the advantage of being able to measure both the concentration of free amino acids and those which remain bound in the same portion of the same sample, and hence give an assessment of the state of preservation of the sample, in terms of the extent of natural hydrolysis.

The aim of this part of the study was to analyse the fossil amino acids from brachiopods and other samples in order to determine:

(i) The state of physical preservation of intracrystalline fossil molecules.

(ii) The potential decay pathways of the molecules.

(iii) The effect of time on the concentration of the amino acids within the soluble organic fraction of the shell.

In this study, several of the common proteinogenic amino acids, Met, Cys and His have not been quantified. Met and Cys are both degraded rapidly during acid hydrolysis by oxidation of the sulphur atom contained within their side chains (**section 3.8.2**). This loss is highly variable, and would introduce unacceptable errors into the results, as it would not be certain which losses are due to time, and which to the effect of the acid hydrolysis. His is a basic amino acid which decomposes rapidly (Vallentyne, 1964), but which is also very rare in intracrystalline proteins from brachiopods (**Chapter Five**), and was therefore not quantified from fossil samples.

7.2 Previous work

Natural hydrolysis (the decomposition of peptide bonds by the action of time, heat or pressure) of proteins and peptides results in amino acid monomer units being released from the protein structure. On the whole, these molecules are much more stable when in proteins, although Akiyama (1980) demonstrated that serine could undergo decomposition reactions whilst remaining peptide bound. Examination of the thermal stability of amino acids in an attempt to understand their distribution in the fossil record was completed in the first study of fossil amino acids (Abelson, 1954), where it was noted that the thermally unstable amino acids were either absent entirely, or present in very low concentrations in the fossil record. Experimentally, the thermal stability of alanine (one of the most common of the amino acids found in fossils) in dilute aqueous solution was examined between 188 and 310°C, and equations produced to determine the expected stability at normal temperatures and pressures. Alanine would require 10 000 Ma at 20°C in order for 63% of the molecule to decay (Abelson, 1954), hence no thermal destruction would be expected at normal temperatures.

The most common method of assessing the stability of amino acids has been completed by heating aqueous solutions of the pure amino acids (pyrolysis experiments). Provided that the decomposition reactions follow first order kinetics, it is reasonable to assume that the reactions which occur relatively rapidly at elevated temperatures can be used to extrapolate back to lower temperatures found under normal conditions. As outlined above, alanine was examined and was relatively stable (Abelson, 1954). Standard solutions of other pure amino acids, vacuum sealed in glass tubes, have also undergone pyrolysis reactions (Vallentyne, 1964; 1968). Following pyrolysis (during which no allowance was made for increased vapour pressure due to evaporation), the solutions were analysed by paper chromatography and amino acid analysis to determine reaction products, i.e. the compounds formed by decomposition reactions. These are summarised in **table 7.1**, which indicates that:

(i) Individual amino acids break down at differing rates.

(ii) The reaction products of some amino acids may be other amino acids.

Differing rates of decomposition of the molecules explains the selective preservation of these molecules in the fossil record. The production of proteinogenic amino acids from decomposition of other amino acids will explain the anomalous concentration of some amino acids noted in the study of fossils (Bada and Man, 1980). Thermal reaction kinetics were also calculated for Ser, Thr and Phe. At 20°C, thermal decomposition of these compounds would take approximately 0.1 Ma, 1 Ma and 10 Ma respectively, indicating the differences in reaction rates. Mixtures of these pure amino acid solutions did not affect the relative stability order of the amino acids as determined with single solutions (Vallentyne, 1964), although the rate of decomposition is higher in these

Amino acid	Temp (°C)	0.37 life	Organic ninhydrin positive products
Asp	230	1 hr	None (PC)
Glu	245	124 hr	γ-aminobutyric acid (PC)*
Gly	252	2-3 hr	Methylamine (PC)
Ala	243	56 hr	Ethylamine (PC)
Val	216	48 hr	Gly (PC, A^3)
	252	0.5-3 hr	Gly (PC, A^3)
Leu	252	0.5-2 hr	Unidentified
Ile	252	0.5-4 hr	Unidentified
Ser	185	15 hr	Gly (PC, A^3)
			Ala (PC, A^3)
			Ethanolamine (A^3)
Thr	180	2.7 hr	Gly (PC, A^3)
Met	216	6-10 hr	Gly (PC, A^3)
			Ala (PC, A^3) trace
Phe	230	40 hr	Phenethylamine (PC)
Tyr	252	1-4 hr	Gly (PC) trace
Pro	282	5-15 hr	None
Lys.HCl	216	0.5-3 hr	Unidentified
Arg.HCl	216	5-15 min	Pro (PC, A^3)
			Ornithine (PC)
			"Near ornithine"

Notes:
Identification methods: PC = Paper chromatography, A^3 = Amino acid analysis
Arg and Lys analysed as HCl salt
* = Apparent only after hydrolysis

Table 7.1: Reaction products of amino acids which have undergone pyrolysis reactions (after Vallentyne, 1964).

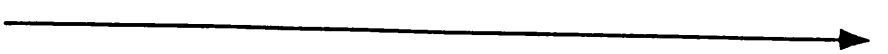
mixtures than in the single solutions, indicating some inter-reaction of the samples. It was important to consider the reactions of amino acids in mixtures, as they will naturally occur in such mixtures within the shell.

Pyrolysis experiments were also undertaken using shell powders heated in the presence of water (e.g. Hare and Mitterer, 1969; Totten *et al.*, 1972), but in this case the added water could only possibly affect the reactions of intercrystalline molecules. The results show that pyrolysis destroys Thr, Ser, Cys, His and Arg totally. In comparison with samples of the same species from the Miocene, the same amino acids were lost, indicating that the extrapolation of the results from high temperature experiments to lower temperatures and pressures is likely to be valid. A decrease in the relative stability of amino acids that are associated with the inorganic mineral phase was also noted, when compared to the analysis of pure compounds.

Amino acids may degrade into other amino acids. This was first noted by Vallentyne (1964; **table 7.1**) as the reaction products of pyrolysis. The majority of these reactions appear to be simple decarboxylation or deamidation reactions, producing products which are not detectable by the PITC derivatization/analysis method utilised in this study. Dehydration of the hydroxy-amino acids, Ser and Thr, produces alanine and α-aminobutyric acid via intermediates (Bada *et al.*, 1978; **figure 7.28**). This reaction also releases water, which will further degrade any remaining peptide bonds, especially when the molecules are trapped, as is the case within the inorganic phase, and reaction products are unable to escape. The presence of non standard amino acids such as α- and γ-aminobutyric acid and β-alanine (Hare and Mitterer, 1967) has been used to indicate contamination of the sample (Schroeder, 1975), as the α-decarboxylation reaction is rare in pyrolysis experiments. However, in closed systems, such as for intracrystalline molecules, the presence of these amino acids will indicate the change, by reaction, of protein amino acids, rather than contamination of samples by extraneous sources.

In summary, amino acids have decomposition rates which are very different from each other, but all of which follow first order reaction kinetics. The amino acids may be grouped in terms of their relative stability (**figure 7.1**), which also hold true for mixtures of amino acids. The reaction may proceed through to non-detectable products, non-standard amino acids or diagenetically produced amino acids which also naturally occur within proteins. The rate of reaction of amino acids is likely to be more rapid (Hare and Mitterer, 1969; Totten *et al.*, 1972) when associated with other features such as the inorganic matrix, than is predicted from the pyrolysis of the pure amino acids. All of these factors must be recognised prior to the investigation of fossil molecules.

Group 1	Group 2	Group 3	Group 4
Aspartic acid	Lysine.HCl	Tyrosine	Alanine
Cystine	Histidine.HCl	Glycine	Proline
Threonine	Methionine	Valine	Hydroxyproline
Serine		Leucine	Glutamic acid
Arginine.HCl		Isoleucine	



Increasing Stability

Figure 7.1: Relatively stability of the amino acids determined by pyrolysis experiments (after Vallentyne, 1964).

7.3 Methods and Materials

Samples were collected from the localities described in **Chapter Two.** The shells were prepared for amino acid analysis by the methods described in **Chapter Three.** In summary, shell surfaces were cleaned, and the shells powdered and bleached to destroy by oxidation any remaining intercrystalline molecules. This was followed by repeated washing with MilliQ™, freezing and lyophilisation of the sample. Powders were decalcified using HCl (2N) at a ratio of 11 μL acid to 1 mg shell powder, centrifuged to remove insoluble compounds, and loaded directly onto the analyser. Amino acid analysis was carried out both with and without hydrolysis in order to determine the proportion of amino acids present in the free state. With the exception of *T. sanguinea*, samples were mostly identified only to the generic level, rather than to the specific.

Analyses with hydrolysis were repeated at least three times, and those without hydrolysis at least twice to ensure reproducability. Results were converted to weights (ng/mg) using the spreadsheet program EXCEL™, and collated by averaging out the multiple analyses. These figures were then graphed using the graphical program Cricket Graph™ to give a visual representation of the results.

7.4 Results

The results are summarised in the tables contained within **Appendix 1.** Analyses are expressed in the following ways; bulk amino acid composition, individual amino acid concentration, %change of each amino acid, concentration of free amino acids and the proportion of the amino acids which are present in the free state (%free). It should be noted at this point that the data for the concentration of combined amino acid could be an underestimate, as the samples are calibrated using derivatization only standards, which do not take into account losses caused by hydrolysis (**section 3.8.2**). In normal circumstances these losses will be negligible, but it is important to identify such possibilities in order to account for anomalies. The significance of the relationship between age and concentration of each amino acid is also assessed by curve fitting and the calculation of correlation coefficients. For a correlation to be considered significant, the value of R^2 needs to be greater than 0.750.

7.4.1 Bulk amino acid composition

Figure 7.2 shows the total amino acids present in samples from the South Wanganui Basin compared with the age of the sample. All samples analysed contain appreciable amounts of amino acid, ranging between 580.79 and 23.22 ng/mg, and show a general overall decreasing trend in the concentration of the amino acid through time, although in all species examined, with the exception of *Notosaria*, there appears to be an

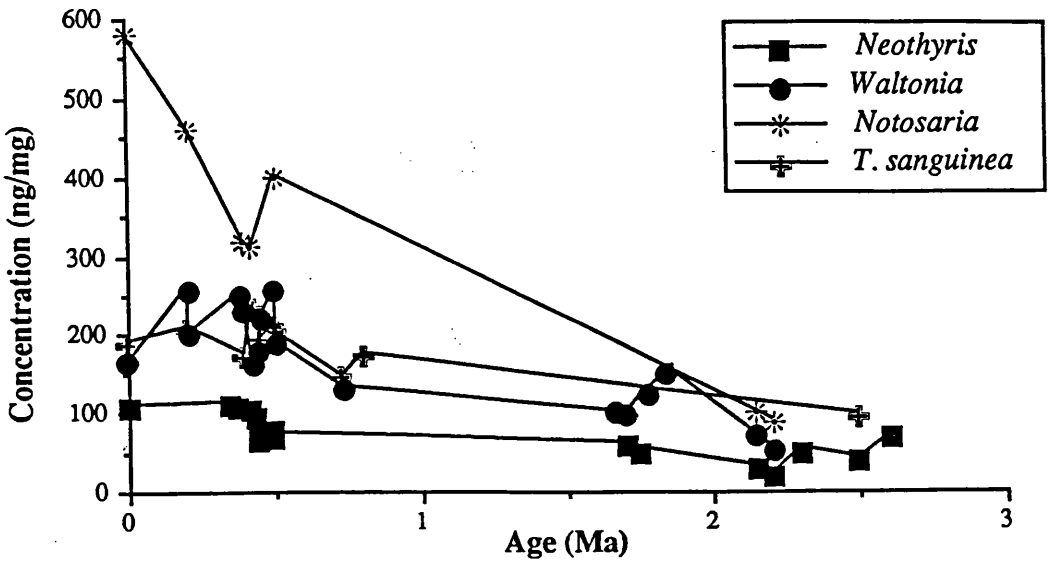


Figure 7.2: Total amino acids present in the acid soluble organic extract from Recent and fossil samples compared to the age of the samples. There is a general decrease in the concentration over time, although some samples show an increase due to the solubilisation of the acid insoluble protein found in Recent brachiopods.

initial rise in the concentration of the amino acids (between the Recent samples and the youngest fossils), before the concentration decreases with time. This initial increase differs in magnitude and in the time which it persists. For example, in *Neothyris* the increase is only to 2%, well within the range of experimental error, which is here estimated to be between 5 and 10%. However, *Waltonia* shows a 59% increase in the concentration of amino acid in the Rapanui Marine Sand and in the Kupe Formation (separated by 0.3 Ma), with slightly lower values between.

The proportion of amino acids which are present in the free state rises from negligible amounts in the Recent to greater than 58% by 0.2 Ma (**figure 7.3**). This suggests that the proteins within the shell are undergoing rapid natural hydrolysis. The %free also fluctuates with respect to time, although remaining greater than 60% in all fossil samples, the value varies up to 100% in *Waltonia*, 98.95% in *Neothyris* and 89.46% in *T. sanguinea*. The lack of pattern in the %free data for the total amount of amino acid contained within the shell, and the corresponding concentration of amino acid which is present in a bonded state (either original or diagenetic bonding), is indicative of a complex system within the shell, whereby individual amino acids react at different rates as a response to degradative factors. The free amino acids present in *Notosaria* are the exception to this rule, as these amino acids decrease in concentration over time, without any such fluctuations (**figure 7.4**), although this may be due to *Notosaria* having fewer data points than the other species.

Samples of *Neothyris* collected from the Rapanui Marine Sand show very low concentrations of amino acid (when compared to the other samples), which is indicative of the samples being older than first thought. This is discussed in greater detail in the

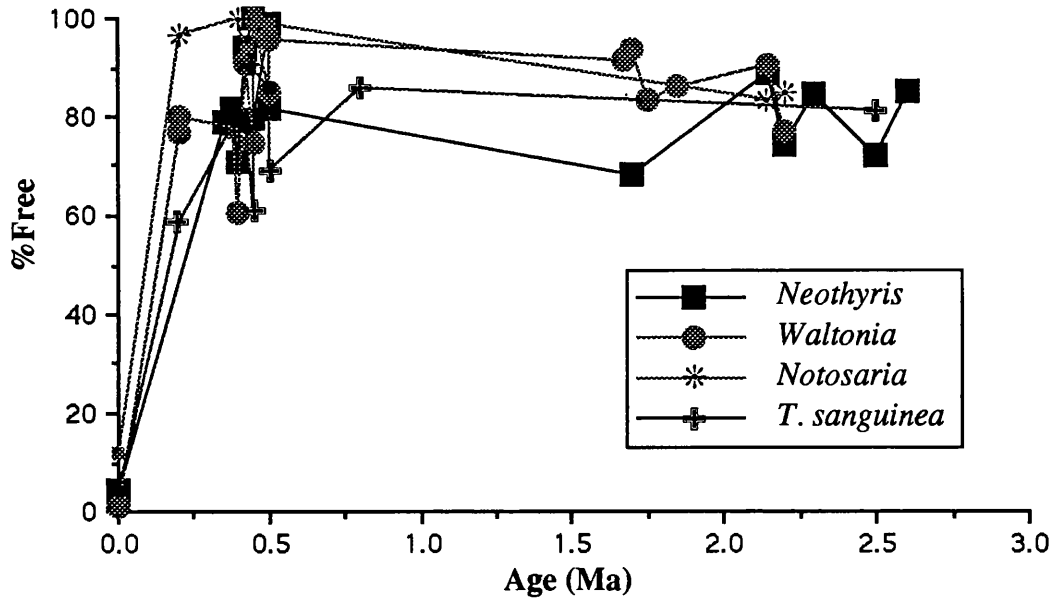


Figure 7.3: Total amino acids %free v time. Note the very rapid rise in the %free of the amino acids, indicating the rapid degradation of the constituent proteins.

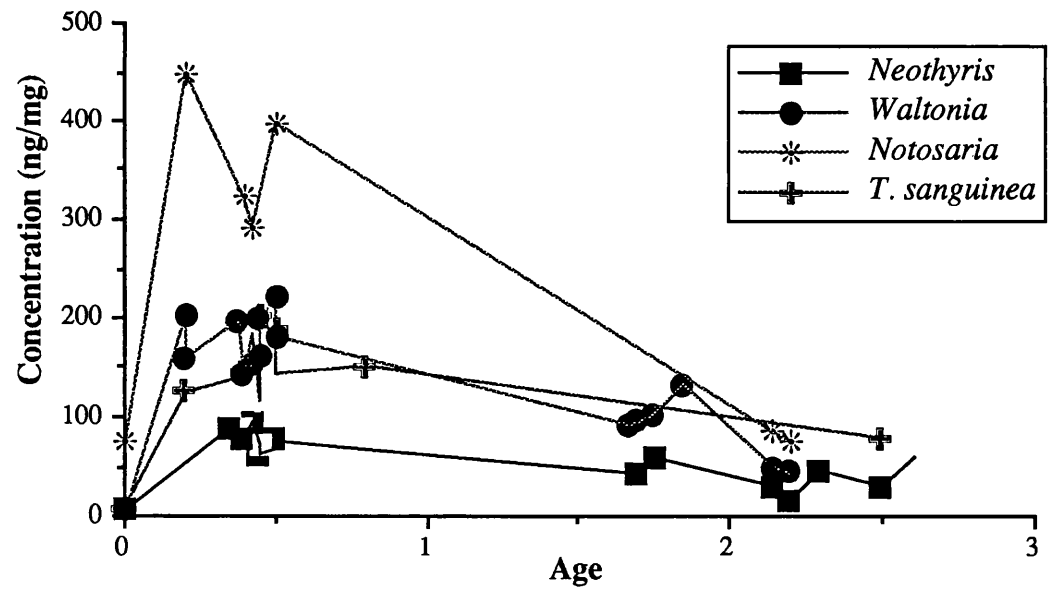


Figure 7.4: Concentration of free amino acids over time. The samples show a rapid increase representing the degradation of the proteins, followed by a slower decrease in the concentration representing the degradation of the amino acids from the free state.

following chapter (**section 8.4**).

The bulk amino acid composition of the samples are subject to change due to the individual amino acids in the proteins. In order to understand these fluctuations, it is necessary to examine the data for each individual amino acid.

7.4.2 Individual amino acids

(i) Aspartic Acid (including asparagine)

Aspartic acid shows an exponential decrease in concentration over time (**figure 7.5**), with correlation coefficients (R^2) being as follows:

Notosaria	0.977
T. sanguinea	0.914
Neothyris	0.894
Waltonia	0.872

All of these figures are highly significant and represent very rapid initial breakdown of the aspartic acid and asparagine molecules, followed by slowing of that rate. More than 80% of the Asp/Asn molecules have been lost from the acid soluble portion of shell organics

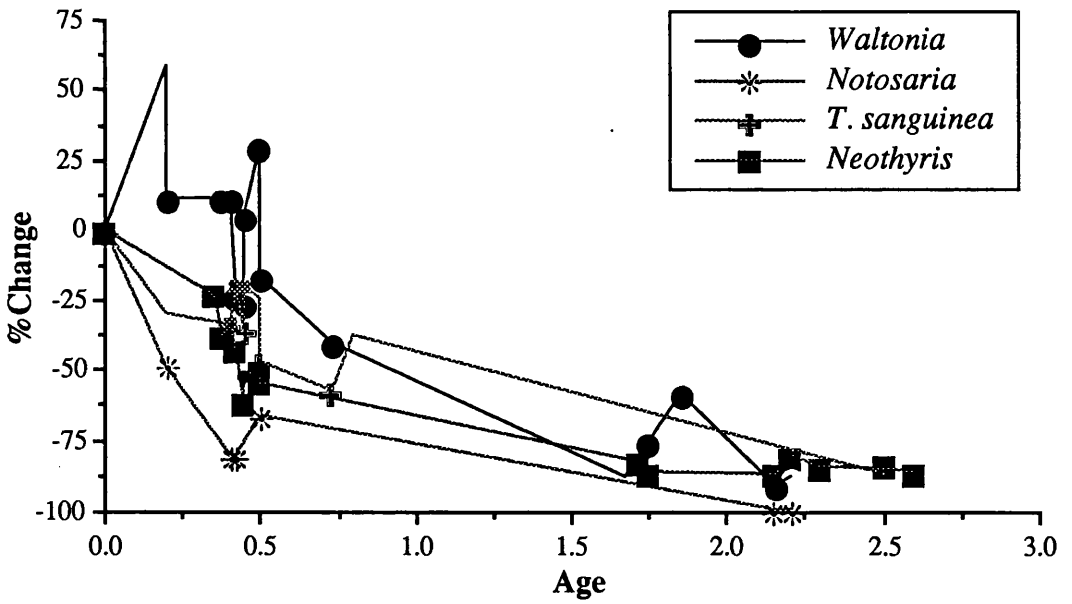


Figure 7.5: %change in concentration of Asp compared with time.

by 2.6 Ma. In *Waltonia*, the concentration of the amino acid rises from 8.23 ng/mg to 13.06 ng/mg between the Recent and the Rapanui Marine Sand, before decreasing to 1.05 ng/mg in the Hautawa Shellbed. This is the only sample to show such an increase in Asp/Asn concentration.

The proportion of the amino acid present in the free state varies between samples. In *Neothyris*, there is a rapid rise to c. 95% free (0.38 Ma), followed by a maintenance of this proportion, even though the overall concentration of the molecule

decreases over this time (**figure 7.6**). This pattern is also shown by the molecules in *Notosaria*, although in both *Waltonia* and *T. sanguinea*, the rise in the rate of increase of the proportion of uncombined molecules is somewhat slower, the same level is reached by 0.5 Ma. This level is maintained throughout the rest of the samples, until complete destruction of the molecule (in the free state) in *Waltonia* takes place by the Upper Okiwa Group (2.15 Ma).

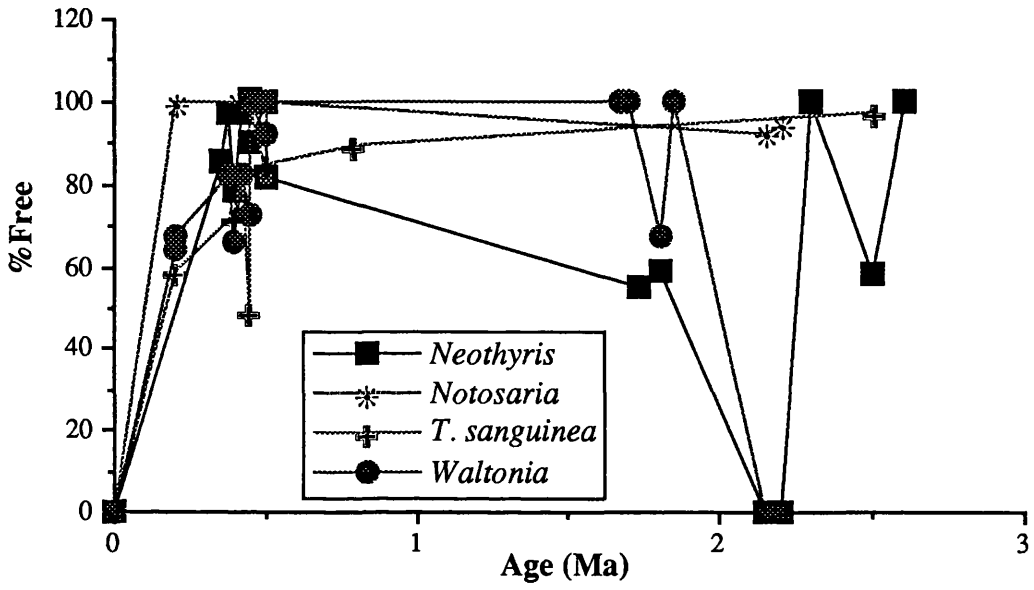


Figure 7.6: Proportion of Asp present in the free state over time.

(ii) Glutamic acid (including glutamine)

Glutamic acid shows an exponential decrease in concentration (**figure 7.7**), with R^2 values as follows:

Notosaria	0.922
T. sanguinea	0.825
Waltonia	0.748
Neothyris	0.740

These figures are again significant, indicating some parity in the behaviour over time of the acidic amino acids, although Glu is a more stable molecule than Asp (seen by the %change data). All of the samples show an increase in the concentration of the amino acid in the youngest samples, followed by rapid breakdown of up to 80% of the initial concentration by 2 Ma. The proportion of free amino acids in the sample is low, in all cases being below 40%, and generally below 30% of the total present. The proportion shows a rapid increase to this level, followed by little fluctuation throughout the rest of the period under study.

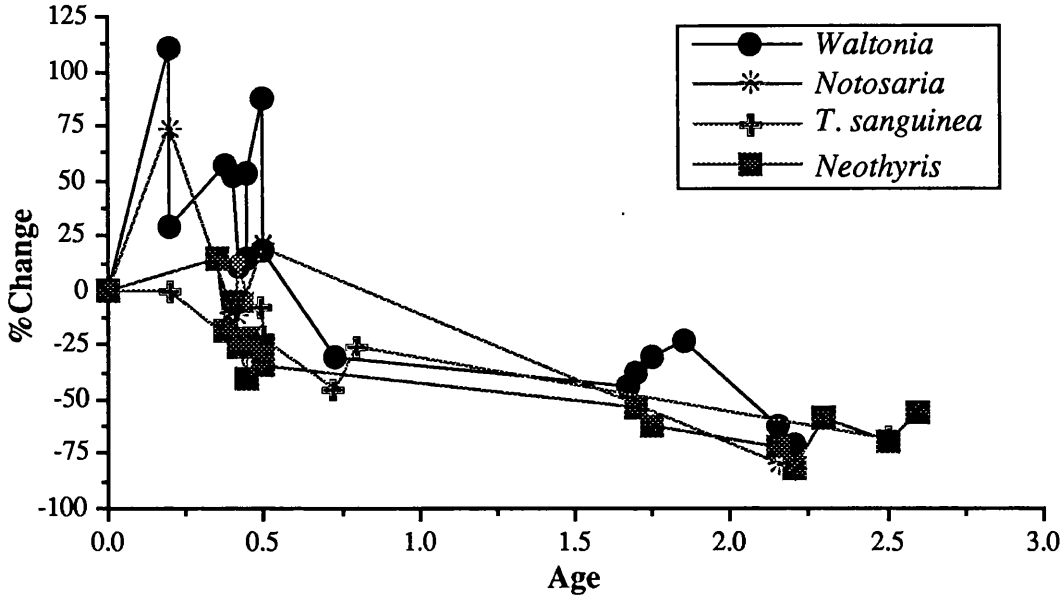


Figure 7.7: %change in concentration of Glu/Gln compared with time.

(iii) Serine

Serine decays rapidly, with the decay curves being almost two component (**figure 7.8**). The data does not follow an exponential curve, but instead follows a logarithmic curve very closely, with values for R^2 as follows:

T. sanguinea	0.924
Notosaria	0.863
Neothyris	0.848
Waltonia	0.684

The concentration of the molecule drops rapidly, until c. 80% of the original concentration has been lost after only 0.2 Ma. Following this rapid initial decrease, however, the level

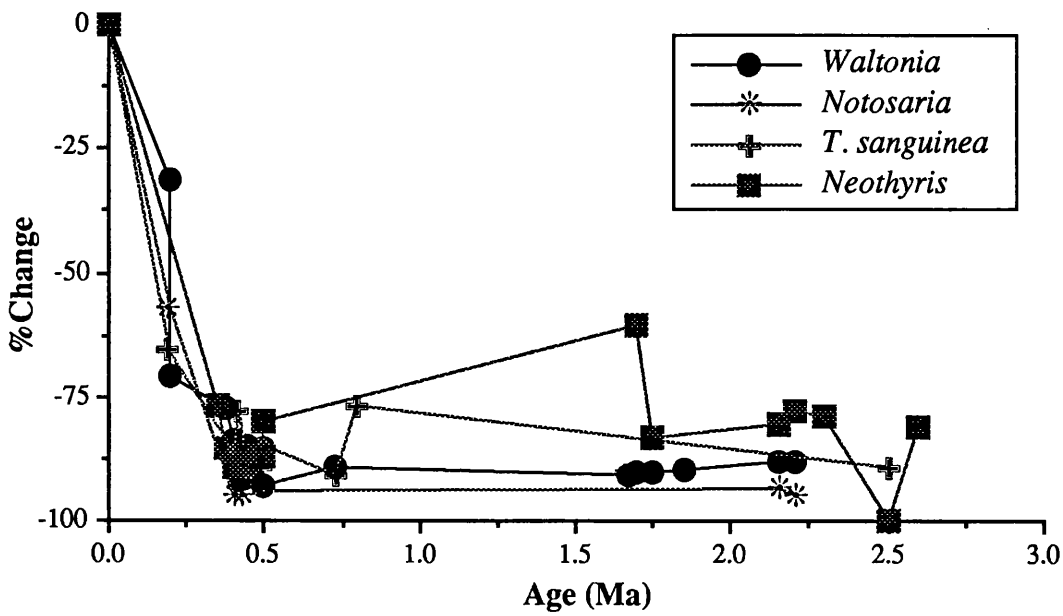


Figure 7.8: %change in concentration of Ser compared with time. Note the very rapid degradation of the molecule over time.

of %change remains at a similar level. This is also seen in the graphs of concentration against time (**figure 7.8**). In *Waltonia*, for example, the concentration (free and combined) decreases from 6.94 ng/mg in the Recent to 1.01 ng/mg by 0.5 Ma, but by 2.3 Ma this has only been reduced to 0.83 ng/mg.

The proportion of Ser present in the free state rapidly increases to a maxima and then decreases, forming a bell shaped curve (**figure 7.9**). In all samples, the %free rises rapidly until 90% of the molecules are free by 0.5 Ma, although in *Waltonia*, the rise does not appear as pronounced. However, the level of Ser in the free state decreases rapidly from this level, until no Ser molecules remain in the uncombined state. The concentration of the molecules therefore indicates that any Ser which remains in the sample after c. 0.7 Ma is bound to another molecule via an HCl sensitive bond.

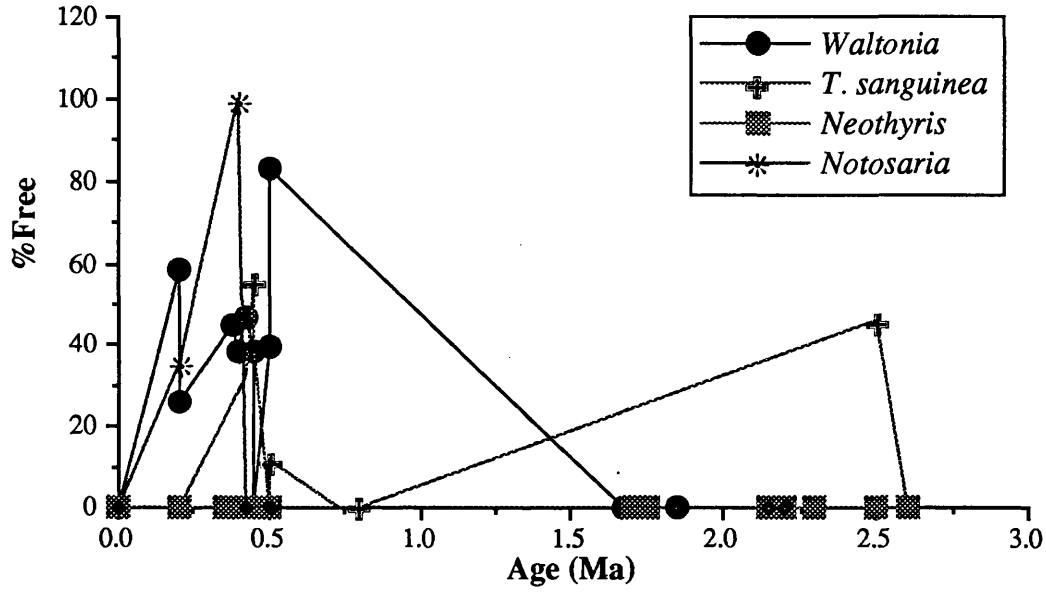


Figure 7.9: Proportion of Ser present in the free state over time. The low percentage of the Ser in the free state in the older samples represents the decomposition of the molecule from the free state rather than when peptide bound.

(iv) Glycine

The concentration of Gly within the samples over time is highly variable (**figure 7.10**). In *Notosaria*, the degradation of the molecules follows an exponential curve, with a correlation coefficient of 0.937, the overall concentration dropping from 218.20 ng/mg in Recent samples to 40.09 ng/mg by the Hautawan (2.2 Ma). However, in *Waltonia*, *T. sanguinea* and *Neothyris* the change in the concentration over time cannot be satisfactorily described by either an exponential or logarithmic curve. All of these samples show a general decreasing trend over time, with up to 80% of the molecules being degraded by 2.3 Ma, although there is considerable variation in concentration over this time period (**figure 7.10**). This indicates that the concentration of Gly within the shell may well be controlled by factors other than the decomposition of the molecule over time.

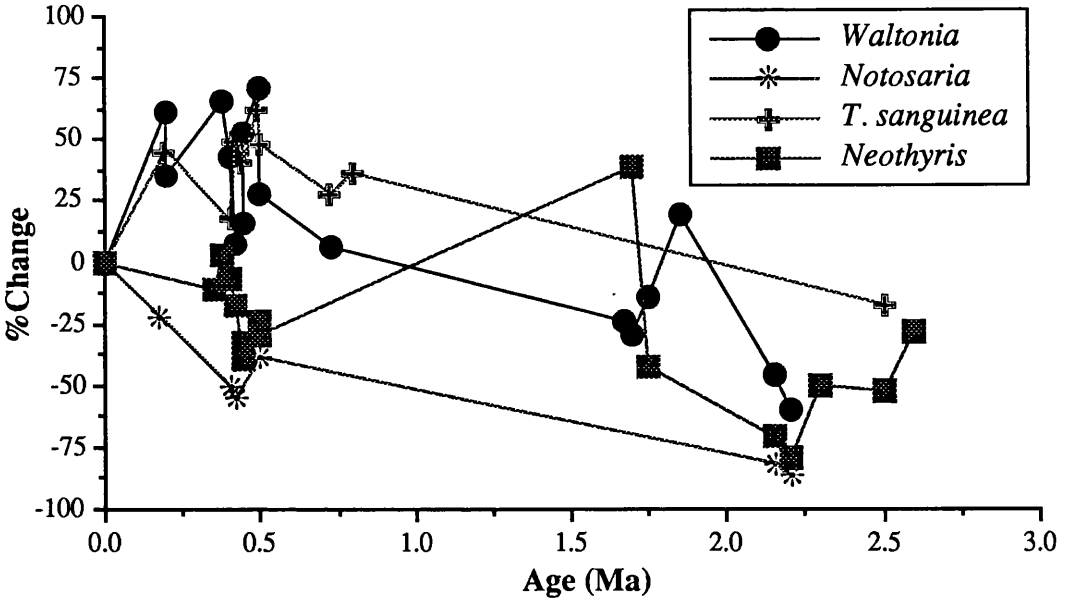


Figure 7.10: %change in concentration of Gly compared with time.

The proportion of Gly molecules that are present in the free state show similar patterns for all of the species examined. This proportion shows a rapid rise to greater than 75% free, followed by a maintenance of this level. The oldest of the samples show a slight decrease in the proportion of free amino acids, indicating that Gly decays whilst in the free state, and that the decomposition of peptide bonds occurs at a similar or slightly lower rate than that of the destruction of the molecule in the free state.

(v) Arginine

Arg decays rapidly in all samples analysed. The pattern of decay, however, is not as obvious as for other amino acids. In both *Neothyris* and *Notosaria* the relationship is clearly exponential, whereas in *Waltonia* and *T. sanguinea* this is not as clear (**figure 7.11**). The values for R^2 are as follows:

Notosaria	0.972
Neothyris	0.941
Waltonia	0.607
T. sanguinea	0.185

Although the R^2 value for *T. sanguinea* is very low, the decay is still very rapid (**figure 7.11**), falling from 6.28 ng/mg in Recent samples to 2.20 ng/mg at 0.2 Ma. The rate of decay then slows, so that by 2.5 Ma, there is still 0.99 ng/mg remaining. When other curves are selected, the degradation of Arg in *T. sanguinea* may be described by a logarithmic curve, with a correlation coefficient of 0.869. Although the pattern of destruction of the molecules are described by different types of curve, they both have essentially the same interpretation. Arg concentration in *Notosaria* shows a slight increase between the Recent and Rapanui Fm.

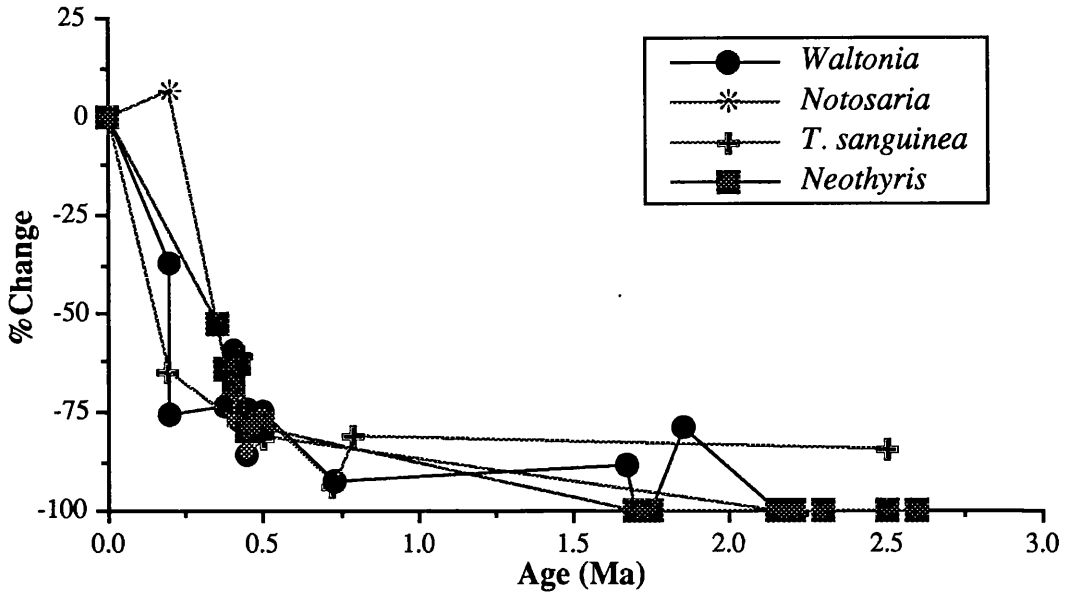


Figure 7.11: %change in concentration of Arg compared with time.

The proportion of free Arg molecules over time have a similar shaped curve for all four samples analysed. There is a rapid increase in the proportion, until almost 100% are free by the Upper Castlecliff Shellbed (0.38 Ma). This proportion decreases rapidly to 0% free by 1.5 Ma (**figure 7.12**). This indicates that any Arg which remains in older samples must be bound, via an HCl sensitive bond, to another molecule, and that this second molecule and bond must act to stabilise the amino acid.

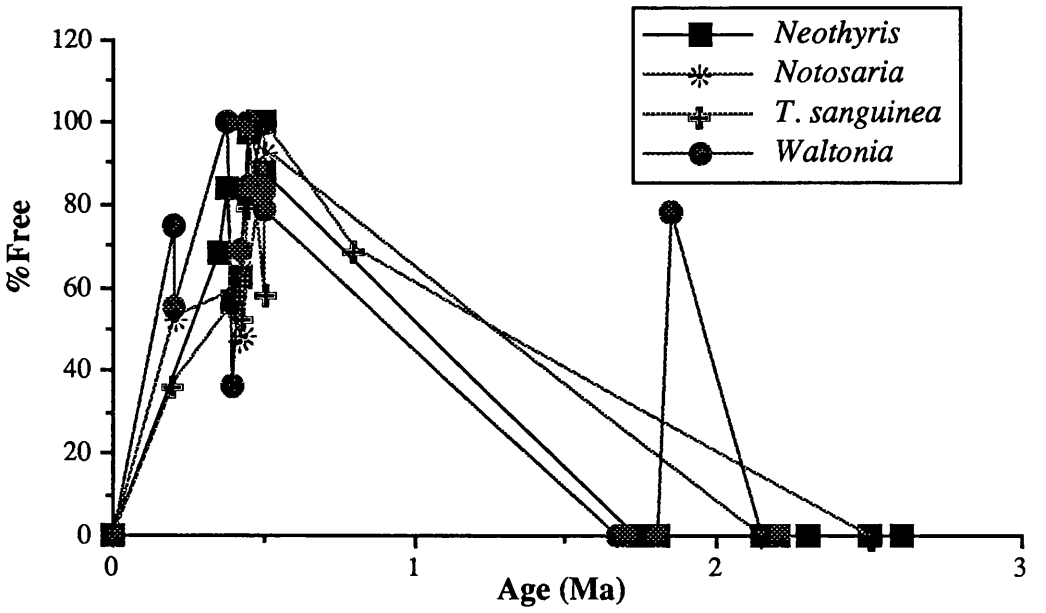


Figure 7.12: Proportion of Arg present in the free state over time. Arg is rapidly decomposed in the free state once released from the peptides by natural hydrolysis.

(vi) Threonine

Thr also decays rapidly in all cases (**figure 7.13**), although neither exponential nor logarithmic curves fit all the data, with only *Neothyris* (0.760) being described by an exponential curve. Other samples show rapid degradation, although the decomposition is not adequately described by curves.

The concentration of Thr drops very rapidly; in *Neothyris*, all Thr is degraded within 0.5 Ma, whereas this level is not reached until the Nukumaruan (1.85 Ma) in

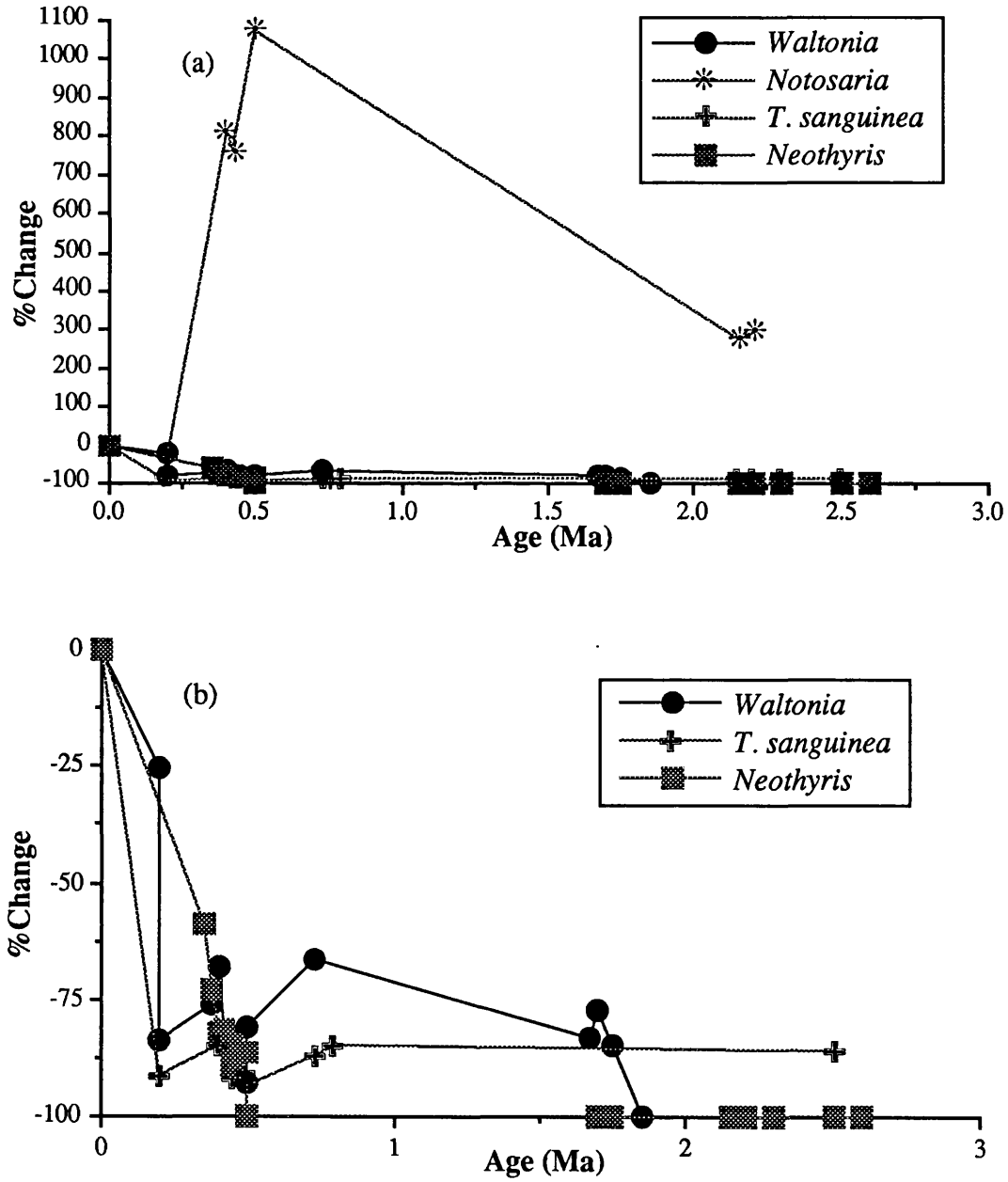


Figure 7.13: (a) %change in concentration of Thr compared with time, with a large increase in the concentration of Thr in *Notosaria*. (b) %change in the concentration of Thr in samples other than *Notosaria*, showing the very rapid decline in the concentration of Thr in the rest of the samples.

samples of *Waltonia*. In samples of *T. sanguinea*, Thr is preserved for much longer in the fossil record, with bound Thr occurring in samples up to 2.5 Ma, at a concentration representing 90% loss of the molecule. *Notosaria* shows an increase in the concentration of Thr (**figure 7.13**), contradicting the observation for the other three samples. Thr increases by 1075.04% (from 5.89 to 69.21 ng/mg) by the Kupe Fm (0.5 Ma), all of which is present in the free state. This is obviously an anomalous situation, which will be discussed in detail below.

The proportion of Thr which is present in the free state rapidly rises to almost 100% (**figure 7.14**), and this level is maintained until the free molecules are completely degraded. There is not such a pronounced downcurve in the %free for Thr as for Arg, indicating that Thr is slightly more stable than Arg in the free state. The increase in the concentration of Thr appears only to occur within the portion of the sample which is present in the free state.

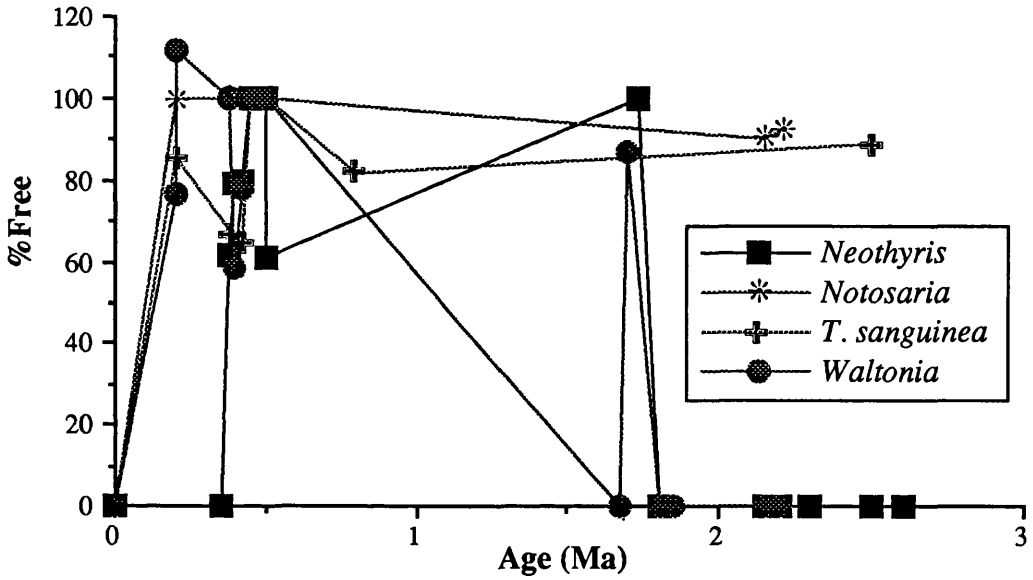


Figure 7.14: Proportion of Thr present in the free state over time. In some cases there is apparently no Thr present in the free state. This is due to the concentration in these samples being lower than the threshold value of the analysis system.

(vii) Alanine

The concentration of Ala within the shells of the samples is an essentially random spread with respect to age. There is a large increase in the concentration between the Recent and the fossil (**figure 7.15**), with a greater than 280% increase in Ala concentration in samples of *Neothyris* from the Landguard Sand (0.35 Ma). In all cases, the concentration undergoes a large and geologically rapid increase on transition from the living to the fossil, with a maxima at approximately 0.5 Ma. This maxima is followed by a decreasing trend, although in all cases the concentration of Ala in the oldest samples is

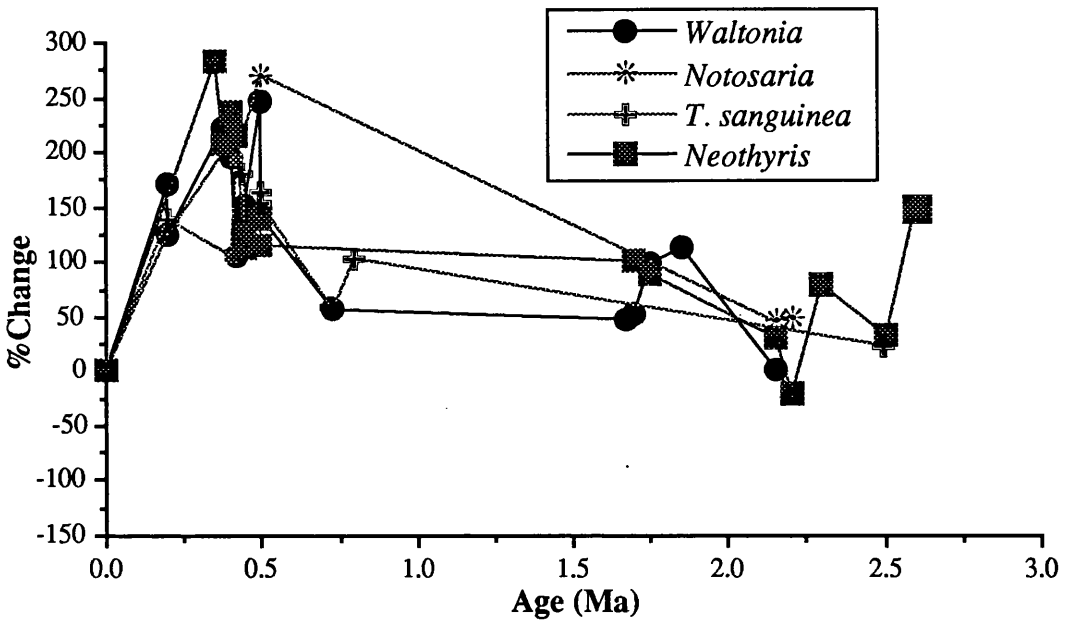


Figure 7.15: %change in concentration of Ala compared with time. All samples show a large increase in the concentration of Ala.

similar to, or greater than, the initial concentration found in Recent samples. This indicates that there are other controls regarding the concentration of Ala.

The maxima of Ala concentration coincides with the highest proportion of molecules present in the free state. There is a rapid increase in the proportion of free amino acids with greater than 90% being free in most cases by 0.5 Ma, and in some cases the proportion reaches 100%. A high percentage of free molecules remains throughout the rest of the samples studied, indicating that the increase in concentration of Ala must occur in the free state.

(viii) Proline

Pro molecules rapidly decay in the fossil record, with the decay being described by an exponential curve (**figure 7.16**). R^2 values for the data are as follows:

Notosaria	0.982
T. sanguinea	0.809
Waltonia	0.780
Neothyris	0.674

These figures are, for the most part, once again highly significant, the only exception being *Neothyris*, which has a large spread of data. The data for *Notosaria* and *Waltonia* show a slight initial increase in Pro concentration. Proline decays to a level representing greater than 80% loss over the 2.5 Ma of the study, although some remains both in the bound and the free state. In *Waltonia*, Pro concentration rises from 12.85 ng/mg in the Recent, to 20.45 ng/mg in the Rapanui Fm, prior to decreasing to 2.93 ng/mg in the Hautawa

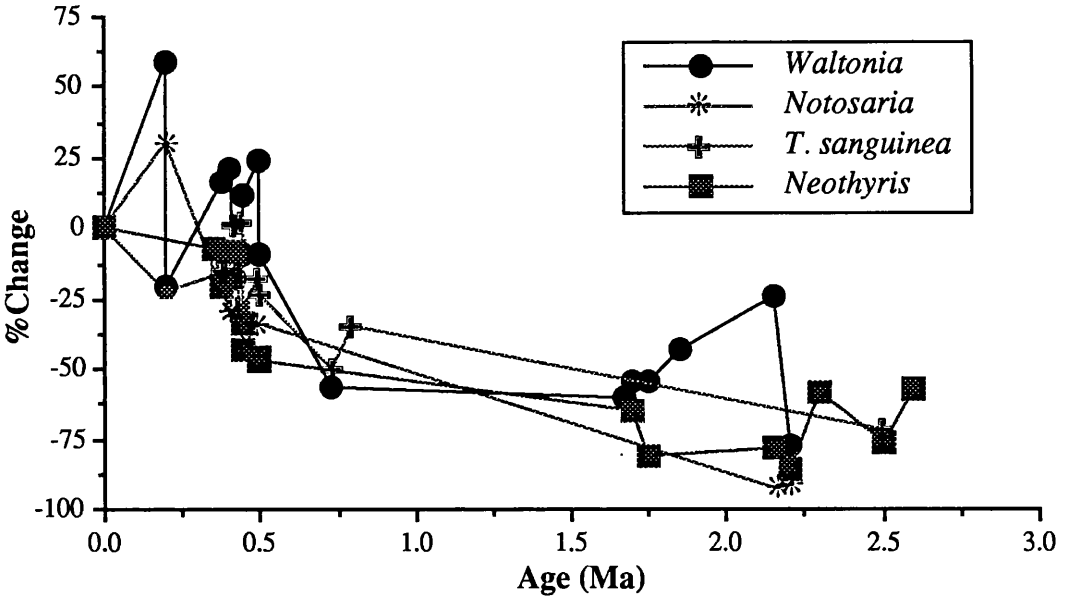


Figure 7.16: %change in concentration of Pro compared with time.

Shellbed.

The proportion of Pro that is present in the free state rapidly increases to greater than 80% in 0.2 Ma, and maintains that proportion for the remaining 2 Ma. The concentration of free amino acids rapidly decreases after this initial increase, indicating that the loss of free amino acids by decomposition must not be exactly mirrored by the release of Pro from the peptide bonds. There must, however, be some relationship, as if Pro were being lost from both the free and peptide bound state, the concentration of free amino acids would remain approximately the same.

(ix) Tyrosine

The concentration of Tyr in fossils follows a similar pattern to that of other amino acids (**figure 7.17**) showing a rapid decrease in concentration over time. In *Notosaria*, there is close correlation to an exponential decay curve (R^2=0.934). The remainder of the samples, however, do not correlate above the chosen level of significance. Although the decay of Tyr may not be described by these curves, there is a general decrease in the concentration over time. For example, the concentration of Tyr in *Waltonia* decreases from 2.94 ng/mg in the Recent to 0.79 ng/mg by the Hautawa Shellbed. All samples except *Notosaria* show an initial increase in the Tyr concentration when compared to that of the Recent. This initial increase peaks at 0.5 Ma, where there is up to 100% increase in Tyr concentration (**figure 7.17**). Subsequent to this, the concentration drops rapidly to greater than 80% loss by 2.5 Ma.

The proportion of Tyr molecules present in the uncombined state rises to 100% by 0.5 Ma in all species studied (**figure 7.18**), although the rate of increase in *T. sanguinea* is slightly slower than that of the other species. The proportion is maintained throughout

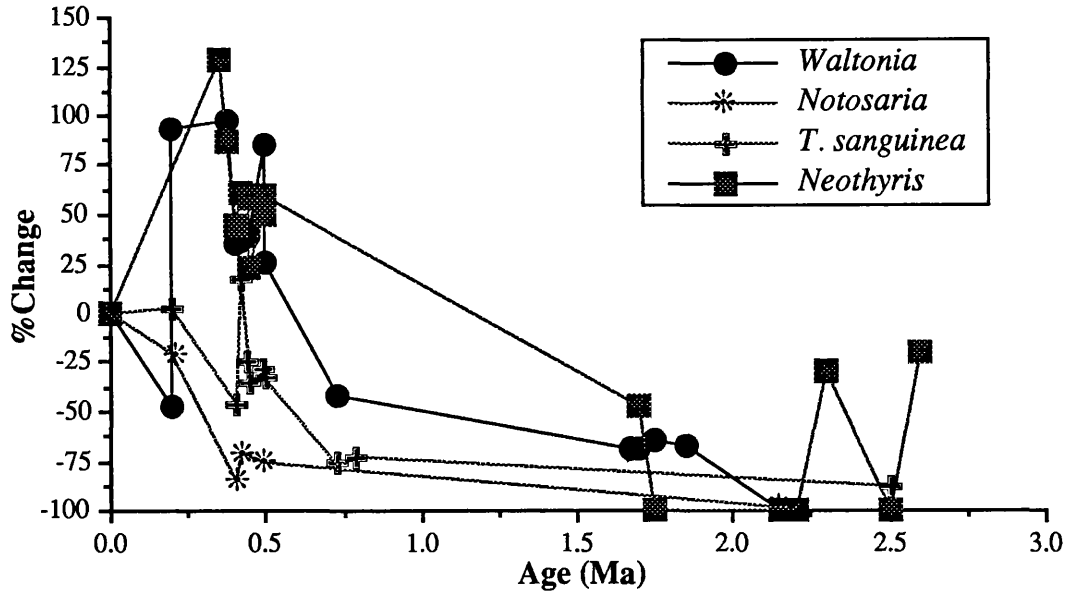


Figure 7.17: %change in concentration of Tyr compared with time.

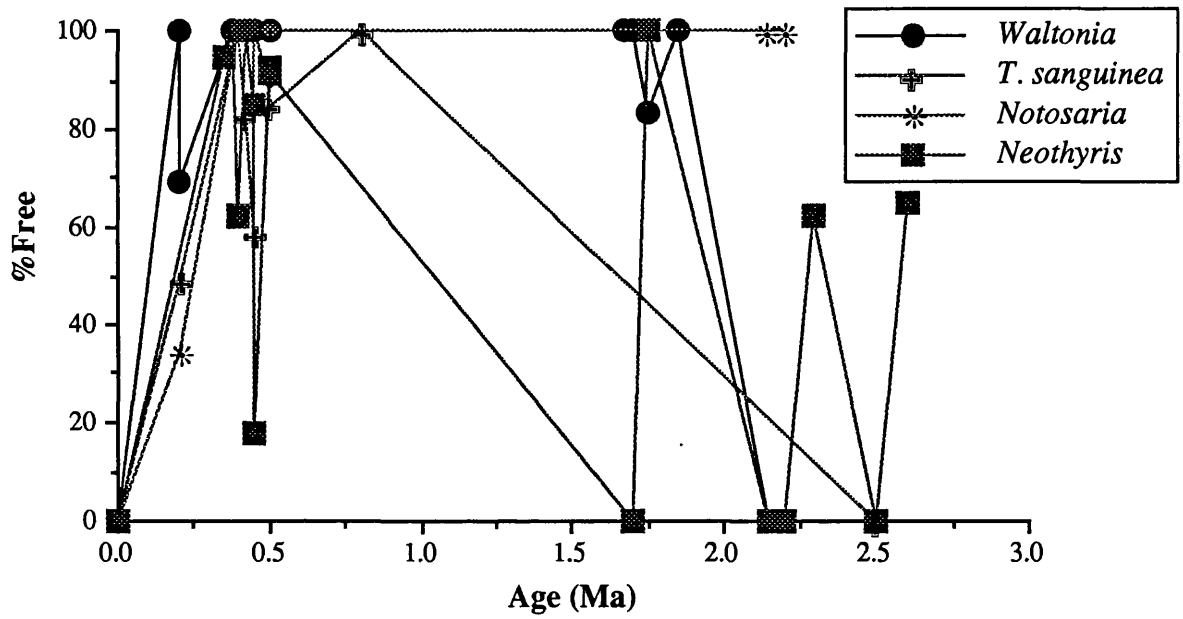


Figure 7.18: Proportion of Tyr present in the free state over time, showing a rapid increase in the %free.

the rest of the samples, except where the concentration of Tyr has dropped so low as to fall below the threshold value of the analysis system.

(x) Valine

Concentrations of Val are variable through time, although the graphs indicate a general decrease (**figure 7.19**). The degradation may be described by an exponential curve in *Notosaria* (R^2=0.810), where the concentration decreases from 7.90 ng/mg to

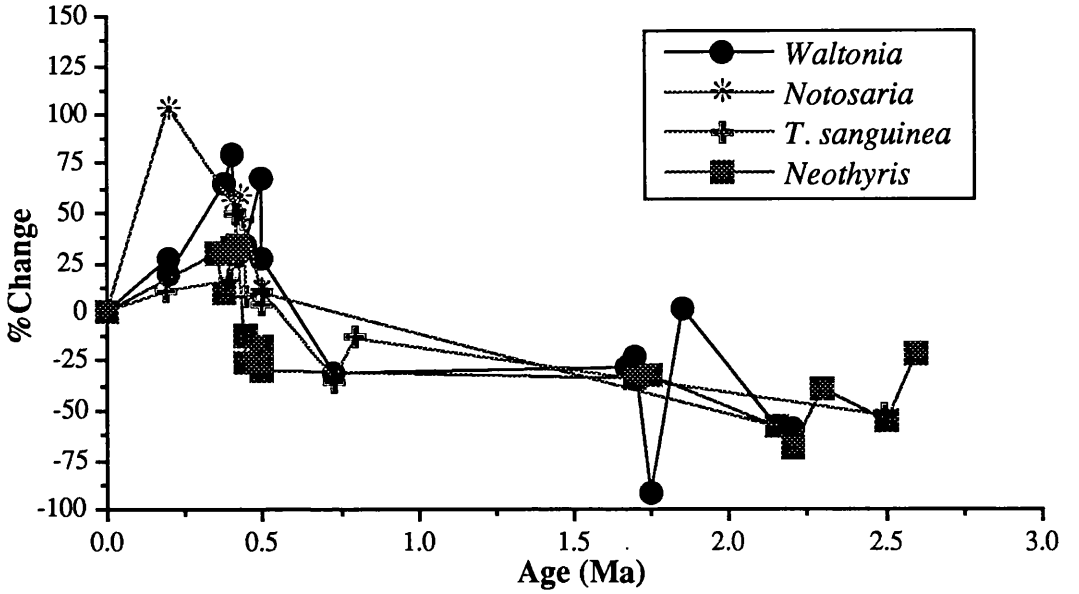


Figure 7.19: %change in concentration of Val compared with time.

3.11 ng/mg, although the remaining samples have a larger spread of data and have correlation coefficients below that considered significant (0.750). All samples show a large (up to 100%) initial increase in Val concentration between Recent and young fossil samples. The maxima of this increase is again at 0.5 Ma, after which there is a decrease to approximately 50% of that found in the Recent (**figure 7.19**). These figures show that Val is one of the most stable of the amino acids.

The proportion of the molecules present in the free state rises to greater than 80% in 0.2 Ma (0.5 Ma in *T. sanguinea*; **figure 7.20**), but then shows a slight decrease over time, indicating that the rate of degradation of Val is more rapid in the free state than in

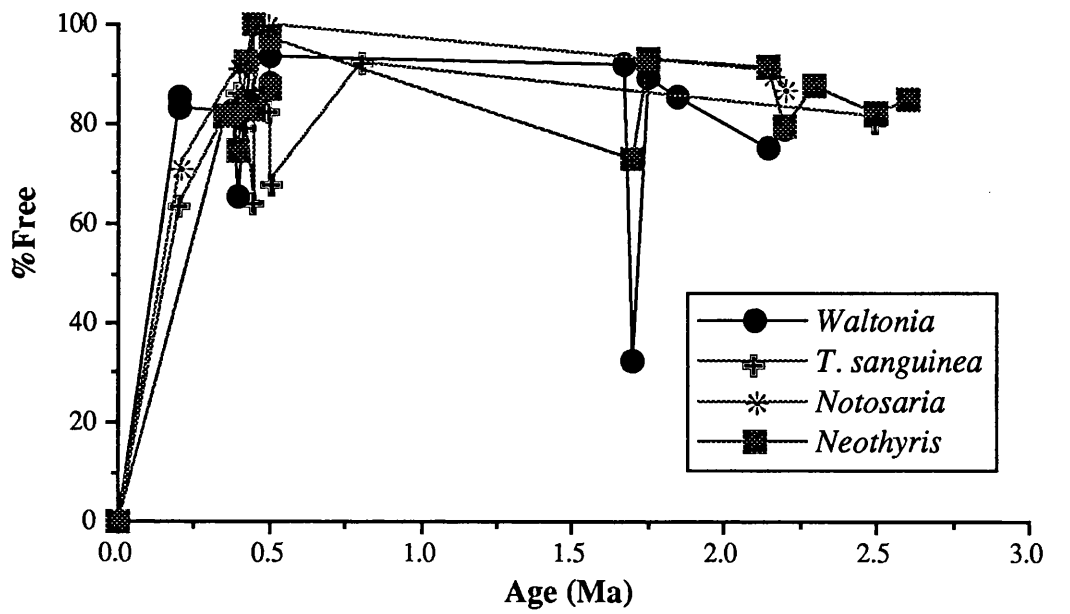


Figure 7.20: Proportion of Val present in the free state over time.

the bound state, and that once the peptide bonds are initially broken, the rate of release slows dramatically.

(xi) Isoleucine

The degradation of Ile over time may be described by an exponential curve (**figure 7.21**). R^2 values are as follows:

Notosaria	0.831
Waltonia	0.768
Neothyris	0.755
T. sanguinea	0.637

With the exception of *T. sanguinea*, these values are highly significant. All samples except *T. sanguinea* show an increase in the concentration of the molecule from the Recent. The concentration of Ile in *Neothyris* is variable between 0.35 and 0.5 Ma, where there is variation between 2.64 ng/mg (Pinnacle Sand) and 1.27 ng/mg (Lower Castlecliff Shellbed). This compares to an initial concentration of 1.78 ng/mg in the Recent sample. Following the initial increase in concentration, the molecule follows the exponential curve of decay, with all samples showing greater than 60% loss of molecules by 2.5 Ma.

The proportion of Ile present in the free state shows a rapid rise, up to greater than 80% free by 0.5 Ma. This proportion is maintained throughout the remaining samples. Such levels of free amino acid indicate that some Ile remains peptide bound in the fossil, even in older samples.

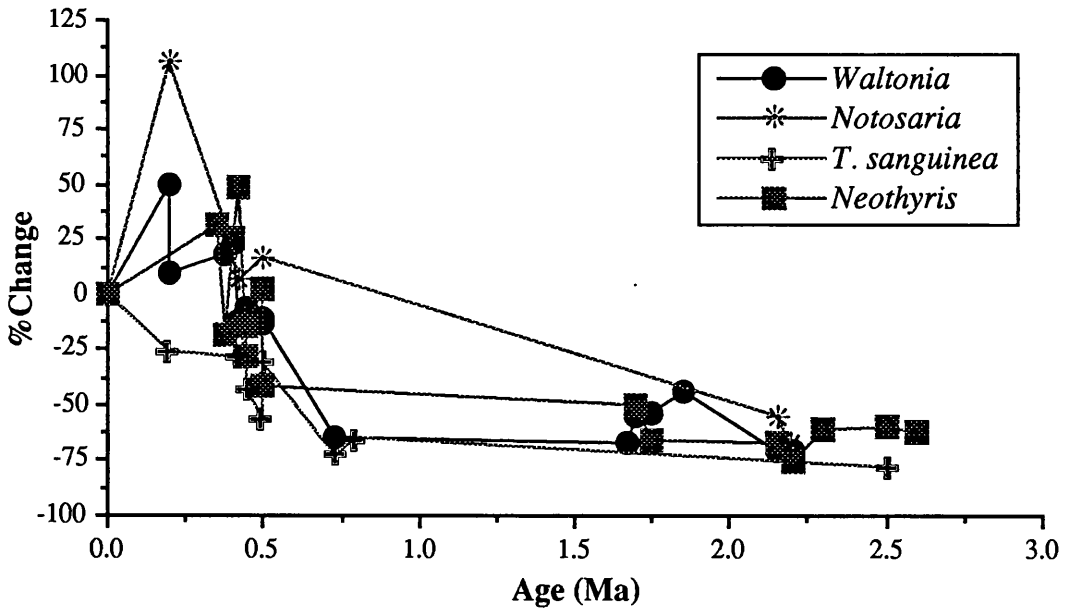


Figure 7.21: %change in concentration of Ile compared with time.

(xii) Leucine

The decay of Leu within the shells is rapid (**figure 7.22**), and three of the species show close correlation with exponential curves. R^2 values are as follows:

Notosaria	0.883
T. sanguinea	0.817
Waltonia	0.783
Neothyris	0.098

Although the data for *Neothyris* cannot be described by either an exponential or a logarithmic curve, it follows a similar pattern as the other samples. All samples (with the exception of *Notosaria*) show a slight initial increase in concentration, up to 150% at approximately 0.2-0.5 Ma (**figure 7.22**). After this increase, there is a decrease in concentration, reflecting the degradation of the molecules. In *Notosaria*, this decrease occurs from the Recent. Molecules are degraded to a level representing greater than 80% destruction.

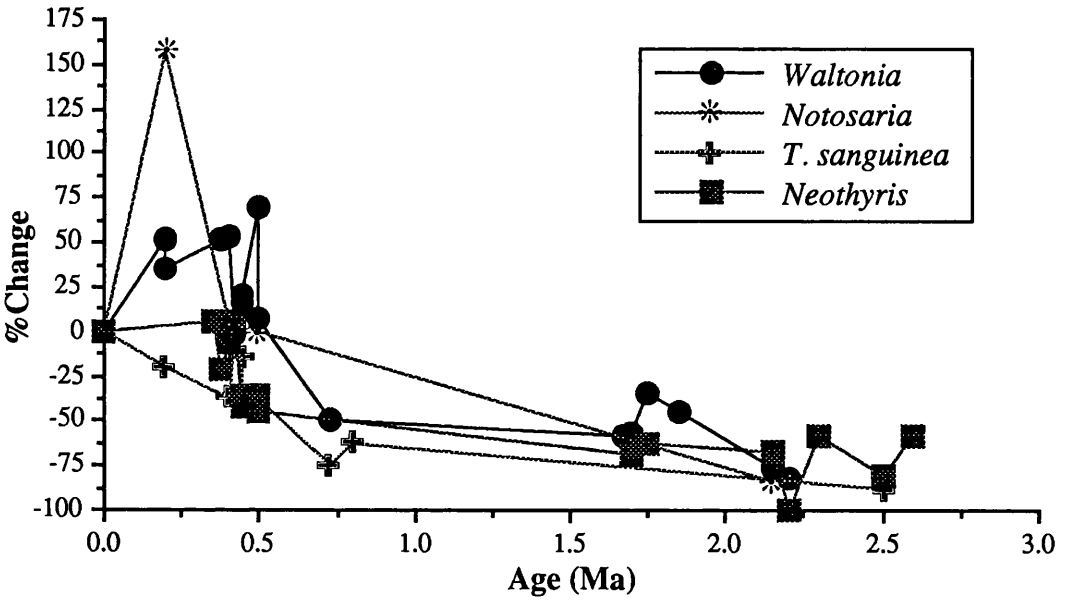


Figure 7.22: %change in concentration of Leu compared with time.

The proportion of free amino acids rapidly rises to greater than 80% by 0.5 Ma, and this level is maintained for most of the period under study, although in older samples the proportion of free amino acids decreases, indicating that the destruction of Leu in the free state is more rapid than if the molecule is present in the bound state, and that the most labile of the peptide bonds containing Leu have been degraded.

(xiii) Phenylalanine

With the exception of *Neothyris*, the concentration of Phe decreases in a pattern which shows good correlation to exponential curves (**figure 7.23**). R^2 values are

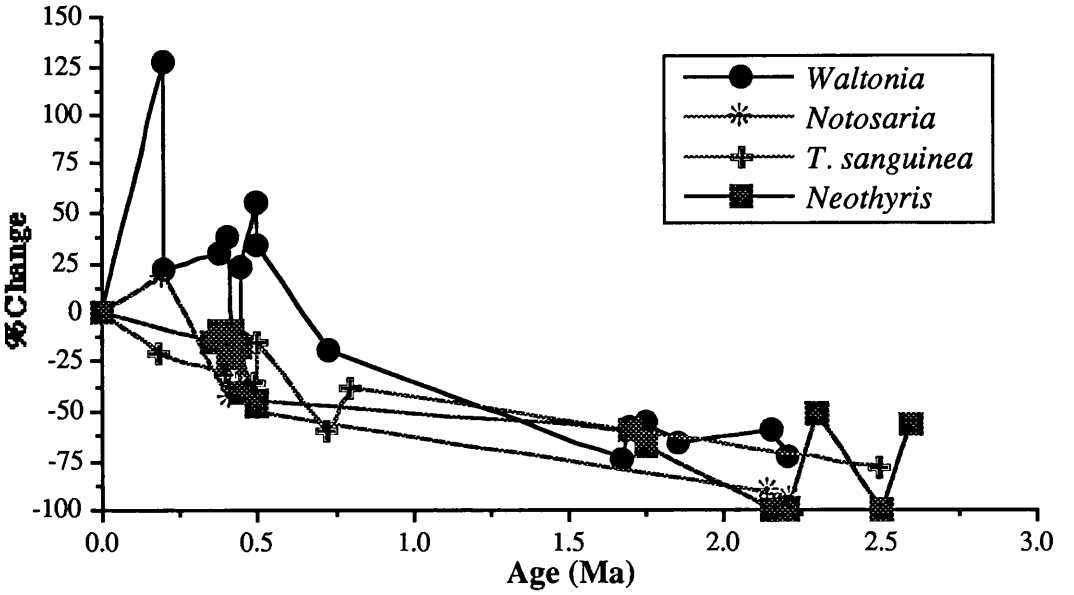


Figure 7.23: %change in concentration of Phe compared with time, including a very high increase in the concentration of Phe in *Waltonia*.

as follows:

Notosaria	0.969
T. sanguinea	0.832
Waltonia	0.819
Neothyris	0.347

Neothyris shows a large variation in the concentration of the older samples, which accounts for the low, non-significant, correlation coefficient. *Waltonia* and *Notosaria* both show a slight increase (from 2.22 to 5.03 ng/mg and 13.45 to 15.96 ng/mg respectively) in the Phe concentration up to 0.2 Ma. After this maxima, the concentration of Phe decreases in all cases until up to 80% is lost from the samples.

Free amino acids account for greater than 80% of the total preserved by 0.5 Ma. The rate of increase up to this level is much slower in *T. sanguinea* than in the other samples, a feature noticed for some, but not all of the other amino acids. This level is maintained throughout the remainder of the samples.

(xiv) Lysine

The concentration of Lys in the fossil record is variable (**figure 7.24**) and, except in the case of *Notosaria* (R^2=0.926), cannot be described by either exponential or logarithmic curves. The concentration, however, following an initial increase in some cases, shows a marked decrease with time, indicating greater than 90% degradation by 2.5 Ma. For example, in *Neothyris* Lys concentration decreases from 1.47 ng/mg in the Recent to 0.53 ng/mg in samples from the Upper Waipipi Shellbed (2.5 Ma).

The proportion of Lys present in the free state rises rapidly to 40% at 0.2 Ma

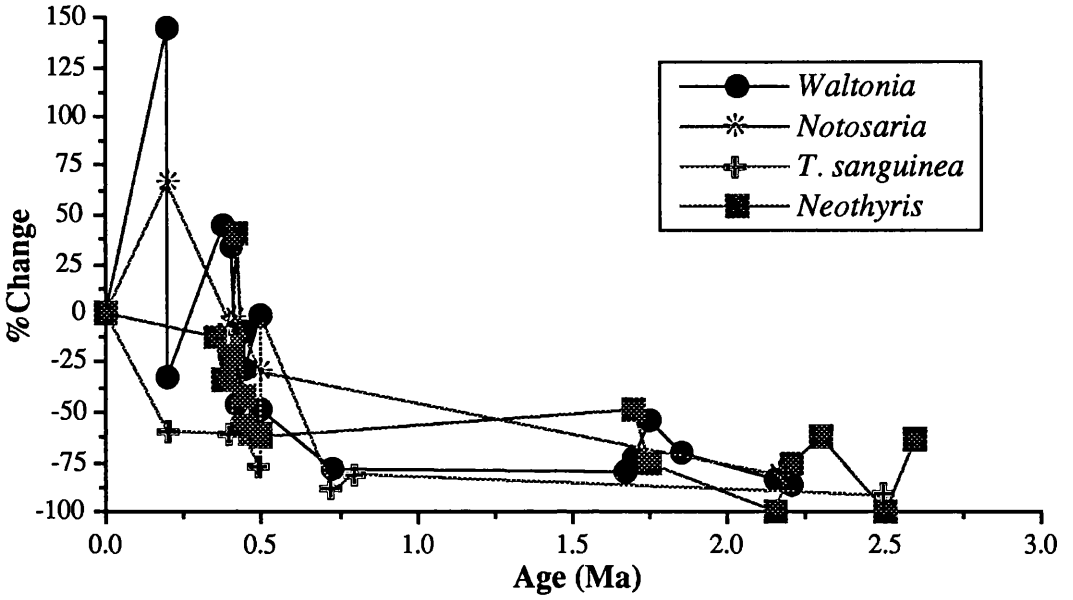


Figure 7.24: %change in concentration of Lys compared with time.

and 60% at 0.5 Ma. The proportion shows a further increase in some cases, up to 80% (**figure 7.25**). With the exception of Glu, the proportion of Lys present in the free state is much lower than that found for any of the other amino acids. It is evident, therefore, that a comparatively large proportion of Lys remains bound to other molecules within the fossil record.

(xv) Summary

Proteins present in the shell during the life of the fossil samples undergo very rapid natural hydrolysis, so that by 0.5 Ma there is generally greater than 80% of any

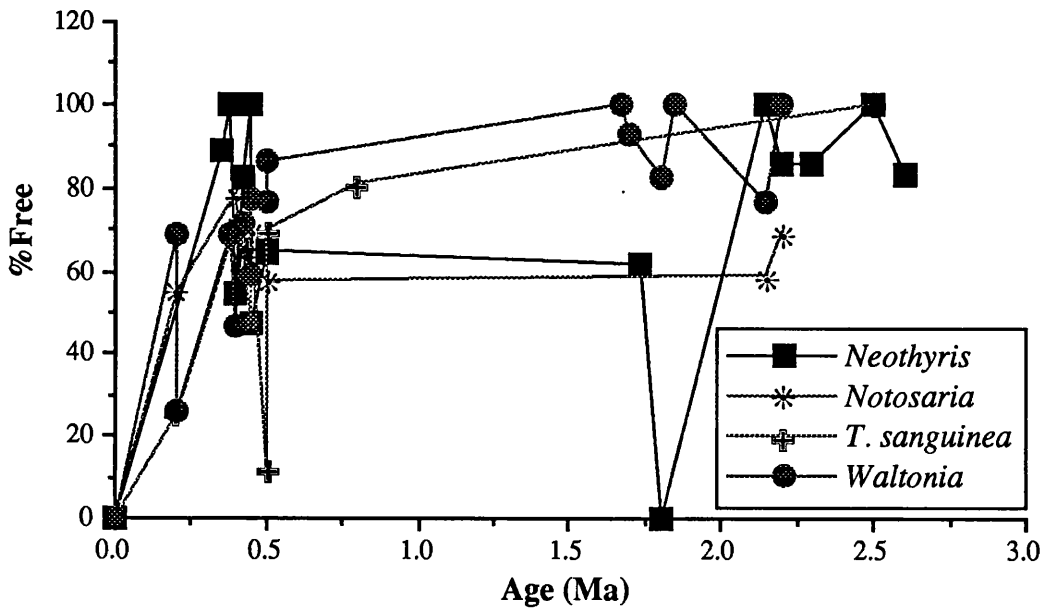


Figure 7.25: Proportion of Lys present in the free state over time, showing a slower rate of increase in the proportion of the molecules present in the free state.

particular amino acid present in the free state. The concentration of individual amino acids in the samples also shows a dramatic decrease over the 2.5 Ma period of time sampled in this study, although in the majority of cases there is a slight initial increase in this concentration. In most cases, the pattern of decrease of the amino acids may be described by either an exponential or logarithmic curve.

7.5 Discussion

7.5.1 Natural hydrolysis of peptide bonds

The equilibrium between the condensation and hydrolysis reactions of a peptide bond is pushed a long way to the left (**figure 1.11**), as its decomposition is thermodynamically downhill, and will therefore be favoured in any natural system, causing hydrolysis of peptide bound compounds. Previous studies have shown that the rate of this natural hydrolysis depends on temperature, water and the nature of the residues on either side of the bond (**section 6.2**).

The proportion of free amino acids, in the samples studied, rapidly increases from almost nothing in Recent samples, indicating that the amino acids are all bound into proteins, to c. 80% in less than 0.5 Ma. This figure does not entirely contradict that of Abelson (1955) who determined that 1-5% of peptide bonds would be broken in 0.1 Ma, although from the figures determined in this study, this would appear to be a rather low figure. The high levels of free amino acid in this study may be due, in part, to the sites of molecular preservation; Abelson (1955) utilised both inter- and intracrystalline molecules for his estimate, and acknowledged that there would be a loss of soluble peptides from the shell due to leaching. Collins (1986) showed that intercrystalline molecules decay in under a year, perhaps due to diffusion between the shell crystallites, and this measurement confirms the potential for loss hypothesised by Abelson (1955). It is likely, therefore, that the low estimate of peptide bond breakage would be due to the loss of some small peptides and free amino acids from the intercrystalline sites within the shells, giving a false impression of the state of preservation; the residual protein would contain a relatively high proportion of peptide bonds.

Hydrolysis therefore occurs naturally at high rates within the shell. For any reaction to take place, however, it is necessary for there to be an input of water. Inclusions of water have been observed within the shell of molluscs (Hudson, 1967), and although no search was made for them in this study, it is possible that similar structures would be present within the shells of articulate brachiopods, as a relict of biomineralisation. Comparison of the proportion of free amino acids contained in the shells of molluscs from the same horizon as the brachiopods are similar, suggesting some relationship between the natural hydrolysis reactions of the two. There is no evidence, as yet, as to the way in which the protein is contained within the shell, nor its conformation; it is possible that they are

contained in aqueous solution, in which case a source of water for natural hydrolysis would not be a problem. However, if the protein is discrete from any large included source of water as suggested by Towe and Thompson (1972), it will be better preserved; and some other source of water is necessary to both initiate and continue the degradative process. It is unlikely that the protein will be totally divorced from such a source, but if it is, water would be available from the decomposition of amino acids and other molecules trapped within the shell. Decomposition of some amino acids yields water, and these reactions are considered further below.

As the fossil proteins are in a highly degraded state, it is reasonable to assume that water has been present in relatively large quantities within the shell, and that natural hydrolysis has proceeded largely unhindered. Samples of different species show different rates of hydrolysis. *Notosaria* shows the most rapid increase in the proportion of amino acids present in the free state, with almost 97% being free by 0.2 Ma. After a similar time, both *Waltonia* (76.85%) and *T. sanguinea* (58.80%) are much less degraded (no samples of such young *Neothyris* were available). After the initial rise to the maxima, the proportion of free amino acids tends to remain at a level, with very little variation. This shows that the overall rate between the loss of free amino acids, through degradative reactions, and the production of other free amino acids by natural hydrolysis reactions must be similar. If the rate of either were to fluctuate significantly, then the proportion of free amino acids present within the shell would also show a high rate of fluctuation.

Previous studies have shown that decomposition of peptide bonds will depend on the nature of the residues on either side of the bond. Protein sequencing (Cusack *et al.*, 1992), immunology (Endo, 1992) and amino acid analysis (**Chapter Five**) have all shown that the amino acid composition of the species under study are different. This will cause species level variation in the rate of natural hydrolysis between the samples as a product of the nature of the residues on either side of each peptide bond. Van Kleef *et al.* (1975) showed that bonds between Ser-Ala, Ser-Ser, Asp-Ala, Asn-Glu and Thr-Ala were particularly susceptible to natural hydrolysis. However, no quantitative analysis has been carried out to determine whether or not it would be possible to predict hydrolysis time for all possible pairs of amino acid within a protein. As the entire primary sequence of intracrystalline proteins from brachiopods are not known, it is not possible to predict the effect which different residues will have on rates of natural hydrolysis.

The data presented here show that some amino acids remain peptide bound even in the oldest of the samples analysed. There is no evidence regarding the size of these peptides, nor their primary sequence. Consequently, there is no direct evidence from this data to enable statements to be made regarding the effect on peptide bonds of the residues which surround them. Van Kleef *et al.* (1975) show evidence of a stepwise degradation from the *C*-terminus, which enables these shortened peptides to be purified. As it was not possible to separate or purify fossil peptides of a size greater than 10 kDa (**section 6.5.3**)

to complete a similar study, it is reasonable to suggest that the degradation of proteins by natural hydrolysis has occurred in many places along the chain of the peptide, releasing greater than 80% of the amino acids into the free state in a short time, and that any remaining peptides will be of short and variable chain length and chemical reactivity.

The natural hydrolysis of proteins has important repercussions for the study of the immune reaction of fossil organic matter. Immunological determinants have been recorded from samples as old as 70 Ma (de Jong *et al.*, 1974). An immunological investigation of Recent and fossil New Zealand brachiopods was completed by Collins *et al.* (1991b), who concluded that it was possible to record the presence of determinants in the fossil record, but at a much reduced level of reactivity (between 1 and 0.1% of the signal obtained for the Recent sample). Natural hydrolysis of the proteins and peptides will cause this loss of reactivity by the destruction of the determinants. If natural hydrolysis is essentially a random process, then determinants will also be destroyed in a random fashion. However, the conclusions of Collins *et al.* (1991b), that there are some preserved determinants provides further evidence of the preservation of short peptides in the fossil record. Simple calculations show that if 80% of the amino acids are present in the free state, then it is possible that some of the peptides which remain in the sample will correspond to the determinants of the molecule. The reduction in immune response of 99-99.9%, shows that only a small proportion of peptides need to be preserved in the samples in order for immunological investigations to prove useful in taxonomic studies in the fossil record.

Collins *et al.* (1991b) suggest that antibodies may be produced against carbohydrates present on a glycosylated protein. If this is confirmed by further work, then it would be likely that a loss of response may also be caused by the destruction of carbohydrates, as well as by the destruction of the peptides.

7.5.2 Reactions of individual amino acids

(i) Aspartic acid and asparagine

Asparagine undergoes rapid and irreversible deamidation to produce aspartic acid and ammonia. The reaction has a half life, at 37°C and neutral pH, of days to years within peptides (Robinson and Rudd, 1974), and consequently it is unlikely that asparagine would persist in the fossil record. Any asparagine which does remain would be converted to aspartic acid during analysis on the 420H amino acid analyser, and hence would not be quantified. No distinction is therefore made between asparagine and aspartic acid, and decomposition reactions will refer to just aspartic acid (Asp).

Aspartic acid undergoes decomposition which may be described by an exponential curve, indicating that the molecule undergoes rapid breakdown with over 80% of the molecule being destroyed in short periods of time. There are two main pathways by which Asp may decompose:

(a) By slow, reversible deamidation to produce fumaric acid and ammonia (Bada,

1971), which has a half life, at 25°C and pH 7, of c.100 ka and 28 Ma at pH 7, 0°C (Bada and Miller, 1968). If the ammonia remains trapped within the shell (i.e. *in situ* decay), an equilibrium will be set up and a mixture of fumaric acid and Asp produced, leading to the persistence of Asp in older samples. This reaction cannot take place when Asp is peptide bound (Bada and Man, 1980).

(b) By decarboxylation of the α- or β-carbons to form the non-protein amino acid β-alanine, or the proteinogenic amino acid alanine. In a series of reactions between pH 1 and 13 and 60-135°C, less than 0.2% decarboxylation to these products was observed indicating that this pathway is unlikely in aqueous solution of the pure compound at these elevated temperatures (Bada and Miller, 1969; 1970).

Evidence for either of these reactions is difficult to obtain, as there may be many sources and sinks for the ammonia produced by deamidation, and fumaric acid cannot be identified on the analysis system. Ammonia reacts with PITC to give the systems peak PTU (see **figure 3.1**). The size of this peak rises in the fossil record, indicating that the concentration of ammonia increases through time. However, this peak height is highly variable and cannot be accurately measured. The rise in the peak height between Recent and fossil samples provides evidence for the existence of ammonia of some source within the shells of fossils.

β-alanine, being an amino acid, reacts with PITC and may therefore be derivatized and separated by hplc in the same way as other amino acids. Analysis of commercially obtained samples of β-alanine produces a peak with an elution time identical to that of the proteinogenic amino acid threonine. In order to assess the likelihood of the α-decarboxylation of Asp (to produce β-alanine) examination of samples with high concentrations of Asp is necessary. *Notosaria* contains by far the highest concentration of Asp in this study, 214.5 ng/mg, and hence it would be expected that if any α-decarboxylation of Asp took place, there would be a rise in the concentration of "Thr" (which in reality would be due to β-alanine). In *Notosaria*, the concentration of Thr shows a very large increase (up to greater than 1000%) over the period 0.2-0.5 Ma, from 5.89 ng/mg to 69.32 ng/mg. Over a similar time period, the concentration of Thr in *Waltonia* decreases from 3.74 ng/mg to 0.70 ng/mg, a loss of 81.28%. These figures show that some diagenetic reaction product of the intracrystalline molecules in *Notosaria* co-elutes with Thr; the only likely candidate for this is β-alanine, thus showing that the α-decarboxylation reaction can take place in geological samples.

When younger samples of the two species are examined (from the Rapanui Marine Sand, 0.2 Ma) Thr loss is -33.45% for *Notosaria* and -25.40% for *Waltonia*. What is evident from this is that the coelution product is not produced in younger samples, production of β-alanine must therefore occur in the period 0.2-0.5 Ma, and this is

confirmed by the rise over time of the concentration of "Thr" (**figure 7.13**). The proportion of "Thr" that is present in the free state rises rapidly to c. 100% and maintains this level, showing that the breakdown product of Asp is present, and was formed, in the free state rather than whilst Asp was peptide bound, again showing that α-decarboxylation is possible (as this process will break any remaining peptide bonds).

When the %change of Asp and Thr are compared for *Notosaria*, there is no direct correlation (**figure 7.26**), indicating that not all of the Asp decays by α-decarboxylation, and that some must decay by other methods. It would appear likely that the majority of Asp must decay via deamidation; the wealth of literature evidence from pyrolysis dictates this fact. β-alanine is also not a very stable molecule as the "Thr" concentration rapidly drops after this increase (**figure 7.26**). As other samples do not have a corresponding rise

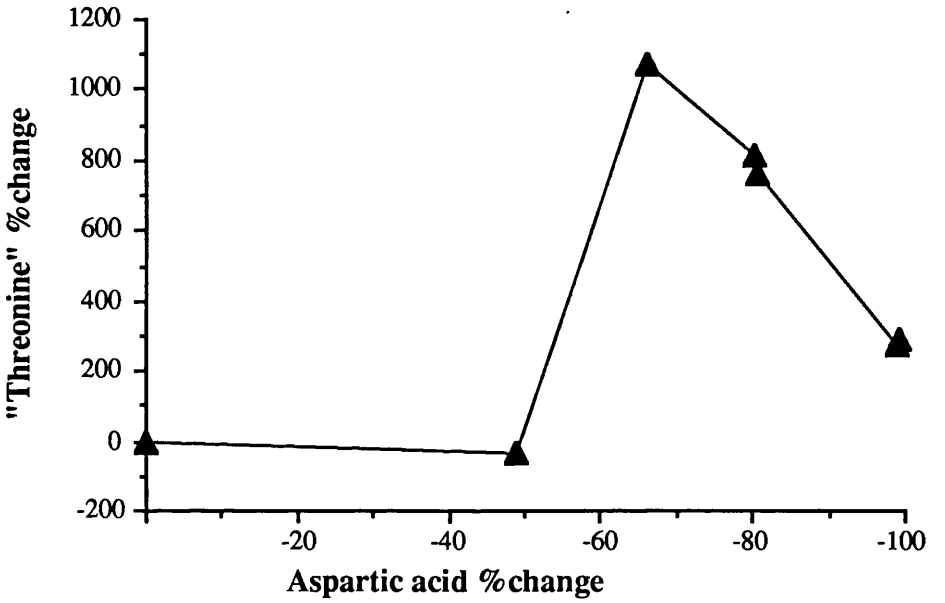


Figure 7.26: Change in the concentration of Asp v the concentration of "Thr" (β-alanine) in *Notosaria*. The increase in the concentration of "Thr" does not occur immediately on fossilisation of the samples. The formation of this molecule takes place after the majority of the peptide bonds have been degraded. The degradation of Asp to β-alanine accounts for a maximum of 45% of total decomposition. The molecules of β-alanine also decay very rapidly.

in "Thr", then only a proportion of molecules need to decay via α-decarboxylation. As other fossil samples do not have as great an initial concentration of Asp, then the same amount of α-decarboxylation will not yield such a high concentration of "Thr", and thus an increase will not necessarily be observed on the chromatogram. This may be measured in the following way. By 0.5 Ma, Asp concentration has dropped by 142.40 ng/mg, and "Thr" concentration has risen by 63.43 ng/mg. At maximum, therefore, only c. 45% of Asp decomposition can be explained by the increase in "Thr", representing the formation of β-alanine (this is likely to be a maximum value, as some of the "Thr" increase will be due

to the remaining original Thr present in the sample).

Previous studies have indicated that, at least in pyrolysis of pure Asp, the α-decarboxylation reaction is rare (Bada, 1971). However, these estimates were based on reactions at elevated temperatures of the pure compound alone, and in this case may not be directly applicable to fossil molecules occurring in mixtures. If there is some α-decarboxylation, as this data suggests, then it is possible that there will also be some β-decarboxylation, resulting in the formation of some α-alanine, i.e. "normal" alanine, which will explain some of the increase in Ala concentration, although high temperature experiments indicate that this is unlikely to occur (Bada and Miller, 1970). Both reactions proceed rapidly in micro-organisms with the aid of enzymatic catalysis (Meisler, 1965), and reactions which are catalysed by enzymes can also take place without the catalysis, but at a very much reduced rate (enzymatic catalysis usually increases the rate of reaction about a million fold). This reduced reaction rate will act over the extended time period available in the geological record.

(ii) Glutamic acid and glutamine

Glutamine, like asparagine, undergoes rapid irreversible deamidation yielding Glu and ammonia, with a similar half life (days to years in peptides at 37°C, pH 7 (Robinson and Rudd, 1974)). Any Gln not undergoing this reaction would be converted to Glu by the analytical procedure, and hence would be indistinguishable from the original Glu.

Glutamic acid is one of the most stable of the proteinogenic amino acids (Vallentyne, 1964), although it may undergo degradation by two reaction pathways.

(a) Lactam formation to produce pyroglutamic acid.

(b) γ-decarboxylation to produce the non-protein amino acid γ-aminobutyric acid.

The kinetics of the formation, in water, of pyroglutamic acid from Glu was studied in detail by Wilson and Cannon (1937). The formation of the lactam (**figure 7.27**) results in the expulsion of water from the structure of Glu, which is then available for natural hydrolysis reactions. Pyroglutamic acid may be converted back to Glu through

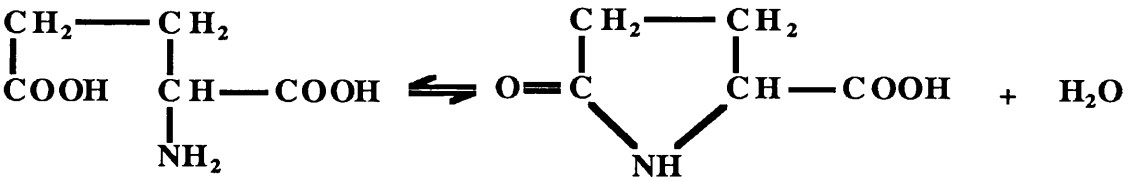


Figure 7.27: Formation of pyroglutamic acid, the lactam of Glu, with the loss of a molecule of water. The formation of the lactam acts to stabilise the molecule over geological time, but also prevents the identification of Glu in samples which have not been hydrolysed (after Vallentyne, 1964).

normal protein hydrolysis, which was demonstrated by the commercial purchase and analysis of pyroglutamic acid. Lactam formation may only take place when Glu is in the free state, hence the protein must have undergone natural hydrolysis prior to lactamisation. In fossil samples, the proportion of Glu present in the free state (measured by derivatization only) is very much lower than would be expected; on average, only c. 30% of the amino acid is free, compared to greater than 80% for other amino acids of similar age. The concentration of Glu in hydrolysed samples is nearer to that predicted by the preservation of other samples. Evidently, one of two possibilities can exist: either Glu remains bound in peptides far longer and those peptides have a -Glu-Glu-Glu- repeating sequence), or lactam formation readily occurs, rendering the free, original Glu (which is now pyroglutamic acid) unrecognisable to the analysis system, until the sample is hydrolysed. From the evidence of the %free data and the analysis of commercial samples, it is likely that lactam formation takes place converting Glu to pyroglutamic acid, and thus acting to preserve this molecule somewhat in the fossil record.

The concentration of Glu decreases in an exponential manner over the 2.5 Ma time period under study. In all samples, the concentration of Glu drops to only c. 20% of that found in the Recent. The γ-decarboxylation of non-lactamised Glu is a possible decay pathway, resulting in the formation of the non-protein amino acid γ-aminobutyric acid. As this compound is also an amino acid, it may be detected by the analysis system (**section 3.7**). Examination of the data generated for fossil samples indicates the presence of a peak in all fossil samples in the correct position for eluted γ-aminobutyric acid. However, this amino acid was not quantified due to presence of a similar peak in all Recent samples, casting some doubt as to whether the peak was truly γ-aminobutyric acid caused by the decomposition of Glu, or was another, as yet unidentified, compound. There is also no direct correlation between the age of the sample, the decrease in concentration of Glu and the increase in the size of the peak at the position of γ-aminobutyric acid.

The data presented here indicates that Glu readily undergoes lactam formation to produce pyroglutamic acid, which cannot be identified unless the sample is hydrolysed, and which acts to preserve the molecule to some degree, although not extensively. This reaction, however, does not explain the decrease in the concentration of Glu, which either occurs via decarboxylation to form γ-aminobutyric acid, or via the decomposition of pyroglutamic acid. There is no direct evidence of a relationship between the decrease of Glu and a corresponding increase of the γ-aminobutyric acid, but this may be due, in turn, to the rapid destruction of this reaction product.

(iii) Serine

The peptide bonds on either side of Ser residues undergo rapid natural hydrolysis to release amino acids into the free state. Over time, the proportion of Ser in this state declines, indicating that the majority of the decomposition of Ser takes place when that molecule is free, rather than when combined with other amino acids in peptides.

Some Ser residues have been reported to decay whilst remaining bound within the peptide (Akiyama, 1980), although as natural hydrolysis takes place so rapidly and the free molecules undergo such rapid degradation when compared to those which are peptide bound, it is likely that the majority of the intracrystalline molecules present in brachiopods decay whilst in the free state, and that decomposition whilst peptide bound is not significant. Overall, Ser shows a decrease, which may be described by an exponential curve.

Pyrolysis experiments have demonstrated that thermal decomposition of 63% of Ser (the so-called 0.37-life of Vallentyne (1964)) would only take 0.1 Ma at 10°C. This is consistent with the data presented here; where a slightly lower percentage loss of Ser over this time is observed. Pyrolysis products from the destruction of Ser were found to be Ala, Gly and ethanolamine, with the latter not being detected by the analysis system. Ala and Gly, however, are common amino acids in fossil samples, and the diagenetic production of these amino acids, by the degradation of Ser will lead to increasing concentrations of these amino acids in fossil samples, distorting any information gained from these amino acids.

Fossil molecules of Ser were examined in foraminiferal tests by Bada *et al.* (1978), based on the results of Vallentyne (1964), and it was concluded that decomposition followed three possible pathways (**figure 7.28**).

(a) Dehydration of the hydroxyl group in the side chain of Ser resulting in the release of a molecule of water (which may be utilised in natural hydrolysis) and the formation of intermediates prior to the formation of Ala.

(b) Aldol cleavage resulting in the formation of Gly and formaldehyde.

(c) Decarboxylation resulting in the formation of ethanolamine.

Bada *et al.* (1978) demonstrated that dehydration was the prevalent decomposition reaction of Ser, and that aldol cleavage and decarboxylation were not so important in fossils. Neither formaldehyde nor ethanolamine could be recognised by the analysis system used in the present study, hence the positive identification of two of these reactions cannot be made. The increase in concentration of Ala in fossil samples (**section 7.4.2**) is suggestive of the occurrence of dehydration reactions. Comparison of the %change and concentration of Ser and Ala show no direct correlation, so although there is an increase in the concentration of Ala, this cannot, therefore, be explained solely by the decomposition of Ser.

Bada *et al.* (1978) also demonstrated that the decomposition of Ser by dehydration followed first order reaction kinetics. By comparison with the concentration of Leu, the results could be described by a straight line on log/normal graphs (i.e. the destruction is exponential) up to 0.6 Ma. The data presented here however, shows that, although this relationship holds true for samples up to 0.6 Ma (**figure 7.29a**), older samples deviate significantly from this pathway (**figure 7.29b**), reflecting the recorded lowering of the rate of decomposition, and giving secondary evidence for the two component curve

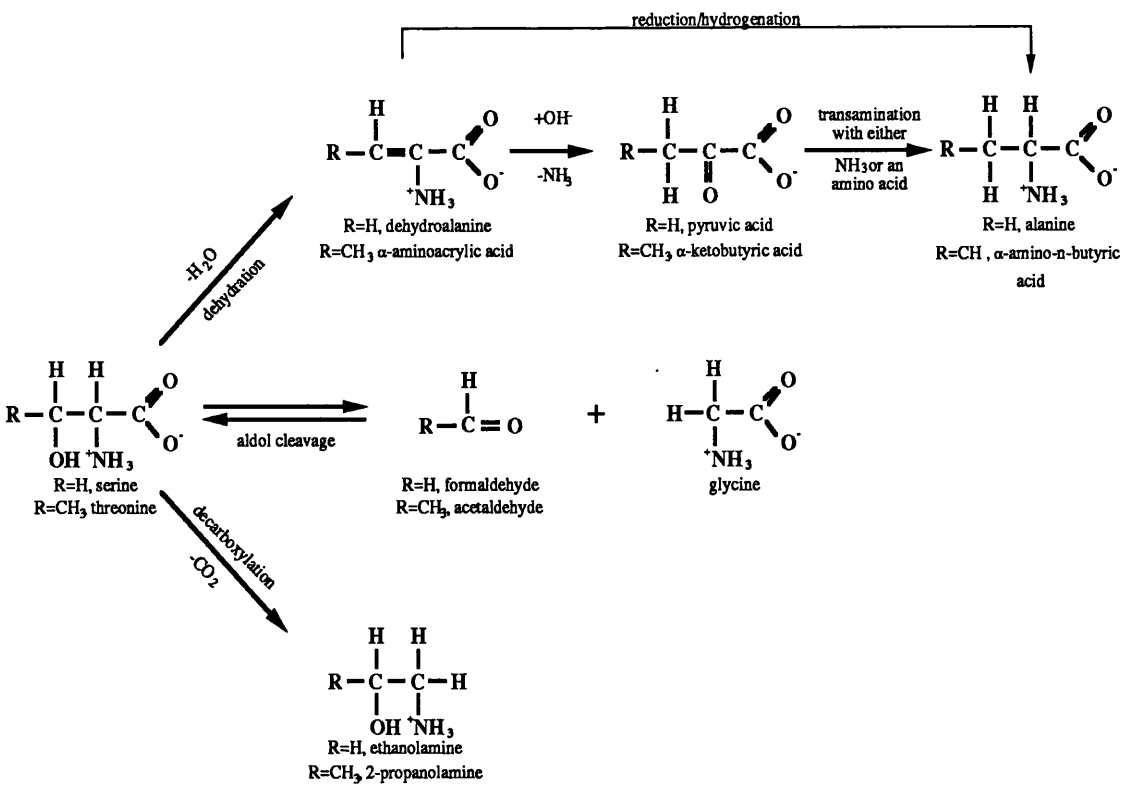


Figure 7.28: Possible decomposition pathways for the hydroxy amino acids Ser and Thr (after Bada *et al.*, 1978).

system (i.e. rapid hydrolysis of susceptible bonds and reaction of free amino acid compounds).

The proportion of Ser present in the free state both rises and drops rapidly, from 0% in the Recent, to c. 90% at 0.5 Ma, to 0% by c. 0.7 Ma. This shows rapid destruction of the amino acid whilst it is in the uncombined state. Older samples still contain some preserved Ser, although this is present in the combined state, and other amino acids to which Ser is bound must act to stabilise the molecule. Bada and Man (1980) state that Ser, when in the free state, decomposes predominantly by dehydration reactions, whereas aldol cleavage is dominant when the molecule remains bound. It is likely, therefore, that decomposition of Ser from these intracrystalline molecules will follow dehydration reactions when the amino acid is present in the free state. Decomposition by reversible aldol cleavage is negligible, as there is only a small rise in the corresponding concentration of Gly. It can thus be expected that most of the Ser in the shell has decomposed to Ala.

The Ser dehydration reaction proceeds via several intermediate compounds (**figure 7.28**). Formation of dehydroalanine and pyruvic acid from Ser follow standard reaction pathways. However, the final stage of reaction requires transamination of the pyruvic acid intermediate, by ammonia or the amino group of another amino acid. Several amino acids are known to contribute to this process (see review in Tan-Wilson, 1983), including glutamate, Val and the non-protein amino acid β-alanine, formed during the

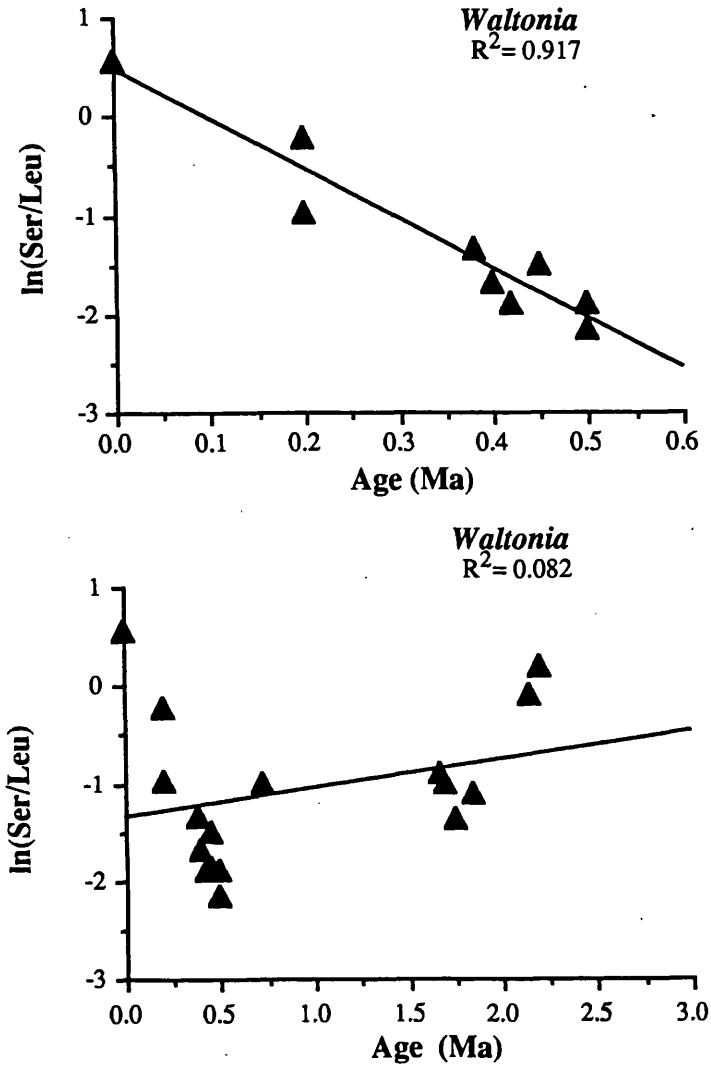


Figure 7.29: (a) Graph of the natural logarithm of Ser/Leu against time (up to 0.6 Ma) showing an almost linear relationship of decay (c.f. Bada *et al.*, 1978). (b) As in (a) but with older data included, showing that the linear relationships does not hold true for older samples, probably due to the peptide bound preservation of Ser or the degradation of the Leu molecules over time.

decomposition of Asp (Tan-Wilson, 1983). There is no corresponding decrease in the concentration of either Glu or Val, indicating that another compound is necessary for the reaction to take place. It is likely that this will either be ammonia, which may be produced by the deamidation of amino acids, or β-alanine, produced by the α-decarboxylation of Asp.

(iv) Glycine

Glycine is one of the most thermally stable amino acids (Abelson, 1954), but still shows considerable degradation over the 2.5 Ma of this study, resulting in a highly variable percentage loss for the four species under study (**figure 7.10**). Pyrolysis experiments have shown that the reaction product of Gly is methylamine (Vallentyne, 1964), indicating that the major pathway for Gly decay is via decarboxylation to produce this product and CO_2. Neither of these products were positively identified on the analysis system, although methylamine may account for one of the unidentified peaks on the chromatogram.

Gly is the most abundant of the amino acids found in the intracrystalline extracts from brachiopods, accounting for up to 50% of the total. Initially, there was some concern that this was due to coelution between Gly and some other compound, giving an over estimate of the concentration of Gly. This may be discounted for two reasons:

(i) Quantification of free amino acids from Recent samples shows negligible Gly concentration. When these samples are hydrolysed, high concentrations of Gly are released, indicating that the molecule is peptide bound.

(ii) There is no known common amide which reacts with PITC to coelute with the Gly peak (P. Jackson, Applied Biosystems, pers. comm.).

The proportion of Gly molecules present in the free state rapidly increases to almost 100%, and the decrease in the concentration of Gly occurs whilst the molecules are present in the free state. As the decomposition reaction is decarboxylation, only molecules which are present in this state are decomposed (as the carboxyl group lost forms part of the peptide bond (**figure 1.11**)).

In most cases, there is an initial rise in the concentration of Gly, indicating that Gly may be a decomposition product of another amino acid present within the shell. Examination of the table of pyrolysis products (**table 7.1**), produced by Vallentyne (1964), shows that this is the case, with decomposition of Val, Ser, Thr, Met and Tyr all possibly resulting in the formation of Gly. These reactions are considered in the relevant sections. Where Gly is the decomposition product, it should also be remembered that the "new" Gly will decompose in the same way as the original Gly, and hence the overall concentration of the amino acid will decrease once production from other sources has ceased. Once again, the diagenetic production of this amino acid will distort information gained from the molecules.

(v) Arginine

Arginine shows the most rapid decomposition of amino acids considered in this study, following an exponential or logarithmic curve and proceeding to complete destruction in less than 1 Ma (with the exception of two results of older samples of *Waltonia* which may be erroneous. Pyrolysis experiments (Vallentyne, 1968) indicate the rate of Arg decay, and extrapolated results (which will have some error as the data was presented only as graphs) indicate that the 0.37-life (the time required for 63% decomposition) for Arg is c.100 years at 20°C. At lower temperatures, which will probably exist in the sediment, this figure will rise, but even at 0°C only to c. 1300 years. This is the lowest 0.37-life value of any of the amino acids studied. The proportion of free amino acids shows a rapid rise up to 100%, and then a decrease as the molecules are decomposed.

Arg decomposes in a reaction which yields urea and the non-protein amino acid ornithine (Murray *et al.*, 1965), a reaction which has already been identified from fossil samples (Hare and Mitterer, 1967). Ornithine is an amino acid which may be detected by the amino acid analyser, and which has an elution time between that of Phe and

Lys (**table 7.2**). The height of this peak is variable, but it does show a relationship between the decrease in concentration of Arg and the increase in ornithine, based on peak height alone, as ornithine was not quantified in the course of the experiments (**figure 7.30**). Low concentrations of ornithine are also identified from Recent samples. The concentration of ornithine rises rapidly and reflects the decrease in Arg in mirror fashion, but is also relatively unstable itself, and decays rapidly (**figure 7.30**), via decarboxylation to form putrescine and CO_2 (Murray *et al.*, 1965). The other decomposition product of Arg, urea, could not be identified on the amino acid analyser.

Amino acid	Elution position	Comments
β-alanine	coelutes with Thr	
ornithine	immediately after Phe	
α-aminobutyric acid	immediately before aniline	
γ-aminobutyric acid	immediately after PTU	
hydroxyproline	coelutes with Glu	
norvaline	coelutes with Met	
citrulline	immediately after His	
α-aminocaproic acid	immediately after Leu	
ε-aminocaproic acid	immediately before Ile	
norleucine	immediately after Leu	
glucosamine	coelutes with Ser	
galactosamine	between Glu and Ser	
pyroglutamic acid	coelutes with glutamic acid	only after hydrolysis

Table 7.2: Retention positions of some non-standard amino acids analysed during the present study. Actual retention times are not given, as these times vary to some degree for all amino acids. Repeated analyses with commercially obtained standards are necessary to confirm retention times.

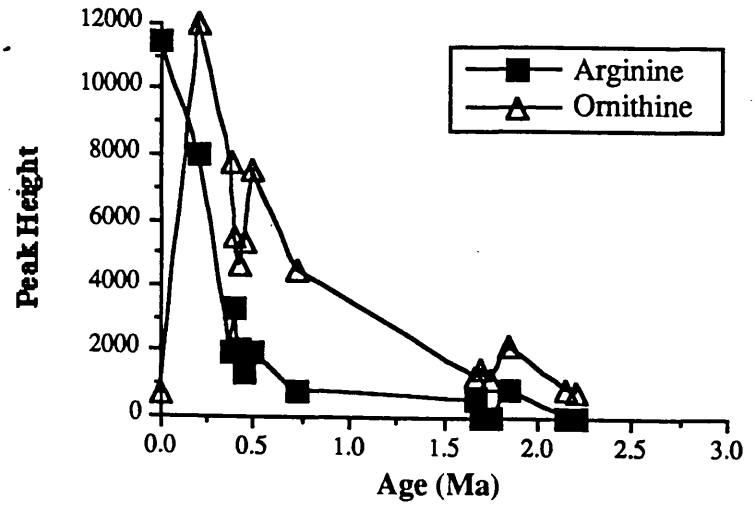


Figure 7.30: Graph of the peak heights of Ornithine and Arg, showing corresponding increases and decreases in the concentration of the molecules. Ornithine is also unstable over time, shown by the decrease in the peak height of Ornithine in older samples.

It should also be noted that it is possible to produce ornithine during the acid hydrolysis of peptides and proteins containing Arg, although at the temperatures and time of hyrolysis used in this study, the work of Murray *et al.* (1965) show that only a negligible percentage change from Arg to ornithine would occur. This could, however, account for some of the ornithine identified in Recent samples.

In the pyrolysis experiments (Vallentyne, 1964; 1968) Arg was shown to decompose to ornithine, ammonia, proline and an unidentified compound which is associated both with ornithine and with Arg. Meisler (1965) indicates that citrulline is an intermediate compound between ornithine and Arg, and it is possible that this compound may correspond to the unidentified compound. However, no citrulline was detected in this study although commercial standards were analysed to determine the elution time (**table 7.2**). The possible origin of the Pro found by Vallentyne (1964) is not certain.

(vi) Threonine

The decay of Thr in the fossil record is a rapid process, with greater than 80% loss of the amino acid occurring in less than 1 Ma. Possible decay pathways are as follows, and are summarised in **figure 7.28.**

(a) Aldol cleavage, resulting in the formation of Gly and acetaldehyde.

(b) Dehydration of the side chain, via the loss of the hydroxyl group, resulting in the release of a molecule of water and the formation of intermediates, prior to the formation of the non-protein amino acid α-aminobutyric acid.

(c) Decarboxylation, resulting in the formation of propanolamine and CO_2.

Pyrolysis experiments to determine the rate of decay of Thr (Vallentyne, 1964) only observed the production of Gly from the decomposition of Thr, with the reaction following first order kinetics. The 0.37-life (63% decomposition) for Thr at 10°C is in the region of 30 Ma (inferred from graphical data), very much longer than is shown by the data presented here (where the corresponding 0.37-life would be 0.5 Ma, in the Kupe Fm). Earlier studies, quoted by Vallentyne (1964), have indicated that the aldol cleavage of Thr occurs more rapidly when the reactions are catalysed by metal ions, or when the Thr molecule is present in peptides. As the pyrolysis experiments of Vallentyne (1964) were carried out on standard solutions of pure amino acids, the peptide effect could not be observed, possibly resulting in the higher estimation of the 0.37-life of the amino acid than in the peptide bound samples.

The decomposition rate of Thr was also analysed by Bada *et al.* (1978), who discovered that the rate of decay in foraminiferal tests would result in complete destruction of the amino acid in around 1 Ma, similar to the results presented here. This work also examined the stereochemistry of α-aminobutyric acid (α-ABA), and concluded it all was formed by the decomposition of Thr. α-ABA has been identified in the fossil species examined in this study, but not quantified, as there was no obvious correlation between the concentration of α-ABA and Thr. However, the presence of α-ABA indicates that some

of the Thr must decompose by the dehydration reaction. This reaction will release water, which may proceed to participate in natural hydrolysis reactions. Calculations by Bada *et al.* (1978) indicate that dehydration accounts for only c. 10% of the decomposition of Thr, and hence a direct relationship between this and α-ABA would not be expected.

A second possible decomposition pathway for Thr is via aldol cleavage, resulting in the formation of Gly and acetaldehyde. The latter compound could not be identified by the analysis system. Gly, however, is one of the most common of the amino acids found in fossil samples. There is an increase in Gly concentration over the first 0.5 Ma of the study period, the time when Thr is decomposing, although there is no direct relationship between the two.

The third possible pathway for the decomposition of Thr is decarboxylation, which would yield propanolamine and CO_2. Neither of these products were identified by the analysis system used in this study, so it was not possible to assess the potential for this decay pathway.

From both the data presented here, and from previous studies, it is probable that the decay of Thr in the fossil record takes place via a combination of all three potential pathways. Dehydration reactions do not account for a large proportion of the decay; the concentration of α-ABA is not great, nor is there a convincing correlation between the concentration of α-ABA and Thr. Some of the molecules decay via aldol cleavage, resulting in the formation of Gly, although it is not surprising that there is no correlation between the concentration of Thr and Gly, as Thr is one of a number of amino acids which decays to Gly. Decarboxylation may be the major decay pathway for Thr, although the reaction products cannot be identified.

(vii) Alanine

Alanine is one of the group of amino acids which are thermally very stable (Abelson, 1954; Vallentyne, 1964). Decay of Ala is via decarboxylation to form ethylamine and CO_2 (Abelson, 1954). Pyrolysis of the amino acid is very slow; it would take 10^{10} years for 63% of the molecule to decay at 20°C (Abelson, 1954), indicating that thermal decomposition may not be important. In the presence of oxygen (the pyrolysis experiments were carried out under nitrogen), however, the rates of this pyrolysis reaction would increase dramatically. At room temperature, oxygen attacks Ala molecules, causing the release of carboxyl carbon with a reaction half life of 20 ka (Conway and Libby, 1958). This indicates that the presence of oxygen is probably of vital importance in the decomposition of these amino compounds.

The data presented here (**figure 7.15**) indicates that the concentration of Ala present in the Recent samples is almost identical to that recovered from samples dated at 2.5 Ma. The intervening period, however, shows major variations in Ala concentration. All samples show a rapid increase in concentration, before it decreases to a level similar to that at which it started. The increase in Ala is due, in part, to its diagenetic formation

from decomposition reactions, including the dehydration of Ser. There is no difference between the original and diagenetic Ala in terms of decomposition pathways, hence some of this diagenetic Ala will also decompose. As there is no direct correlation between the decomposition of Ser and the production of Ala, it is likely that other diagenetic reactions may produce Ala, for example the β-decarboxylation of Asp, or the cleavage at the ring of Phe and Tyr. These reactions must play a part in the increase in concentration of Ala, as this cannot be explained by the decomposition of Ser alone.

The rate of decay of Ala appears, from both this and previous studies (Hare and Mitterer, 1967; 1969), to be much more rapid that that recorded for pyrolysis of a solution of the pure amino acid. This could be due to a number of effects, notably the presence of oxygen, surviving peptide bonds within the shell and the effect of mixtures of amino acids (Vallentyne, 1964).

(viii) Proline

Proline is the most stable of the amino acids tested by pyrolysis at temperatures below 212°C (Vallentyne, 1968). In the earlier pyrolysis experiment (Vallentyne, 1964), Pro did not have an organic ninhydrin positive reaction product. In the later study (Vallentyne, 1968), Pro decomposed to form ammonia (recognised on column chromatography), indicative of decay by deamidation. Calculated 0.37-lives (63% decomposition) at 0°C and 20°C (inferred from graphical data) are 126×10^{12} years and 178×10^{9} years respectively, indicating that no thermal decomposition would be expected at these temperatures. The data presented here shows that up to 80% of the Pro has decomposed by 2.5 Ma, indicating that there is an increased rate of decomposition in fossil samples, which may not be assessed by the use of pyrolysis of the pure compound. This is indicative of the influence of other factors, such as the inorganic component of the shell or the effect of mixtures of amino acids, on the rate of decomposition.

(ix) Tyrosine

After an initial increase in the concentration of Tyr, there is rapid degradation of the molecules over time. Pyrolysis of this amino acid has not been studied in detail, although in a preliminary study Vallentyne (1964) discovered that Gly was formed in small quantities. The amino acid undergoes almost complete destruction in 2.5 Ma, but it is not apparent by which reaction pathway. The structure of the Tyr side chain (**figure 1.5**) includes an unsaturated ring structure (containing a hydroxyl group). In order for Gly to be formed during decomposition of Tyr, the structure would undergo cleavage of the C-C bond in mid chain location away from the ring (**figure 7.31**). Although mid chain breakage is more likely, it is possible for the side chain to break at the junction between the ring and the remainder of the molecule, rather than in mid chain. This pathway has not been explored in previous studies. Either of these methods would appear to be likely, as the bond energies to be overcome are the same (**table 7.3**) although mid chain cleavage is sterically favoured.

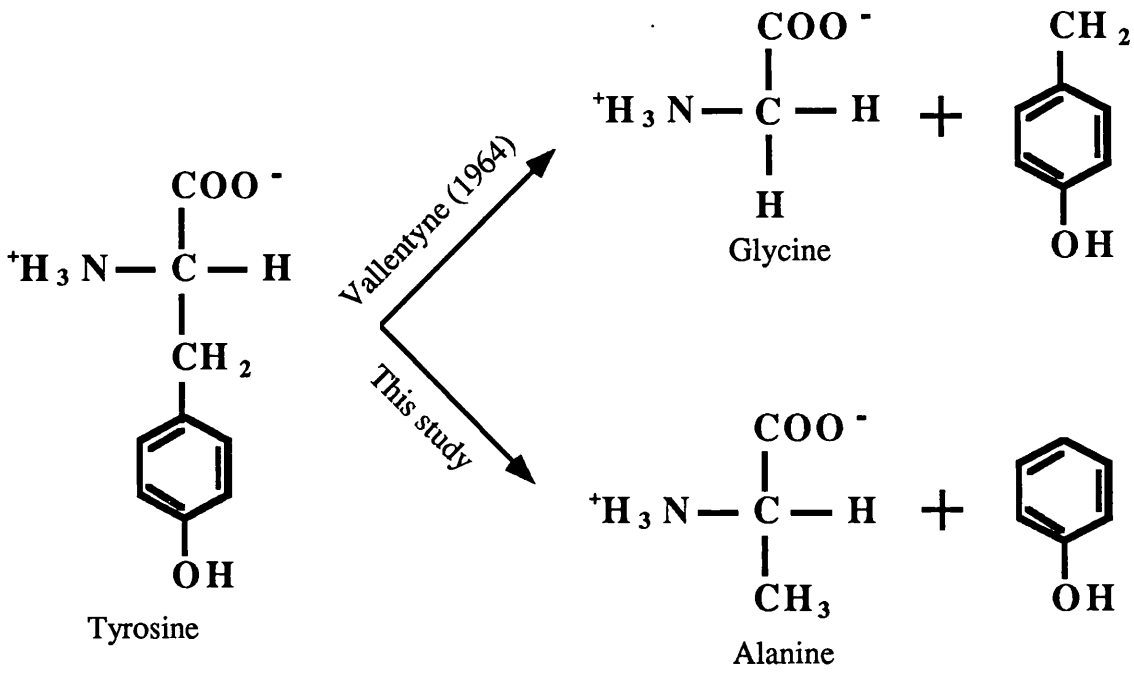


Figure 7.31: Possible pathways for the decomposition of Tyr, proposing a possible mechanism for the increase in concentration of Ala.

When the molecule begins to decompose, the ring structure will tend to be unaffected, as the C=C double bonds within the ring are stronger than other bonds. It is possible that the hydroxyl group attached to the ring may be released via dehydration reactions, allowing more water molecules to be utilised in the process of natural hydrolysis. If the ring structure were to dissociate totally from the remainder of the amino acid molecule, then the part which contains the α-carbon will form a molecule of Ala.

As the increase in the concentration of Ala cannot be accounted for by the dehydration of Ser alone, other decomposition reactions must also have an influence on the increase in concentration of this amino acid. Tyr, with the potential for cleavage of the ring from the remainder of the molecule must be considered as a candidate for the parent molecule of some of the diagenetic Ala found within the shell. Other reactions, such as decarboxylation and deamidation could proceed in Tyr, although further work is required to determine potential products and to confirm reaction pathways.

(x) Valine

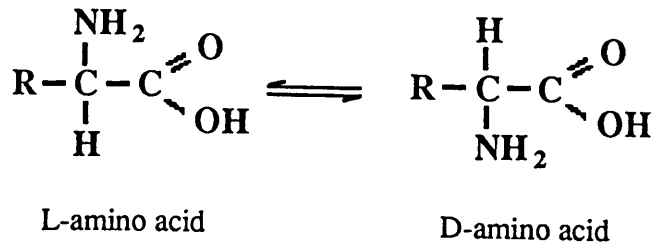
Valine decomposition proceeds to 50-60% destruction over the 2.5 Ma of this study, although pyrolysis experiments do not identify the decomposition pathway. Vallentyne (1964) tentatively identified Gly as the only product of the reaction. Decarboxylation of Val produces 2-methylpropylamine and CO_2 (Meisler, 1965), neither of which were identified by the analysis system. It is likely that the decomposition will take place either by decarboxylation or deamination, although neither could be examined in this study. 0.37-life (63% decomposition) calculations for the decomposition of Val,

based on a regression line derived from the figures of Vallentyne (1964), are 632.69×10^{12} years (at 0°C) and 1.00×10^{12} years (at 20°C). These figures are extremely high when compared to the data presented here, and it is therefore necessary to postulate on some factor which will increase the rate of reaction, such as elevated temperatures, due to the burial of the samples.

(xi) Isoleucine

Ile decomposes readily over the 2.5 Ma of this study until up to 80% of the molecule has decomposed, although it is not certain by which pathway. Vallentyne (1964) found small amounts of ninhydrin positive molecules on pyrolysis of Ile, but these were not identified. Once again, the most likely decay pathways are via decarboxylation and deamination, which produce products not recognised by the analysis system.

The racemisation of the Ile molecule to its stereoisomer, alloisoleucine (aIle), is one of the most documented and intensively studied of the amino acid reactions (for a review, see Schroeder and Bada, 1976). The conversion, over time, of Ile to aIle (**figure 7.32**) will not affect the separation or identification of the individual amino acids; the change around the chiral centre will not affect the reaction of Ile or aIle with PITC, and hence the two molecules will have identical retention times, be eluted from the column at the same time, and be identified as the same molecule. The formation of racemic molecules in amino acids occurs after hydrolysis, during the stage when the molecules are degraded to diketopiperazides (Steinberg and Bada, 1983); hence the majority of the isoleucine degradation is likely to take place after the molecules have been released from the peptide bond. This reaction will also be true for every other amino acid, with the exception of Gly, which does not have an assymmetrical carbon atom, and consequently does not have a structural stereoisomer.



L-amino acid D-amino acid

Figure 7.32: Conversion of Ile to its diastereoisomer aIle. For a review see Schroeder and Bada (1976).

(xii) Leucine

Leu, although recognised in previous studies as one of the most stable of the amino acids (e.g. Abelson, 1955; Bada and Man, 1980), decomposes rapidly in the fossil record until approximately 80% of that found in the Recent samples of the same species has decomposed by 2.5 Ma. Pyrolysis experiments (Vallentyne, 1968) demonstrated that

Leu decomposed to form ammonia (no other ninhydrin positive compound was formed), with extrapolated (from graphical data) 0.37-lives (63% decomposition) of 317.01×10^{12} years (0°C) and 563.89×10^{9} years (20°C). The conversion of Leu to ammonia was only 25-40%, when the decomposition of Leu had reached 85-98%, indicating that deamination was not the sole pathway of decomposition. The formation of ammonia arises from the deamination of the amino acid. The low rates of conversion indicate that other pathways of decomposition are also important, but which cannot yet be identified.

The data presented in this study indentifies a much more rapid rate of decay of Leu than is indicated by the pyrolysis reactions, perhaps suggesting again an increase in the rate of reaction of the molecule. The 0.37-life of Leu, shown by this data is c. 1.8 Ma. Such rapid decomposition of this amino acid (when compared to pyrolysis experiments) has broad implications for previous studies, some of which (e.g. Bada and Man, 1980) have represented the data as being Leu equivalents, or have used the concentration of Leu as a constant on the basis of the stability of Leu in pure aqueous solution (Bada *et al.*, 1978). Leu is very stable when pure solutions of the amino acid are pyrolysed, but this does not appear to be the case when the amino acids are peptide bound, in free mixtures or when associated with an inorganic phase. Comparisons with the results of pyrolysis of pure compounds need to be drawn with caution in these samples, and this is confirmed by the pyrolysis of oyster shell powders (Totten *et al.*, 1972), where Leu is also rapidly decomposed.

(xiii) Phenylalanine

In the 2.5 Ma time period studied, the Phe molecules contained within fossils decompose until approximately 80-90% of the level found in Recent samples are lost. Pyrolysis experiments (Vallentyne, 1964) have shown that the main pathway of decomposition for Phe is via decarboxylation to form phenethylamine, which is then further decomposed to benzylamine. The reaction follows first order kinetics, with a 0.37-life (63% decomposition) of 100 Ma at 10°C, much slower than that found in the present study. The decomposition of Phe in the pyrolysis of mixtures of amino acids is more rapid than the single compound (Vallentyne, 1964), and it is this effect which may cause this rapid decomposition.

Phe, like Tyr, consists of a ring structure at the end of the amino acid molecule. As argued for Tyr, the ring structure is likely to be preserved in the fossil record, as the bond energies of the double bonds are higher than that of the C-C single bonds. Decarboxylation is one of the main pathways of decomposition for Phe but it is also possible that the C-C bond between the ring and the remainder of the molecule will be cleaved, resulting in the formation of alanine, or that there is a mid chain cleavage, resulting in the formation of Gly. This is a third possible pathway for the diagenetic formation of Ala which has not been previously considered. There is no direct evidence in this study to be certain of the pathway of decomposition.

The proportion of Phe that is present in the free state rises rapidly to 80-90%, but then maintains that level. The %change of the molecule decreases at a lower rate than that of the %free, indicating that the decomposition of the molecule is likely to take place whilst Phe is in the free state.

(xiv) Lysine

The concentration of Lys in these fossil samples decreases until c. 90% of that present in the Recent samples has decomposed by 2.5 Ma. Pyrolysis experiments (Vallentyne, 1968) revealed that the decay of Lys follows a pattern of non-linear decay, and hence the data could not be extrapolated further to allow assessment of the likely rates of decomposition at lower temperatures. However, in an earlier study (Vallentyne, 1964), Lys was found to be one of the least stable of the amino acids. This is not unexpected, due to the presence of a secondary amino group, which is easily degraded by deamidation. The bond energies for the decomposition of C-N bonds are somewhat lower than those for C-C (**table 7.3**). Column chromatography detected the presence of ammonia in the pyrolysed solutions, indicating that either primary or secondary deamination is occurring.

Average bond energies for common bonds in amino acids

Modified from Templeton (1978)

Bond	Average bond energy (kJ per mole)
C-H	415
C-C	345
C=C	610
C=C*	505
C≡C	835
C-O	360
C=O	750
C-N	305
N-H	390
O-H	465

*In Benzene

Table 7.3: Values for the amount of energy required to break bonds which are common in amino acids and proteins. The bond energies will vary depending on their position in the molecule.

The earlier study (Vallentyne, 1964) used paper chromatography to identify the pyrolysis products, and discovered two unidentified products (probably representing primary and secondary deamidation); these were not identified in the later study (Vallentyne, 1968) and it was thought that they may have co-eluted with ammonia. Lys molecules could undergo either primary or secondary deamination, producing amines with two different chemical characteristics, which would in turn produce two peaks on chromatograms. There is no evidence in this study to indicate which decay pathway Lys follows, nor the hplc elution time of Lys decomposition products, although it is possible that two of the unidentified peaks represent the decomposition products. No search was made for these compounds.

7.5.3 The effect of carbohydrates on the destruction of amino acids

The reaction between carbohydrates and amino compounds is well documented (e.g. Maillard, 1913; Hoering, 1973; 1980; Furth, 1988), and results in condensation reactions and non-enzymatic glycosylation during the Amadori rearrangement. The reactions begin with interactions between the reducing groups of sugars and amino groups of other compounds to form glycosylamines. The end product of the reaction is a dark compound referred to as melanoidin, which has properties similar to that of humic acid (Hoering, 1973). In this study, no studies regarding potential reactions of carbohydrates with amino acids were undertaken. Carbohydrates, including glucose (H. Clegg, pers. comm.) are present within the shell (Collins *et al.*, 1991b), but it is not certain whether or not they are attached to the protein (M. Cusack, pers. comm.). Studies regarding the amino acid content of acid insoluble compounds from the fossils of this study are detailed in **Chapter Nine**.

Although not examined in this study, carbohydrates have an effect on amino acid decomposition. Pyrolysis reactions involving amino acids and glucose (Vallentyne, 1964) indicate that the higher the concentration of glucose in a standard solution of alanine, the faster the decomposition rate of the amino acid (**figure 7.33**). The 0.37-life of the alanine at 167°C without glucose is just over 10 years, but is reduced to approximately 2 hours when 0.05 M glucose is present, an increase in the rate of decomposition of some 40 000 times.

Carbohydrates clearly have a role in amino acid decomposition, but it is outwith the scope of this study to analyse these features in detail. Current parallel research (H. Clegg) is completing further analyses of some of the carbohydrates over a similar time period from similar samples, and has revealed that glucose is present within the shells. No measurement of concentration of glucose alone in Recent samples has yet been made, although it is clear that the presence in fossils indicates that the carbohydrates were present in the living shell.

Collins *et al.* (1991a) measured the protein and carbohydrate content of Recent samples of *N. lenticularis* and *W. inconspicua*, by colourimetric assays of the EDTA

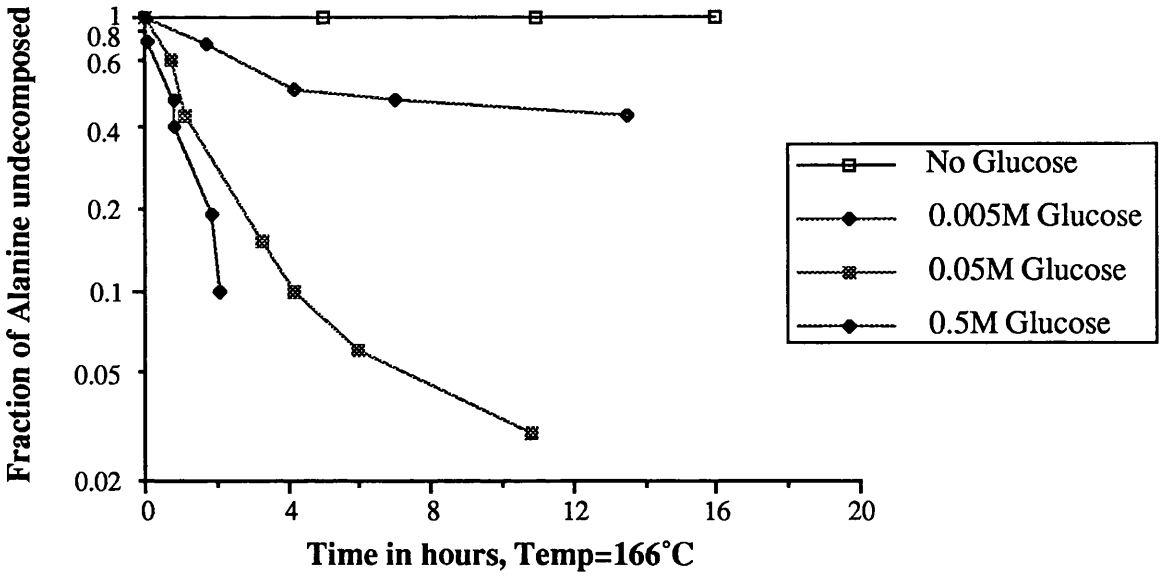


Figure 7.33: The influence of differing concentrations of glucose on the decompositon of Ala based on pyrolysis experiments (after Vallentyne, 1964).

soluble fraction. The method used by Collins *et al.* (1991a) lost small peptides and free amino acids, as indicated by the low levels of recorded protein (less than half of that found in the present study, which did not use filtration). Colourimetric analyses, however, are not the best method of assessing the protein content of these samples, which are bright red in solution (Curry *et al.*, 1991a; 1991b; Cusack *et al.*, 1992), and which will absorb the UV light to give erroneous results. Carbohydrate levels were relatively high, indicating that free sugars are available for the degradative reactions of amino acids.

7.6 Conclusions

7.6.1 The use of pyrolysis as a predictive tool

Pyrolysis experiments have been the only method by which amino acid decay over a long time scale has been examined prior to this study. Several authors have considered the pyrolysis of shell powders, in addition to the pure solutions of amino acids (e.g. Jones and Vallentyne, 1960; Hare and Mitterer, 1969; Totten *et al.*, 1972). Some of the amino acids most stable to pyrolysis, such as Phe, Lys and Asp were some of the least stable of the amino acids when fossilised (Jones and Vallentyne, 1960). The decomposition of Ala was studied by the pyrolysis of shell powders, where it occurred more rapidly than in an aqueous solution of the pure compound, a finding confirmed by this study. This is likely to be due to one of three possibilities (Jones and Vallentyne, 1960; Vallentyne, 1964):

(i) The act of heating does not mimic the effect of time accurately. Implicit in this statement is the effect of pressure on the decomposition of solutions of pure amino acids. The reaction vessels were sealed prior to heating, and the evaporation of the water from the aqueous solution would increase the

pressure. No measure of this pressure was made, although Vallentyne (1964) noted that the vessels frequently shattered during heating, indicating the increase in pressure. The possibility of the increased pressure increasing the rate of reaction has not been considered in any of the pyrolysis experiments.

(ii) The stabilities of the amino acids may be affected by factors other than temperature. For example, metal ions catalyse oxidative deamination, by chelation of the amino acid (Ikawa and Snell, 1954; Nyilasi and Pomogáts, 1964, in Bada, 1971). It is possible that the Ca^{2+} and Mg^{2+} ions present within the shell carbonate could act as a chelation site for the amino acids.

(iii) Carbohydrates increase the rate at which amino acids are degraded. Brachiopods contain intracrystalline carbohydrates (Collins *et al.*, 1991a), which may react with the amino compounds and degrade the molecules more rapidly than may be predicted from pure solutions alone.

Pyrolysis of oyster shells (Totten *et al.*, 1972) showed that more than 90% of the protein is decomposed when shells are heated at 130°C for 1 week, but that a stable residue remains which is little changed when the heating is continued for 9 weeks. The study of the oyster shells considered only the insoluble components of the shell, and cannot therefore be directly compared to this study, which is concerned with the survival of the soluble components. It is also interesting to note that Totten *et al.* (1972) considered soluble intercrystalline components of the oyster shell to be degradation products, and the insolubles to be the true protein. This contrasts to that found for Recent brachiopods, where the intracrystalline protein is soluble in an aqueous solution of acid.

The burial history of the horizons containing the fossils will have a marked effect on the rate of decomposition reaction. Arrhenius equations derived from pyrolysis data (Vallentyne, 1964; 1968) show that the rate of reaction will increase by an order of magnitude between 20 and 40°C. Such an increase in temperature represents burial of less than a kilometre, a possibility which exists for the samples under study. The results of pyrolysis experiments cannot therefore be directly applied to fossil amino acids unless a detailed burial history for each bed is known. Without this, the pyrolysis results cannot be used as a predictive technique for assessing the age or the molecular state of preservation of fossil molecules.

An average temperature cannot be applied to the basin as it has such a complex history. The 0.37-lives for the data are accurate, although it cannot be certain what temperature the sample has been subjected to, hence direct comparisons with pyrolysis results can only be made as a guide to the likely behaviour of the molecule over time in samples which have had a complex thermal history. In this study in particular, the effect of burial is likely to be very important. The South Wanganui Basin has undergone successive periods of subsidence and uplift, which will have buried samples up to 1 km+. If the geothermal gradient is higher than that of the crustal average (25°C/km), then the samples are likely to have been heated to 30-40°C. This will increase the rate of decay of

the molecules by an order of magnitude, and will be the cause of some of the increase in the rate of decomposition.

These results indicate that pyrolysis experiments of pure compounds can act as a useful predictive tool for the study of the decomposition of amino acids in fossils, but that this must be followed by pyrolysis of shells in order to discern the effects of the carbonate, and other factors associated with the shell. Factors other than temperature also appear to be important in pyrolysis, the most obvious of these being pressure, but these have not been considered in pyrolysis experiments. The pyrolysis experiments therefore act as basic information which should be built upon by the detailed study of the pyrolysis of shells, in order to take into account additional features not present when pure solutions are pyrolysed alone. Burial of fossil samples will also have a major influence on decomposition of amino acid molecules, changing reaction rates as the depth of burial fluctuates.

7.6.2 The molecular state of preservation

The proteins of fossil shells are almost completely decomposed, with only a limited number of peptide bonds surviving fossilisation. These bonds are associated with most of the amino acids (identified from the difference between the total and free amino acid concentration), although Thr and Arg are absent from older samples. The presence of unstable molecules such as Ser in peptide bound compounds is likely to be due to the stabilising effect of the peptide bond, the stability of which is a function of the nature of the residues on the other end of the bond. Relatively stable amino acids, such as Ala, Gly, Leu and Ile will have a stabilising effect on the bond, helping to preserve some of the more unstable molecules. All samples have a proportion of free amino acids (although individual amino acids may not be present) which rapidly (within 0.5 Ma) rises to greater than 80% free. The rate of natural hydrolysis then slows and plateaus out, corresponding to the complete destruction of the most labile peptide bonds and the preservation of less labile ones.

Individual amino acids also undergo degradative reactions, the majority of which take place when the molecule has been released from the protein and is present in the free state, resulting in lower concentrations of the original amino acids. Degradative reactions produce a range of reaction products, including diagenetic production of other amino acids which may be either proteinogenic or non-standard molecules. There can be no distinction between the original and diagenetic amino acids, a factor which will distort any taxonomic relationship through time. For example, Jope (1967b) found that the insoluble fraction of the intercrystalline protein from fossil brachiopods showed marked differences from the nearest living relatives. This study also found raised Ser in the insoluble fraction, possibly indicating either post-mortem alteration or a sink for some Ser. Both the present study, and that of Jope (1967b) contrast with the results of Kolesnikov and Prosorovskaya (1986), who recognised "very familiar" compositions between Recent

and fossil brachiopods (dating as far back as the Jurassic). Such results need to be treated with caution, as this study has shown that, at least in the soluble fraction, the amino acids are unstable to differing degrees, which results in changing amino acid ratios over time, and fossil samples, as a consequence, will have different amino acid ratios to those of extant species.

Due to the level of both natural hydrolysis and amino acid decomposition that has taken place in the fossil samples, it is highly unlikely that proteins from these samples will survive in a state whereby they may be routinely separated and analysed for primary sequence data. The separation of intercrystalline molecules from 0.3 Ma samples of foraminifera (Robbins and Brew, 1991), represents remarkable molecular preservation. The use of dialysis indicates that polypeptides must survive in the foraminifera tests, although at lower concentrations than in the Recent, and that some of the molecules are no longer present. This decay indicates that the same processes found in brachiopods are occurring in the foraminifera, although at an apparently slower rate, which may be caused by the lack of burial or heating of the samples. Amino acids are being lost from the intercrystalline proteins, probably due to leaching. The advantage of using intracrystalline molecules in the present study is therefore to be certain of *in situ* degradation, where the products of degradative reactions are also contained within the shell and are not lost. Although the bulk amino acid composition of samples represents an oversimplification of the nature of the proteins, it is likely to be the sole method of analysis in older samples, where the proteins have been totally degraded. If the molecules are undergoing *in situ* decay, then the loss of molecules will be from degradation of the molecules (i.e. the loss of amino or carboxyl groups) rather than diffusion of original amino acids out of the shell, as is the case for intercrystalline molecules.

The concentration of amino acids in the shells increases in some cases between the Recent and the youngest fossils (**figure 7.2**). Three hypotheses may be considered to explain this phenomena:

(a) Intercrystalline molecules could become entombed by recrystallisation of the shell, or by carbonate cement from the sediment, and may then be protected from digestion by sodium hypochlorite, and hence be added to the intracrystalline molecules, increasing their concentration.

(b) Metal ions within the shell could act as chelation sites for free amino acids (i.e. amino acids released by natural hydrolysis of the proteins), increasing their concentration at certain points within the shell. This could also help to explain some of the variation found in amino acid concentration.

(c) Examination of the insoluble fraction of Recent samples (**Chapter Nine**) reveals that they contain acid-insoluble proteins. These proteins will undergo natural hydrolysis, releasing amino acids into the soluble fraction, thus increasing the amino acid concentration.

Possibility (a) may be ruled out as the state of preservation of the majority of

the shells is excellent, with no evidence of either recrystallisation or cementation. Possibility (b) may be ruled out as metal ions will also affect the stability of amino acids. Chelation of the molecules will take place either when they are bonded, or when they are present in the free state, although this will only increase concentration at certain points in the shell. This will be averaged out throughout the shell. Possibility (c), the solubilisation of the Recent insoluble intracrystalline molecules, is therefore the most likely of the processes. As acid insoluble samples from Recent samples age, they release peptides and free amino acids which are taken into solution by decalcification, effectively increasing the concentration of the amino acids.

7.6.3 Presence of non-protein amino acids within the shell

As described above, both β-alanine and γ-aminobutyric acid are present as intracrystalline molecules in fossil brachiopods. From kinetic studies and the examination of foraminifera, Schroeder (1975) concluded that the presence of these amino acids was indicative of contamination of the foraminifera matrix from the sediment. The data presented here contradict this. Low concentrations of γ-aminobutyric acid are present in the sample, indicating low rates of conversion from Glu. For samples with relatively low concentrations of Asp, there is very little or no recorded conversion to β-alanine. *Notosaria*, however, has a high concentration of Asp, and older samples show an increase in the concentration of β-alanine, although this only accounts for c. 45% of the decomposition of the Asp. As the amino acids have been released during decalcification from within the shell, it is unlikely that the non-protein amino acids were derived from the sediment in which the fossils are contained, and are more likely to be indigenous to the fossil, formed as a result of the diagenetic degradation of other amino acids. The presence of these amino acids should therefore not be regarded as being due to contamination of the organic extract of the sample.

In summary, the following general conclusions may be made:

(i) All samples analysed contained relatively high concentrations of amino acids, some of which remain peptide-bound.

(ii) Peptide bonds decay rapidly in the fossil record, due to the influence of temperature, water and the nature of the residues on either side of the peptide bond.

(iii) Any proteins that were originally present in the sample are highly degraded, with up to 80% of the intracrystalline amino acids occurring in the free state by 0.5 Ma. Some peptide bound amino acids still occur within the shell, but the size of these remaining peptides is not known. These proteins are more highly degraded than would be expected from the results of previous studies (e.g. Jope, 1980; Armstrong *et al.*, 1983).

(iv) Certain unstable amino acids may be preserved for longer in the fossil record when they are peptide bound, rather than when they are in the free state.

(v) The method described in **Chapter Three** allows the analysis of high concentrations of amino acids from small samples of shell (c. 20 mg), without requiring filtration and concentration of samples, and hence small peptides and free amino acids remain within the sample.

(vi) Individual amino acids decay more rapidly in the free state than they do when they are peptide-bound, presumably due to the stabilising effect of the other amino acid attached to the peptide bond, stabilising the entire bond.

(vii) Pyrolysis experiments on decay of amino acids provide a basic framework for amino acid decay reactions, but are difficult to apply to fossil samples due to changing rates of reaction during the burial of the samples.

Potential uses of fossilised proteins and amino acids

8.1 Introduction

The existence of fossil peptides and amino acids within fossil shells has been demonstrated both in the present and earlier studies (e.g. Abelson, 1954; Jope, 1967b; Wyckoff, 1972; Weiner, 1976; Collins *et al.*, 1991b). The geological and palaeontological significance of these compounds, however, are not obvious from the data derived from these fossil compounds alone, and most reports concerning these molecules have been purely descriptive. Molecular biology has shown that intact proteins contain high levels of taxonomic information in their primary sequence, and the present study has shown that this information can also be gained from statistical analysis of the amino acid composition of Recent samples (**Chapter Five**; Cornish-Bowden, 1979; 1983).

Molecules which have undergone fossilisation, however, are highly degraded and have lost much of their original information through peptide bond and amino acid degradation (**Chapter Seven**). Previous studies, whilst being mainly concerned with Recent materials, have examined the taxonomic uses of fossil amino acids in foraminifera (King and Hare, 1972; Haugen *et al.*, 1989) and a preliminary study of brachiopods (Curry *et al.*, 1991b). The aim of this section is to synthesize the information regarding the molecular survival of amino acids gained during the present study, and to examine the potential palaeontological and geological information which may be gained from this data.

8.2 Methods and Materials

Amino acid data was recovered from fossil samples as described previously (**Chapters Three and Seven**). The concentration of each amino acid was converted to a weight% in order to normalise the data, and remove the effect of the large concentration of Gly from the calculations. If the values are not normalised, then the extremes of concentration within the sample outweigh the smaller differences in residues between the samples.

Following conversion, the data was analysed by the statistical analysis program DATADESK™, for the Macintosh microcomputer. This program uses multivariate statistical techniques to summarise the data into a manageable form (**section 1.6**), from 14 variables down to 2 or 3. Once completed, the PCA and cluster analyses were examined in order to elucidate taxonomic information from the fossil amino acids alone.

8.3 Palaeontological Uses

8.3.1 Introduction

The stated long term aim of molecular palaeontology is the direct sequencing of fossil peptides (Curry, 1988). It has long been recognised that fossil peptides were likely to be highly degraded and that the resulting mixtures would probably be so complex as to defy further purification (Abelson, 1954; Akiyama, 1971; Curry, 1988). The use of intracrystalline macromolecules, it was hoped, would allow the recovery of fossil peptides from the shell, as the inorganic phase would protect the proteins from decay (Towe, 1980; Curry, 1988; Collins *et al.*, 1988; Endo, 1992). The present study has indicated that the level of degradation of the peptides is much higher than initially thought, and that the purification of specific fractions of the degradation products of peptides is not possible. A similar conclusion was independently reached by Collins *et al.* (1991a) in a study of macromolecules of a similar age. It is doubtful whether the direct sequencing of fossil peptides would in fact reveal any useful information regarding the taxonomy or phylogeny of the organism, as there would be no information concerning the location of the fossil peptide in any of the corresponding proteins recovered from extant representatives. Unless a mosaic of overlapping fossil peptides could be used to reconstruct a fossil protein, the rates of amino acid substitution in proteins could not be measured. The decomposition of amino acids will distort any reconstruction and hence short chains of uncertain location will not be of any taxonomic value in molecular palaeontology.

The evolution of proteins in a sample may be traced by the change in amino acid residues. As these substitutions only affect relatively few sites in brachiopod proteins (Cusack *et al.*, 1992), then it is likely that these changes would not be observed in fossil peptides. The decay of unstable amino acids from the sample will also distort the relationships; if the substituted amino acids were Ser or Arg, then preservation in fossils would not be as likely as if the amino acid were Val or Gly.

Indirect examination of fossil peptides through immunology has a better chance of revealing useful taxonomic information, as this method relies on the reaction of antibodies, produced in response to intact proteins, with their degraded counterparts in the fossil record. If the antibody is made against a specific protein, then this protein may be recognised in the fossil record, even if only a small portion of the protein is preserved (i.e. the determinant). As the determinants are small, there is an increased probability that these short chains will be preserved within the fossil record (Endo, 1992). Collins *et al.* (1991b) demonstrated that this was indeed the case, and that immunologically active material could be recognised in fossils up to 2.3 Ma, although reactivity was at a much reduced level, corresponding to the decay of the majority of the determinants. The present study has shown that rapid degradation of all amino acids takes place, and that this is coupled with rapid natural hydrolysis over the 2.5 Ma studied. This indicates that immunology would not be routinely successful in older samples. Other studies (e.g. de Jong *et al.*, 1974;

Westbroek *et al.*, 1979) have detected the immunological reaction in older samples, although the acquisition of this information is not available on a routine basis. Immunology is a very specific detection technique, although it is possible that the antibodies are produced against EDTA/peptide or protein complexes rather than the organic molecules alone. The immunological reaction of fossil samples may therefore be directed towards the EDTA/protein complex, rather than the peptide alone.

However, immunology represents the only realistic method of recognising and characterising specific proteins within the fossil record, although to date, antibodies have only been produced against crude shell extracts which will include molecules other than proteins (Collins *et al.*, 1988; 1991a). Immunological studies would prove to be more effective if the crude extract were fractionated and the homogeneous components used to elicit the immune response. This would prevent the production of antibodies against the non-proteinaceous molecules, improving the specifity of the reaction. New methods in the preparation of intracrystalline molecules from Recent samples (Cusack *et al.*, 1992) will increase the likelihood of the production of these specific antibodies.

Previous studies have used the amino acid content of Recent and fossil foraminifera to complete 'chemotaxonomy', the classification of the organisms by amino acid content alone (e.g. King and Hare, 1972; Haugen *et al.*, 1989). These studies have also utilised PCA to summarise data derived from the samples, and hence group samples with similar positions on principal component axes. These studies have included both fossil and Recent data in the same calculations, which has led to a large spread of data within the analyses. Haugen *et al.* (1989) also recorded an increasing Gly/Ala ratio, which is indicative of the diagenetic production of Ala from other amino acids, similar to that described in **Chapter Seven**. Haugen *et al.* (1989) used ratios between Asp/Glu, Glu/Gly, Glu/Ala, Gly/Ala, and Leu/Val in addition to the diastereoisomeric ratio of isoleucine as variables. From the data presented in **section 6.6.3**, it is obvious that samples of different ages would have undergone decay to different levels, and that the ratios would be correspondingly different for samples of the same species, but of different ages, introducing a diagenetic, as well as a genetic component to the analyses, leading to the spread of data in the analyses.

King and Hare (1972) approached the method in a different way, concentrating mainly on the classification of Recent foraminifera, but also examining fossil samples in order to determine whether or not the genetic differences between species are preserved within the fossil, or destroyed during fossilisation. It was concluded that there were changes in the amino acid composition in the fossil record, but that this did not affect the relationships between some of the more stable of the amino acids. The study concluded that evolutionary changes could be traced through studies of the biochemical content of the fossils, although the authors were not able to say how far back in the fossil record such changes could be utilised.

In the present study, the aim was to attempt to extract palaeontological

information from the amino acids and to use statistics to attempt to relate samples on the basis of their amino acid composition alone. Interpretation was made in two ways, both within and between horizons, in order to determine how time will affect the separation of groupings identified in **Chapter Five**, which are described below.

8.3.2 Palaeontological information gained from within horizons

(i) Introduction

Samples collected from the same horizon will be of approximately the same age, and will have been subjected to the same geological processes during their history. The effect of this is to render the horizon as a time plane (similar to that of the Recent, a 'snapshot' of geological time), whereby changes in the amino acid content due to diagenetic alteration of amino acids will be of approximately the same order in all samples, and hence differences between the amino acid compositions will be due to the initial biochemical composition of the species alone. This is obviously an over simplification of possible relationships, and the amino acid composition of the fossils will be distorted over time by, for example, the rate and degree of diagenetic production of some amino acids, which will in turn depend on the initial concentration, and the effect of carbohydrates and of different mixture of amino acids in the sample. However, as the amino acids are contained within a single time plane, and provided that there has been no homogenisation of the amino acid composition of the samples in the horizon through time, similar methods of taxonomic discrimination can be utilised as for the Recent samples (**Chapter Five**).

Principal Component Analysis relies on sufficient information being present in the variation of the datasets in order to calculate meaningful latent vectors which summarise this information (**section 1.6**). However, this requires that there are more samples in the dataset than there are variables (i.e. that there are more samples than there are amino acids). For this reason, not all of the horizons from which the fossils were collected may be analysed by PCA.

(ii) Results

(a) Recent samples

Taxonomic information extracted from Recent samples is considered in **Chapter Five**, and will not be discussed further here.

(b) Rapanui Marine Sand (Waipipi Beach; **locality 3, figure 2.2**)

The Rapanui Marine Sand is the youngest of the horizons considered in the present study, dated at 0.2 Ma. The first three principal components (**table 8.1**) contain 86% of the total variation of the dataset, mainly due to Glu, Gly, Thr, Pro and Leu for the first, Asp, Ala, Tyr, Val, Ile and Phe for the second, and Ser, Pro, Phe and Lys for the third. Graphical representation of the first three principal components (**figure 8.1**) shows that separation of samples by this method is good to at least the subordinal level. The specimens

EigenValues

	Values	Variance Proportion
e1	8.792	62.8
e2	2.713	19.4
e3	1.708	12.2

EigenVectors

	V1	V2	V3
D	0.222	0.442	0.007
E	-0.246	0.180	-0.218
S	-0.304	0.035	0.271
G	0.329	-0.096	0.116
R	-0.216	-0.408	0.262
T	-0.288	0.227	0.213
A	-0.240	0.371	-0.227
P	-0.076	-0.142	-0.709
Y	-0.131	-0.534	-0.025
V	-0.249	-0.217	-0.409
I	-0.310	-0.185	0.048
L	-0.330	-0.026	0.072
F	-0.325	-0.136	0.070
K	-0.329	-0.018	0.129

Table 8.1: Principal component analysis of the relative proportion data calculated for samples of fossils from the Rapanui Fm, Waipipi Beach. Only the first three eigenvectors and eigenvalues are shown.

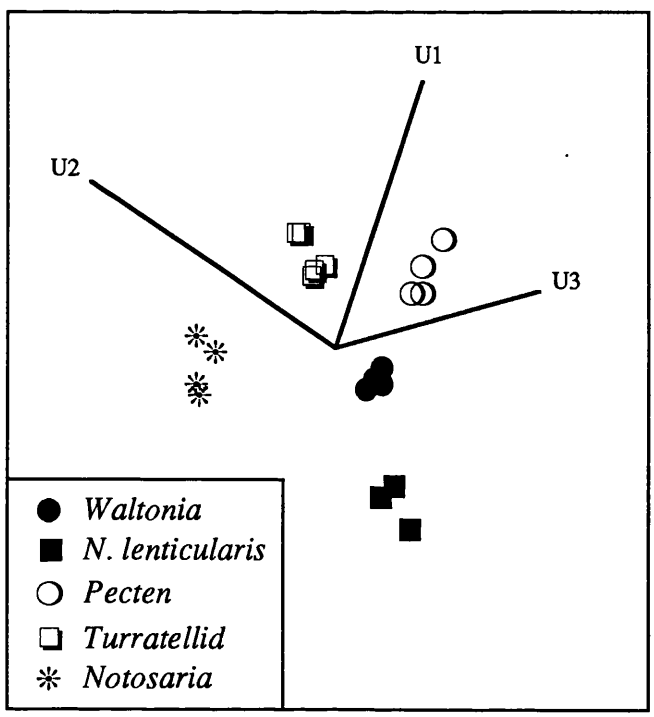


Figure 8.1: Graph of the first three principal components calculated for the relative proportions of amino acids from the Rapanui Fm, Waipipi Beach. All samples are well separated along the axes indicating that some taxonomic information is preserved in the fossil amino acids.

of *Neothyris* present in the sample appear to be older than the remaining samples in the horizon (**sections 8.4.2 and 8.4.3**), and this difference in age may distort the analysis to some degree. Arg is the only amino acid not to have a major effect on the dataset. As all of the other amino acids have some effect on the variation, it may be concluded that any analysis which does not include the vast majority of amino acids will be discarding some important taxonomic information.

(c) Tainui Shellbed (The Buttress (Fleming, 1953), Castlecliff Beach; **locality 6**)

The Tainui Shellbed contains a diverse fauna, amongst which the brachiopods *Notosaria*, *Waltonia*, *T. sanguinea* and *Neothyris* are all present. Outgroups of a turritellid (gastropod) and a pectenid were also chosen for further study. PCA (**Chapter Two**) of the samples shows that 89.5% of the variance was contained within the first three eigenvalues (**table 8.2**), making taxonomic conclusions possible (Sneath and Sokal, 1973). The variability of the first principal component is caused mainly by Ser, Ala, Pro, Ile, Leu and Lys (**table 8.2**), the second by Asp, Glu, Gly, Tyr, Val, and Phe, and the third by Arg, Thr,

Eigenvalues

	Values	Variance Proportion
e1	6.535	46.9
e2	3.586	25.6
e3	2.381	17.0
		Total 89.5

Eigenvectors

	V1	**V2**	**V3**
D	0.072	-0.490	0.176
E	-0.370	-0.020	-0.086
S	-0.247	-0.296	-0.296
G	0.306	0.190	-0.316
R	-0.269	0.009	-0.108
T	0.204	-0.266	0.432
A	-0.093	-0.411	-0.289
P	-0.288	0.307	0.196
Y	0.118	0.340	0.329
V	-0.069	0.401	-0.368
I	-0.357	0.025	-0.105
L	-0.356	0.090	0.221
F	-0.296	0.047	0.384
K	-0.370	-0.124	0.048

Table 8.2: Principal component analysis of the relative proportion data calculated for samples of fossils from the Tainui Shellbed, Castlecliff Beach. Only the first three eigenvectors and eigenvalues are shown.

Val and Lys. This shows that the differences between the samples are expressed in all of the amino acids present within the shell, and that the variation is not dependant on only the high concentration of Gly. Three dimensional plots of the samples (**figure 8.2**) show that there is good separation of the genera in space. Although this is not apparent from the 2D representation shown in **figure 8.3**, the *Pecten* samples plot away from the brachiopods, as do the turratellids. These results are to be expected, and show that there has been no homogenisation of the amino acid composition in samples through the horizon.

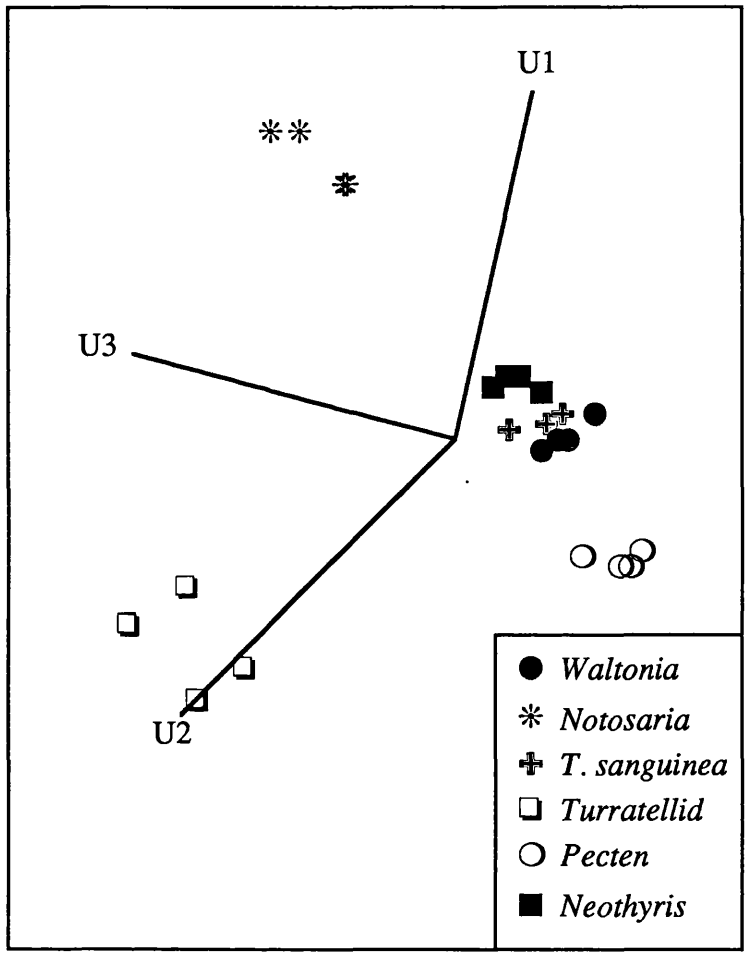


Figure 8.2: Graph of the first three principal components calculated for the relative proportions of amino acids from the Tainui Shellbed. Separation of the species is at the subordinal level, although the Terebratulida plot close together, they may still be separated.

The brachiopod samples are well separated at the ordinal level, with *Notosaria* (Rhynchonellida) plotting well away from the three species assigned to the Terebratulida. In **figure 8.3** the Terebratulida are grouped together, due to the lack of 'fine tuning' given by the third principal component in **figure 8.2**, which shows that *Waltonia*, *Neothyris* and *T. sanguinea* may be separated. This method of separating samples on the basis of their amino acid composition alone may therefore be applied to samples of 0.4 Ma, and be accurate to at least the subordinal level. The difference between the 2D and 3D plots

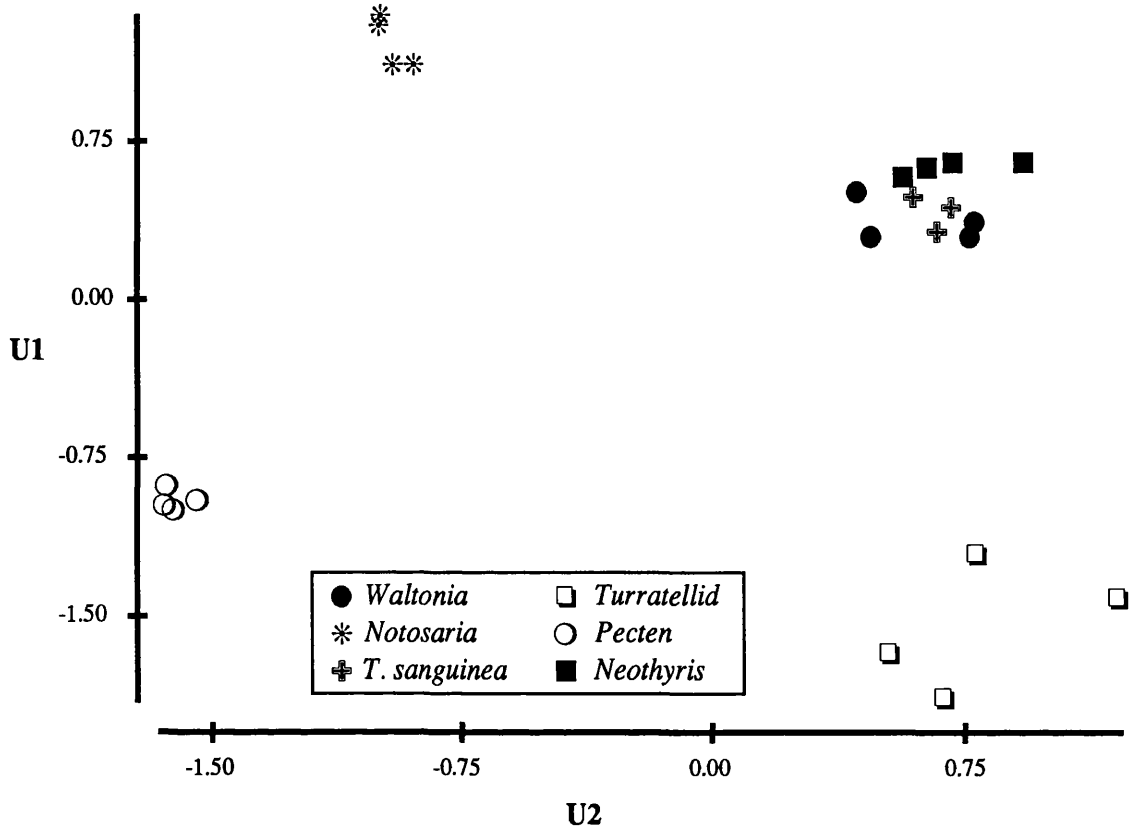


Figure 8.3: Graph of the first two principal components calculated for the relative proportion data for amino acids from the Tainui Shellbed. Separation on this graph is only possible to the ordinal level, showing the loss of sensitivity provided by the third principal component.

indicate the importance of ensuring that sufficient of the variability of the dataset is contained within the eigenvalues to be plotted. In **figure 8.3**, there is only 72.5% of the variability represented and that further details could be elucidated with the use of more of the sample variation.

(d) Pinnacle Sand (The Pinnacles, (Fleming, 1953), Castlecliff Beach; **locality 7**)

The Pinnacle Sand contains a fauna similar to that of the overlying Tainui Shellbed, but is not as rich or diverse. The same species were analysed as above, and PCA completed. The first three principal components contain 92.6% of the variation of the samples (**table 8.3**). The first principal component has variation mainly due to the concentration of Gly, Arg, Ala, Ile and Lys, the second due to Asp, Glu, Ser, Thr, Pro, Leu and Val, and the third to Gly, Thr, Tyr, Val and Phe. Both 3-dimensional and 2-dimensional plots (**figures 8.4 and 8.5**) show good separation of all the samples. The additional variation contained within these components has allowed separation of *Waltonia* from the other members of the Terebratulida in the 2D plot. Once again, the 3D plot is able to separate all of the samples considered, indicating that the genetic, and therefore taxonomic signal, is preserved in the fossil record to 0.45 Ma.

Eigenvalues

	Values	Variance Proportion
e1	6.109	43.6
e2	3.849	27.5
e3	3.003	21.5
		Total92.6

Eigenvectors

	V1	V2	V3
D	-0.205	-0.391	0.211
E	-0.279	0.347	0.081
S	-0.269	-0.306	-0.224
G	0.323	-0.080	-0.332
R	-0.336	-0.042	-0.155
T	0.070	-0.316	0.433
A	-0.366	-0.053	0.023
P	-0.036	0.500	0.050
Y	0.266	-0.044	0.364
V	0.159	0.311	-0.374
I	-0.313	-0.100	-0.268
L	-0.311	0.288	0.147
F	-0.098	0.280	0.456
K	-0.388	-0.038	-0.023

Table 8.3: Principal component analysis of the relative proportion data calculated for samples of fossils from the Pinnacle Sand, Castlecliff Beach. Only the first three eigenvectors and eigenvalues are shown.

(e) Lower Castlecliff Shellbed (combination of coastal and Waipuka Road; **localities 8 and 9**)

Samples from this horizon are beginning to show the influence of geological time on the information available from the PCA analysis (**figure 8.6**). The first three principal components contain 83.7% of the dataset variation (**table 8.4**), and hence the conclusions gained from the 3D plot are valid (Sneath and Sokal, 1973). The variation is due to Asp, Gly, Ala, Ile, Leu and Lys in the first principal component, the second by Glu, Ser, Pro, and Phe, and the third has variation mainly due to Arg and Tyr. From **figure 8.6**, it can be seen that the outgroups are well separated from the remainder of the data, although the brachiopod samples assigned to the Terebratulida are coming closer together, and samples are beginning to merge. At a higher scale the consensus positions for the samples are separate, showing that some of the taxonomic information is preserved, although the identification needs to take into account experimental error, which still might cause this consensus to be inaccurate.

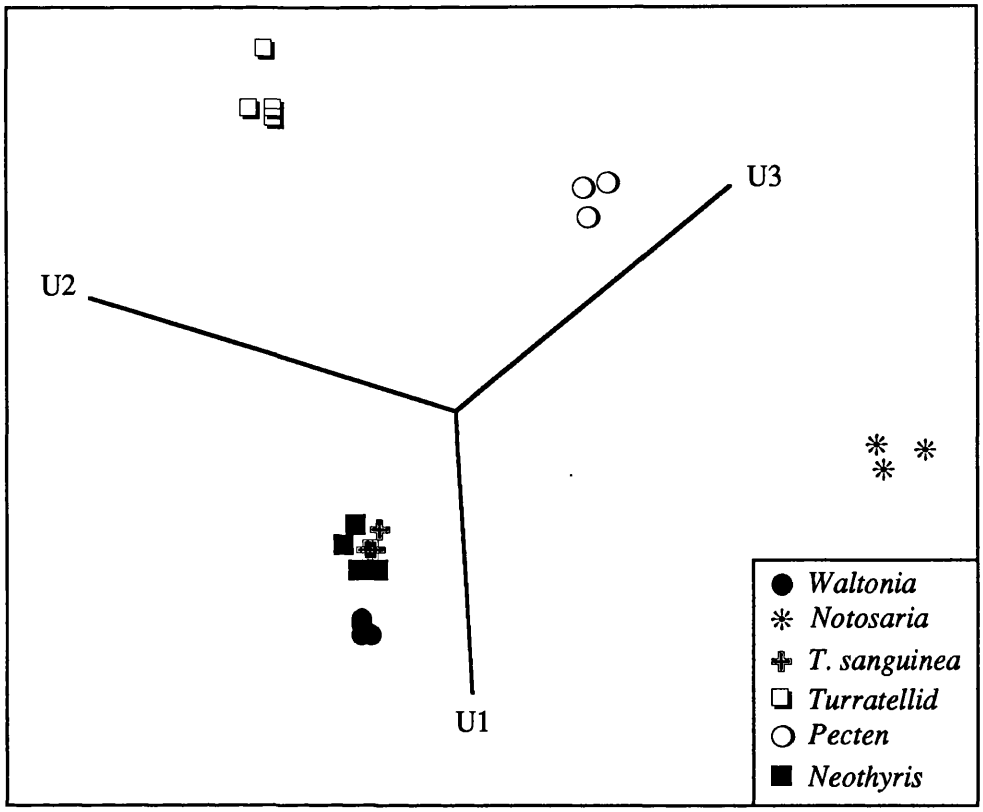


Figure 8.4: Graph of the first three principal components calculated for the relative proportions of amino acids from the Pinnacle Sand. Separation of the samples is still at the subordinal level.

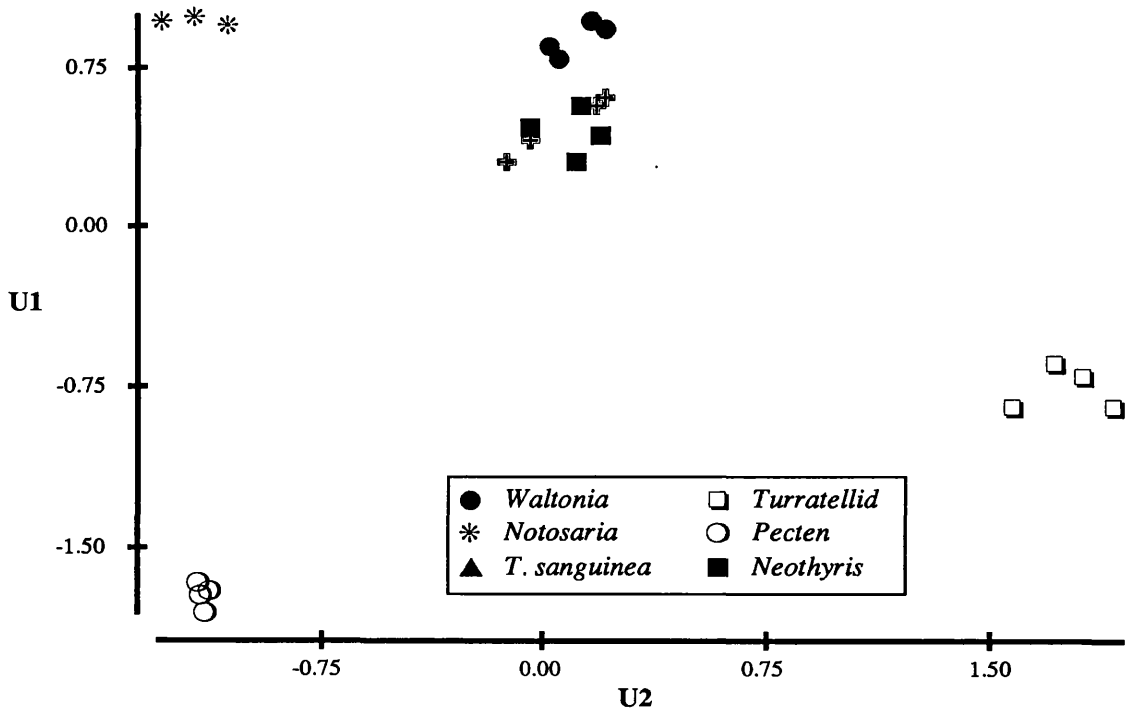


Figure 8.5: Graph of the first two principal components calculated for the relative proportion data for amino acids from the Pinnacle sand. Separation on this graph is possible to the subordinal level, reflecting the higher levels of variation contained within the first two principal components (**table 8.3**).

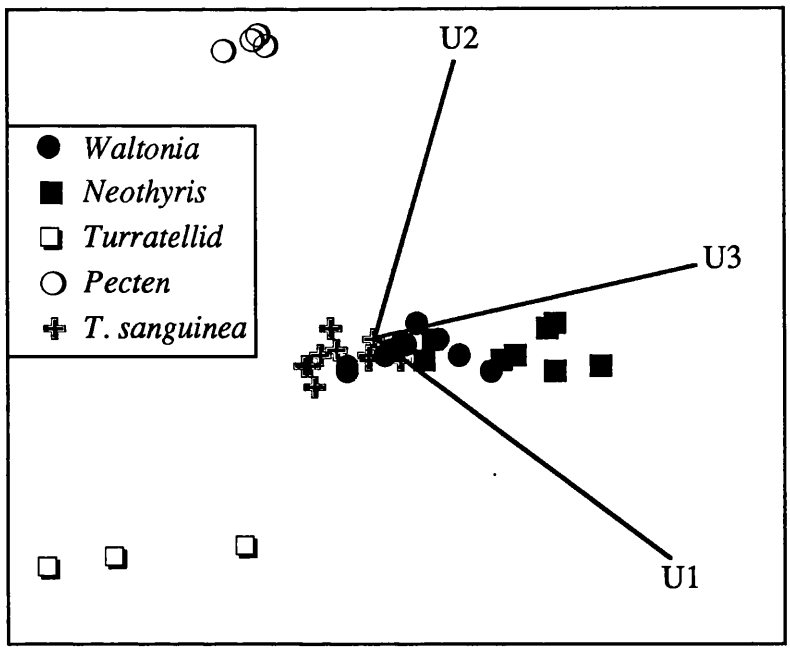


Figure 8.6: Graph of the first three principal components calculated for the relative proportions of amino acids from the Lower Castlecliff Shellbed. Separation of the species is no longer possible at the subordinal level. The samples representing the Terebratulida are grouped together illustrating the close relationship of these samples, and the loss of the differences between the samples.

Eigenvalues

	Values	Variance Proportion
e1	7.355	52.5
e2	3.184	22.7
e3	1.196	8.5
		Total83.7

Eigenvectors

	V1	V2	V3
D	-0.311	0.275	0.059
E	-0.269	-0.349	0.006
S	-0.266	0.316	0.113
G	0.350	0.147	-0.016
R	-0.196	-0.226	0.565
T	-0.276	0.110	-0.060
A	-0.301	0.077	0.238
P	-0.030	-0.522	-0.284
Y	0.179	-0.208	0.635
V	0.186	-0.139	-0.233
I	-0.314	0.137	-0.198
L	-0.320	-0.270	-0.070
F	-0.217	-0.404	-0.108
K	-0.342	0.145	-0.103

Table 8.4: Principal component analysis of the relative proportion data calculated for fossil samples from the Lower Castlecliff Shellbed, Waipuka Road and Castlecliff Beach. Only the first three eigenvectors and eigenvalues are shown.

The merging of the datapoints for the brachiopods must represent the decay of the unstable amino acid molecules which are important in differentiating between species, and which has an endpoint with the amino acid content being similar in all samples (although not the same). This equates with the substantial decay of the unstable amino acids and the diagenetic production of others. The data in this form is not useless, and unknowns from the same horizon could still be analysed for taxonomic position.

(f) Kupe Formation

(1) Samples erroneously collected as the Rapanui Formation, Castlecliff Beach (see **section 8.4.3**; **locality 1**)

The first three principal components derived from the data for this of the Kupe Fm represent 87.9% of the total variability of the dataset (**table 8.5**), mainly due to variation in the relative concentration of Glu, Pro, Ile, Leu and Lys for the first principal component. The second has variation mainly due to Asp, Ser, Gly, Thr, Ala and Val, and the third due mainly to Ser, Gly, Thr and Phe. The data shown in **figure 8.7** shows some similarity to that for the Lower Castlecliff Shellbed, in that the data for the Terebratulida is grouped closely together and that *Notosaria* and the two outgroups are well separated. The data for this horizon shows increased merging of the data for the Terebratulida, equating with the further decomposition of amino acids. The samples of *Notosaria* still plot well away from the other brachiopods, indicative of separation to the ordinal level.

Eigenvalues

	Values	Variance Proportion
e1	7.644	54.6
e2	2.826	20.2
e3	1.829	13.1
		Total 87.9

Eigenvectors

	V1	V2	V3
D	0.161	0.510	-0.153
E	-0.320	-0.054	-0.083
S	-0.175	0.346	0.459
G	0.232	-0.331	0.383
R	-0.283	-0.043	0.004
T	0.242	0.303	-0.371
A	-0.239	0.366	0.259
P	-0.293	-0.232	-0.290
Y	0.272	-0.131	-0.150
V	-0.214	-0.373	0.130
I	-0.305	-0.013	0.202
L	-0.334	-0.005	-0.267
F	-0.287	-0.015	-0.411
K	-0.315	0.263	0.061

Table 8.5: Principal component analysis of the relative proportion data calculated for fossil samples collected from the Kupe Fm (collected as the ?Rapanui Fm, Castlecliff). Only the first three eigenvectors and eigenvalues are shown.

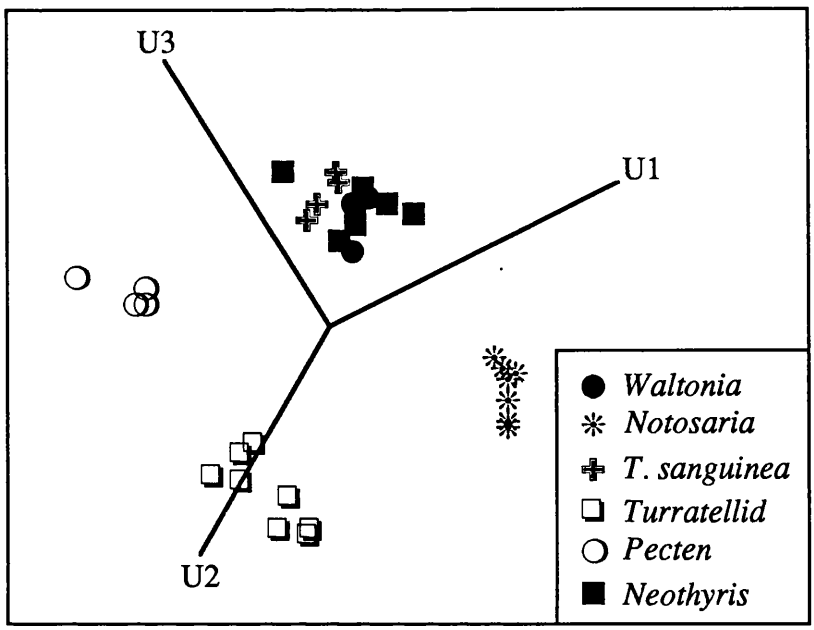


Figure 8.7: Graph of the first three principal components calculated for the relative proportions of amino acids from the Kupe Fm (collected as Rapanui Fm, Castlecliff). Separation of the species remains at the ordinal level with *Notosaria* samples plotting away from the rest of the samples.

(2) Castlecliff Beach (**locality 10**)

These samples did not include either *Notosaria* or a pectenid, which has allowed the samples to be spread further out in the graphs produced by the PCA. The first three principal components contain 95.0% of the variation of the dataset (**table 8.6**), due

Eigenvalues

	Values	Variance Proportion
e1	7.428	67.5
e2	1.901	17.3
e3	1.127	10.2
		Total95.0

Eigenvectors

	V1	V2	V3
D	-0.032	-0.363	0.779
E	-0.347	0.168	0.149
G	0.366	-0.029	-0.057
A	-0.330	-0.281	0.087
P	-0.362	0.015	-0.062
Y	0.101	0.524	0.508
V	-0.026	-0.653	-0.123
I	-0.308	0.232	-0.282
L	-0.365	0.052	0.026
F	-0.366	0.047	0.020
K	-0.363	-0.021	0.056

Table 8.6: Principal component analysis of the relative proportion data calculated for fossil samples collected from the Kupe Fm, Castlecliff Beach. Ser, Arg and Thr are excluded from the analysis as they produce unacceptable variations in the analysis due to the low levels of amino acid present. Only the first three eigenvectors and eigenvalues are shown.

mainly to the variation of Glu, Gly, Ala, Pro, Ile, Leu, Phe and Lys for the first principal component, Asp, Tyr and Val for the second, and Asp and Tyr for the third. The separation of the samples by PCA in this horizon (**figure 8.8**) is better than that found for the Kupe Fm from a different locality. This is due to there being fewer samples that have a very different amino acid composition, in the analysis. In this case, differences in the Terebratulida are more apparent, with *Neothyris* being separated from members of a different subfamily (see **Chapter Five**). A second difference between this sample and that of the above is the omission of Ser, Arg and Thr from the original data. This is due to variation in the proportion of these amino acids. As the concentration of amino acids falls in older samples, the detected level comes closer to the threshold value of the analysis system. Once the level falls below a few picomoles, then the data recorder does not record that amino acid, even though it may be present in lower concentrations. As some samples include Ser, Arg and Thr and some do not, these amino acids act to distort the true analysis are are hence omitted. Taxonomic discrimination of the samples is therefore possible in the Kupe Fm (0.5 Ma) to at least the subfamilial level, on the basis of amino acid content alone.

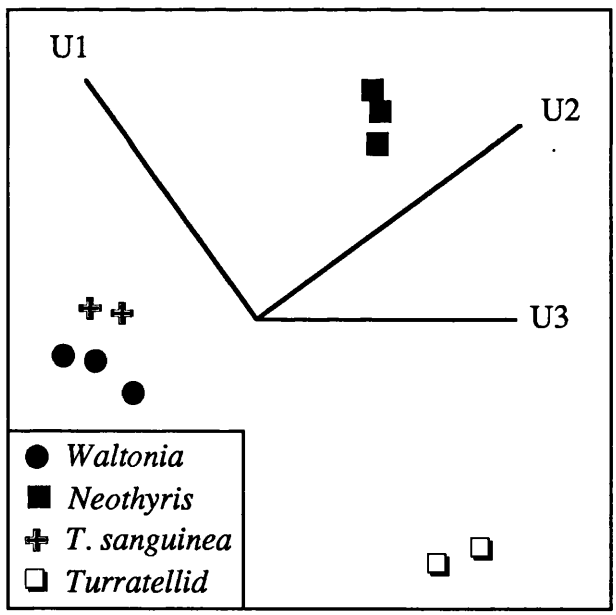


Figure 8.8: Graph of the first three principal components calculated for the relative proportions of amino acids from the Kupe Fm. Separation of the samples is still at the ordinal level.

(g) Upper Okiwa Group (Parapara Road, E of Wanganui; **locality 17**)

The data for the Upper Okiwa Group shows that 91.8% of the variation of the dataset is contained within the first three principal components (**table 8.7**). This is due mainly to Asp, Glu, Pro, Ile, Leu, Phe and Lys for the first principal component, Gly, Ala, Tyr and Val for the second and Asp, Ala, Phe and Lys for the third. It should be noted that the proportions for Ser, Arg and Thr were variable, and for reasons discussed above were ommitted from the PCA. The 3D and 2D plots of these principal components (**figures 8.9 and 8.10**) show that the samples are well separated by the amino acid data. Both the outgroups and *Notosaria* plot away from the Terebratulida. Within this latter group,

Eigenvalues

	Values	Variance Proportion
e1	6.765	56.4
e2	3.347	27.9
e3	0.905	7.5
		Total 91.8

Eigenvectors

	V1	V2	V3
D	-0.300	0.191	0.448
E	-0.375	0.054	0.137
G	0.258	-0.390	-0.171
T	0.229	0.436	-0.046
A	0.161	0.381	0.519
P	-0.372	-0.069	0.028
Y	0.228	0.402	-0.036
V	-0.189	-0.417	0.282
I	-0.328	0.032	-0.148
L	-0.336	0.048	0.192
F	-0.298	0.269	-0.362
K	-0.299	0.240	-0.458

Table 8.7: Principal component analysis of the relative proportion data calculated for fossil samples collected from the Upper Okiwa Group, Parapara Road. Ser, Arg and Thr are excluded from the analysis as they produce unacceptable variations in the analysis due to the low levels of amino acid present. Only the first three eigenvectors and eigenvalues are shown.

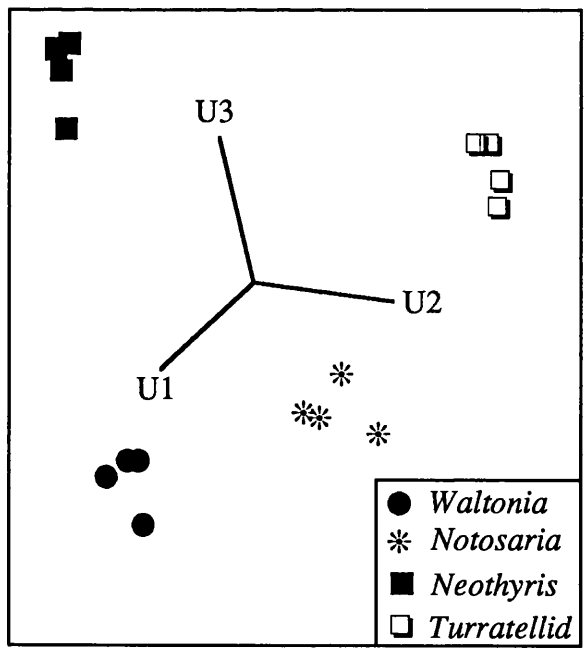


Figure 8.9: Graph of the first three principal components calculated for the relative proportions of amino acids from samples from the Upper Okiwa Group. Separation of the species is still basically at the ordinal level, although within the Terebratulida *Waltonia* and *Neothyris* are well separated. This may indicate the fewer datapoints available for this horizon.

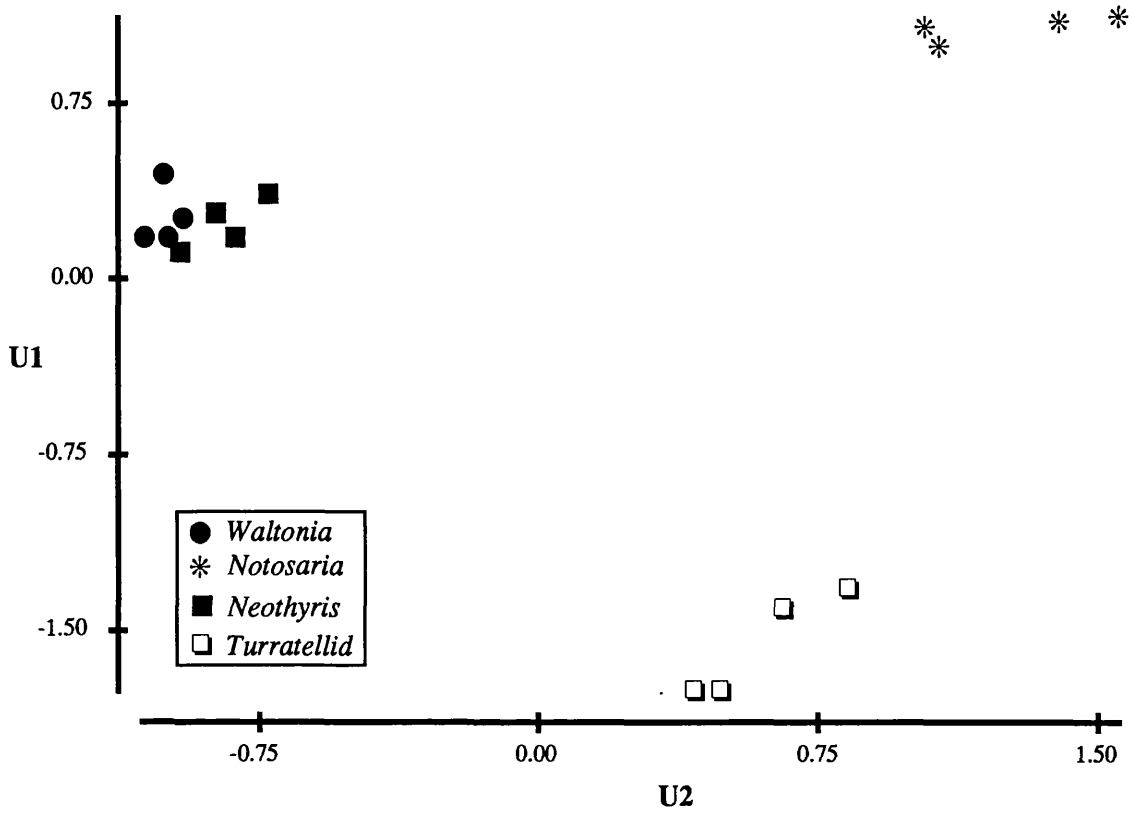


Figure 8.10: Graph of the first two principal components calculated for the relative proportion data for amino acids from samples collected from the Upper Okiwa Group. Separation on this graph is possible to the ordinal level.

Waltonia and *Neothyris* are also well separated. This shows that separation is possible in these samples to the subfamilial level. This increased level of differentiation may be due to the inclusion of only one representative outgroup. Sufficient information is present in the first two principal components to use a 2D scatterplot, which show groupings of the Terebratulida distinct from the Rhynchonellida and from the outgroup (turratellid).

(iii) Discussion

The amino acid compositions of fossils are complex datasets, containing up to 14 variables. Information contained within datasets of this size are difficult to assimilate, and it is difficult to understand the relationships between amino acids, as these are between every member of the dataset. PCA has the advantage of summarising the data into fewer, derived, variables which may then be used to differentiate the samples. The results from the PCA of the amino acid data extracted from fossils show that, within horizons, samples may be separated to at least the subordinal level in all samples, on the basis of their amino acid composition alone, and in some cases this separation is to the subfamilial level. These diagrams may be considered to be analogous to geochemical discrimination diagrams, as the majority of the groupings described above would be recognised, even if morphologically derived groupings were not known.

Older samples contain amino acids and proteins which have been degraded to

a higher degree than have those of younger samples, and this is recognised by the merging of the representative points for similar samples. The data shows that brachiopods have amino acid compositions which are very different to that of the outgroups, members of a different phylum. Separations at the ordinal level, between the Rhynchonellida and the Terebratulida, are also very distinct. One of the major drawbacks of this method is that the samples are separate groupings, and although similar samples tend to be grouped close together, this relationship is not necessarily phylogenetic. What is obvious however, is that each grouping has a characteristic amino acid signature, and that this signature is sufficiently different in all the analysed horizons to allow separation of the samples. This signature must reflect original genetic differences between the samples (c.f. **Chapter Five**), although the level of available information is not as high as it would be if sequence data for the parent proteins were available. The merging of the samples in the plots of the principal component data show that the differences between the amino acid compositions are being lost. As degradation proceeds, then differences between the relative amino acid composition will be reduced (by the loss of the less stable molecules), and, as a consequence, the sample data points will begin to merge.

8.3.3 Palaeontological information derivable between horizons

(i) Introduction

Conventional taxonomic and phylogenetic analysis relies on following a morphotype through geological time, and assessing the changes in morphology and growth in relation to speciation events. Molecular palaeontology, to assist in such evolutionary studies, must be able to complement this process. The previous sections have revealed the state of molecular preservation of these fossils, the intracrystalline proteins and amino acids degrade relatively rapidly over the 2.5 Ma of the study. This degradation of amino acids will lead to increasing and decreasing relative proportions of the molecules, changing the data for the PCA. As a consequence of this, it is to be expected that the resolution of the PCA will decrease as samples of increasing age are analysed. As PCA is derived from a specific dataset (i.e. the amino acid content of fossils from a horizon), then the graphical representation of each horizon cannot be directly compared (as the information in each diagram is sourced from different data). In order to compare data from different horizons it is therefore necessary to recalculate the data, and to complete a new PCA.

Molecular evolution is determined on the basis of the nucleotide substitution in the DNA molecule which acts as the template for the manufacture of that protein (**Chapter One**). Changes in the genetic composition will manifest themselves in changes in amino acid composition, and hence it would be expected that speciation would be marked by a deviation of the sample away from the general grouping of samples which have a similar amino acid composition. At present, however, separation of fossil species on the basis of their amino acid compositions has not been demonstrated, the method has

only been used at the ordinal level. It should prove possible to trace the development of the major lineages through their amino acid composition. The most direct investigation of molecular speciation would be in the sequence of the fossil protein, although as discussed above, this is unlikely.

(ii) Results

All samples analysed in this study were incorporated into the same dataset and a new PCA completed, in order to ascertain whether or not a taxonomic signal was preserved through geological time at a high enough level to cause similar samples to plot close together. The relative proportions of Ser, Arg and Thr were omitted from this calculation, as they have decomposed in some of the older samples do not contain them, due to decomposition reactions lowering their concentration to below the threshold value of the analysis system. The first three principal components contain 78.5% of the total variation present in the dataset, thus indicating that the derived variables contain sufficient information to make taxonomic conclusions (Sneath and Sokal, 1973). **Figure 8.11** shows the plot of the first three principal components, and four groupings may be identified. As this is a 2D representation of multidimensional space, some of these groupings are line of sight, and a full rotation is necessary to split up the samples. However, the groupings show that some degree of taxonomic separation on the basis of amino acids alone is possible from the dataset, including both Recent and fossils, back to 2.5 Ma.

The graphical representation of the PCA data shows a clear separation of the samples into four major groups. The two outgroups, pectenids and turratellids, form distinct groupings, as would be expected from members of a different phylum. The brachiopods form two groups, with the Rhynchonellida grouping away from the Terebratulida. Within the Terebratulida, no differentiation can be made, as the variation in the data causes a spread which encompasses the data from the entire order. Several of the samples plot away from their respective groupings, which is caused by spurious data. Within each of the groupings, there is considerable spread of the samples. This is the effect of time on the samples, and will be discussed in further detail below.

(iii) Discussion

In the above section, it was demonstrated that taxonomic information could be derived for samples within individual horizons. This is due to the amino acids and proteins within the samples having undergone similar rates and degrees of decomposition. When samples of different ages were compared, however, it was suspected that the taxonomic signal present in each of the samples would become mixed up due to time, i.e. that the effects of differential time on the amino acid composition of different samples would render the information provided by the amino acids useless. The data in **figure 8.11**, however, suggests otherwise, and that some level of taxonomic significance occurs in each of the samples, which allows grouping of similar organisms to be made.

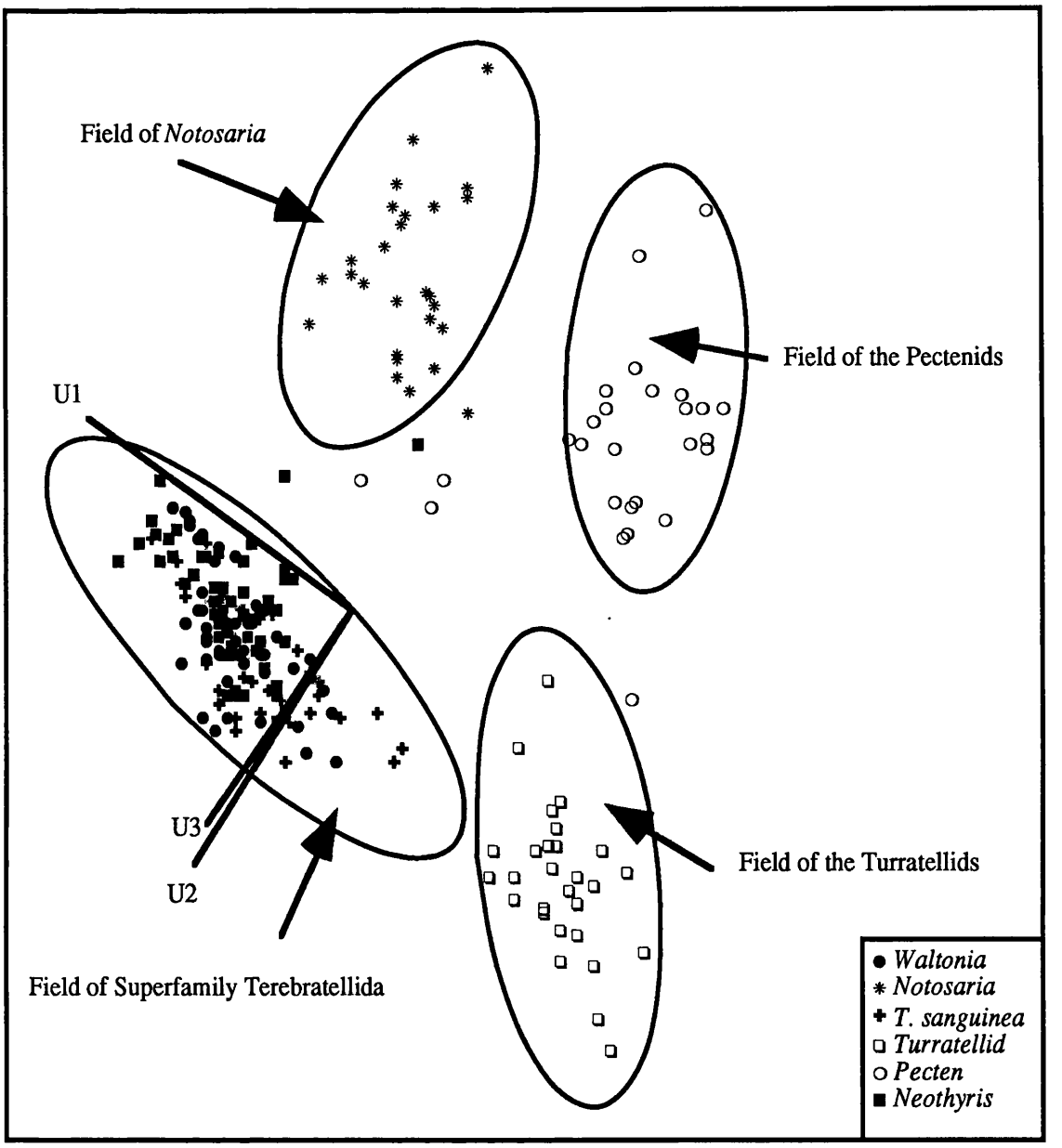


Figure 8.11: Graph of the first three principal components calculated for the relative proportion of amino acids from all samples analysed. The graph shows separation of the samples at the ordinal level, although some samples do not fall within the general fields allocated to the four major groups. This shows that separation is possible to the ordinal level in fossil samples, which may assist in the ordinal assignment of problematica.

It is obvious that the degradation of the amino acids has an influence on the datasets, causing the spread of the sample data, probably reflecting the loss of different concentrations of unstable amino acids. This degradation does not distort the amino acid signature of the sample to a level where it is similar to others from a different order. Samples of a particular type have an initial ratio of individual amino acids, and degradation of these amino acids over time follows a pattern which is similar for all brachiopod species analysed, and it is likely that the same will hold true for other samples, as, once free from their proteins, the amino acids will behave as individual molecules rather than proteins,

provided that they remain within the shell and are not released by shell recrystallisation etc. Degradation of the amino acids occurs within the samples, but the relationships between these amino acids must not change significantly over time, thus allowing similar samples to be grouped together. There is some change due to the effect of time on the samples, indicated by the spread of the samples within the groupings, representing the decay and diagenetic production of amino acids. This relationship is more complex than the simple ratio between two amino acids, and incorporates all of the amino acids present, with the exception of those which are readily decomposed. If Ser, Arg and Thr are incorporated into the dataset, a similar grouping is obtained, although the amount of variation contained within the first three principal components drops to c. 73%, just below the acceptable level of significance.

Ratios between pairs of amino acids have been used in previous studies of amino acid taxonomy (Jope, 1967b; Bada and Man, 1980; Haugen *et al.*, 1989), to differentiate between species. From the data presented in this study, it may be seen that the ratio between the pairs of amino acids range over a wide scale, and that these ratios overlap. Clearly, and as would be expected from such a complex mixture of natural compounds, the relationships of all amino acids to each other must be taken into account in order to ascertain if any taxonomic information is preserved. PCA is one of the methods by which this may be completed.

The level of discrimination of this method, however, is not as high as was initially hoped. In Recent samples, the method can distinguish between genera in all cases, and also possibly to the specific level (investigated within the *Neothyris* genus), although more samples need to be completed prior to reaching any firm conclusions regarding this. In fossil samples, the amino acid data tends to merge for members of the same phylogenetic order, representing the loss of the amino acids which discriminate between them. This study has demonstrated that samples may be separated to at least the ordinal level between horizons on the basis of their amino acid composition alone. The fossil samples plot close to Recent samples of a related or similar genus, whereas fossil samples of other phyla or orders plot further away (**figure 8.11**). It is not yet certain how far back into the fossil record this method may be extended. The results from this study indicate that the samples may be classified to at least the ordinal level, which will be of use in the ordinal assignment of problematica, although it is unlikely to assist in specific/generic assignments.

The method of preparing shells to release incarcerated amino acids (**Chapter Three**) used in this study have two advantages over most previous studies into the amino acid composition of fossils:

(a) The decay of intercrystalline molecules releases peptides and free amino acids, which may be lost from the shell by leaching processes. As decay of the intracrystalline proteins takes place *in situ*, then the taxonomic signal will be preserved for longer in fossil samples.

(b) The use of the entire extract, without the need for concentration and filtration

also preserves the small peptides and free amino acids, which would be lost during this step.

The method relies on sufficient concentrations of amino acids being liberated from the shell powder to provide taxonomic information and that the amino acids sampled are representative of the original starting concentration i.e. free amino acids and petides are not lost. In this study, an acid concentration of 2N was used, as this successfully decalcified the powders, but did not damage the shell proteins too badly (**Chapters Three and Five**). Older samples have already undergone severe degradation, resulting in the hydrolysis of the majority of the peptide bonds, and hence there is no need to prevent damage of any preserved peptides or proteins, as there are very few remaining. For older samples, where amino acid degradation has proceeded further, the concentration may not be high enough to register on the analysis system. Preliminary studies with 6N HCl indicate that this decalcification solution effectively trebles the concentration of the recovered amino acids, as the volume of acid required is only a third of that used in 2N decalcification.

8.4 Geological Uses

8.4.1 Introduction

The sections above have dealt with the uses of intracrystalline amino acids from fossil organisms in a palaeontological context. Molecules within the shells of fossils can also provide information on a range of other geological topics. The site of these molecules, incarcerated within calcium carbonate, limits the potential for contamination of the samples and thus it can be assumed, provided that care is taken, that information gained from within the shells is a true reflection of the organic composition of the sample. The following sections describe potential geological uses for these amino acids and proteins. Several of these uses have already been applied during the course of this study, and examples are given for these, whereas others are simply possible uses for the molecules.

8.4.2 Source and age of derived material

Samples which are not indigenous to the horizon from which they were collected will have an amino acid composition very different to that predicted by the state of preservation of other samples within the same horizon. Derived samples may be older than others from the same shellbed, and consequently their intracrystalline molecules may have undergone a higher degree of molecular degradation. An example of this may be seen in samples of *Neothyris* collected from the Rapanui Marine Sand at Waipipi Beach. The sample was collected as broken fragments, and is clearly derived from another source. This is confirmed by the amino acid composition of the sample (**figure 8.12**), which contains no Arg, and very low concentrations of Thr and Ser (present only in peptide

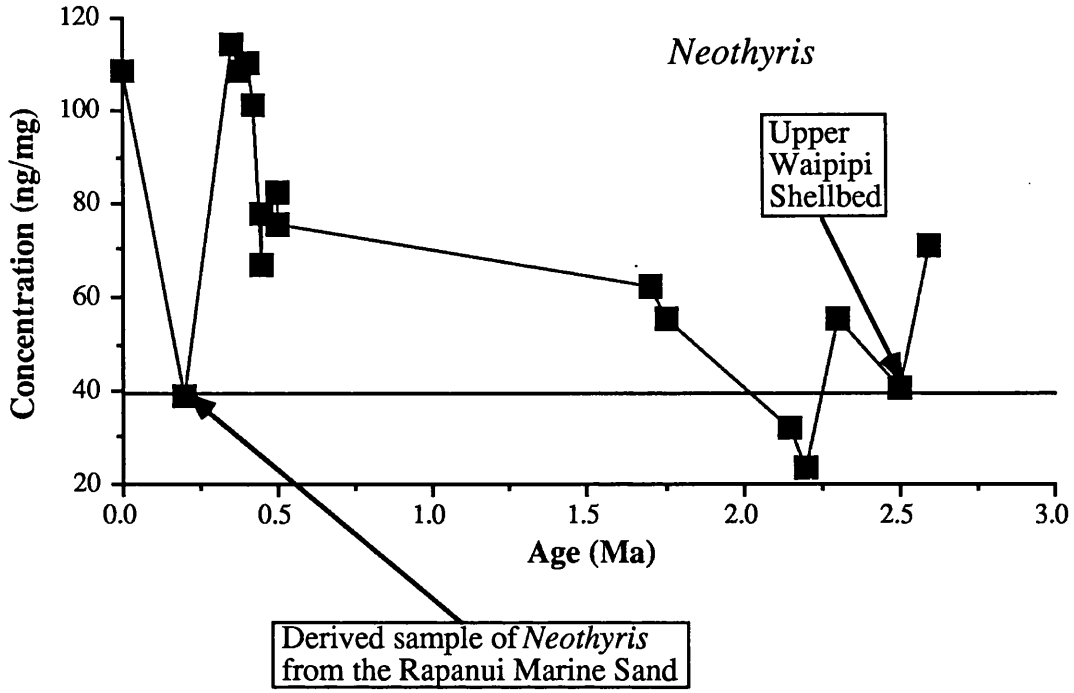


Figure 8.12: The sample of *Neothyris* collected from the Rapanui Marine Sand has a much lower concentration of amino acids than would be expected for such a young sample. By comparison, it is possible to assign the derived sample to its correct source.

bonded samples). The %change of the other amino acids are very different to those of other fossils from the same horizon. For these reasons, it may be concluded that *Neothyris* from the Rapanui Marine Sand is not indigenous to that shellbed. Comparison of the amino acid composition of *Neothyris* from the other shellbeds indicates that the likely source for the *Neothyris* from the Rapanui Marine Sand is either the Middle or Upper Waipipi Shellbed, dated at approximately 2.4 Ma, and that one of these shellbeds was being eroded as the Rapanui Marine Sand was deposited.

8.4.3 Amino Acid Dating Techniques

(i) Amino Acid Racemisation

Although not completed in this study, the quantification of amino acid enantomers is the most studied of the reactions pertaining to fossil amino acids (for a review, see Schroeder and Bada, 1976). With the exception of Gly, all amino acids may occur in one of two forms, the D- or L-form, depending on the direction which they rotate a beam of plane polarised light. Only L-forms are present in biological molecules, although over geological time, rearrangement of the structure of the molecule occurs about the centre of asymmetry. This leads to the production of D-amino acids (the mirror image of the L-form) in increasing quantities, until an equilibrium, where L-form ≈ D-form is set up. The ratio between the two has been used as a dating tool (Schroeder and Bada, 1976), calibrated by the use of ratios from deep sea cores. The time scale produced from the reaction is, however, not absolute (Bada and Man, 1980) as changes in temperature causes

changes in the rate of racemisation reaction. This will effect the results of such a dating technique, although it may be used to provide a rough indication of the age when there are no other methods available. As this method has not been considered in this study, it will not be discussed further.

(ii) Peptide bond degradation dating

Several studies (e.g. Bada and Man, 1980) have examined the decomposition of the peptide bond, and the time scale for these bonds to be decomposed in the fossil record. In carbonate rich sediments, the decomposition of half of the peptide bonds (estimated from the composition of the protein, peptide and free amino acid fractions) takes approximately one million years, whereas those in clay rich horizons this process takes 1-2 Ma. At the beginning of this study, it was decided to attempt to relate the decomposition of the peptide bond to the age of the sample, as it was expected that the proportion of free amino acids within the sample would increase as a direct result of the age of the sample, and that the changing proportion of amino acids which are free could be used as a guide to date sequences.

In general terms, the proportion of free amino acids (on the basis of individual molecules) does increase as a function of time, allowing age comparison. In greater detail, however, the concentration and proportion of individual free amino acids may be variable. This is a result of the decomposition of amino acid molecules in the free state occurring more rapidly than either the release of molecules from peptides or the decomposition of amino acids which remain bound. The cleavage of the peptide bond is not simply a function of time, as temperature and the nature of the residues on either side of the peptide bond will also effect the rate of decomposition, thus not allowing generalisations regarding the rate of peptide cleavage. Peptide bond cleavage will therefore be species dependent (as the proteins within the species will have different amino acid residues present) to a certain extent. The dating of samples by peptide bond cleavage is possible within species, and, to a lesser extent, between species. The major problem with this method is the later stage decomposition of the free amino acids, most notably seen in Arg. Arg decreases relatively rapidly once the peak of 100% free has been reached. Comparisons with the proportion free of other amino acids are necessary in order to ascertain whether the composition of such amino acids are on the upward or downward curve.

(iii) Amino acid decomposition dating

(a) Introduction

This study has shown that amino acids present within the shells of fossils decompose at a rate which may be described (in most cases) by exponential or logarithmic curves. As the rate of amino acid decay is dependent on a number of independent factors, this method of dating will have some error associated with it. Provided that the samples studied have extant representatives with a known amino acid composition, and that

samples have a similar burial history, then samples of unknown age may be dated by the decomposition of their constituent amino acids. This will be especially useful for correlation across basins. The method will be illustrated by the use of specific examples from the South Wanganui Basin.

(b) *Neothyris* from the Rapanui Marine Sand, Waipipi Beach

As considered above, the amino acid composition of this sample has indicated that this sample is not indigenous to the Rapanui Marine Sand, and is over 2 Ma older than that shellbed. This was determined by the state of preservation of the amino acids (**figure 8.12**).

(c) Samples collected as the Rapanui Marine Sand, Castlecliff Beach

During the field season of 1991, a landslip occurred from the cliffs at Castlecliff Beach, close to the mouth of Omapu Creek (**figure 8.13**). This landslip oc-

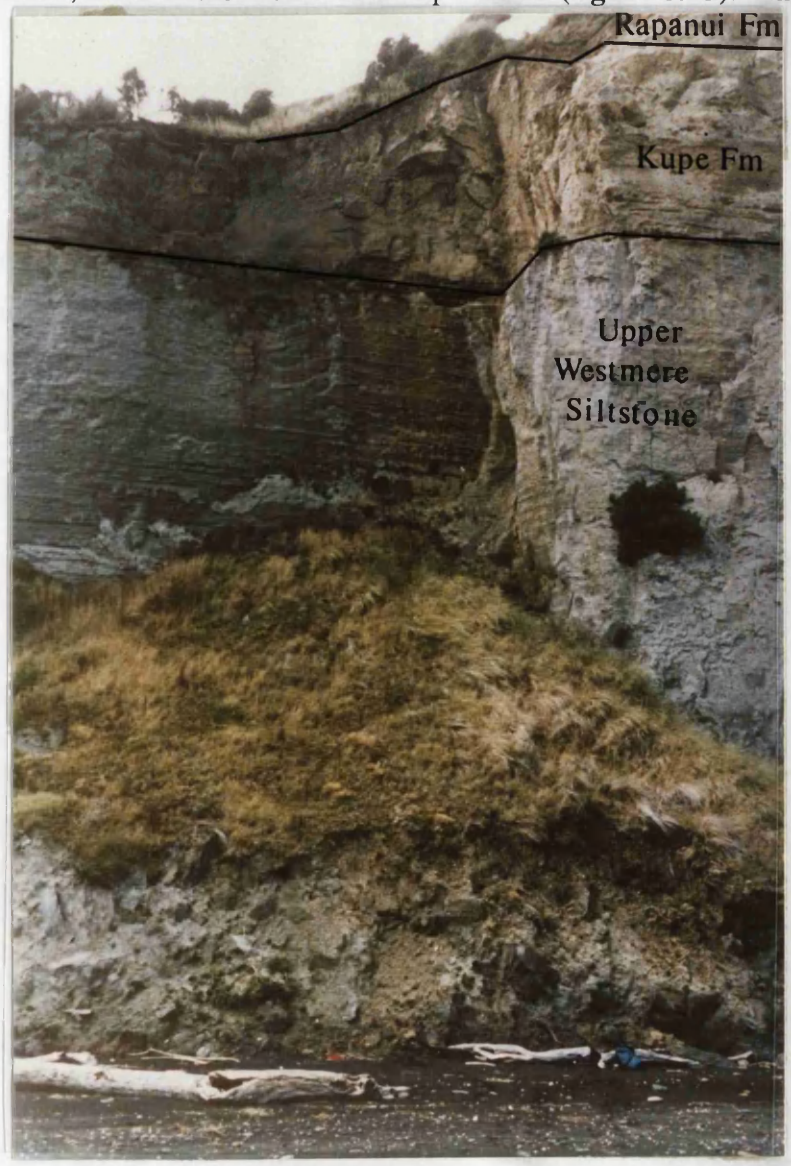


Figure 8.13: Photograph of the landslide which occurred during the field season of 1991, from which the ?Rapanui Fm, Castlecliff samples were collected.

curred in an area where fossiliferous rocks of both the Kupe Formation and the Rapanui Marine Sand were present. In the field, the state of preservation of these samples led to identification of the samples as being sourced from the Rapanui Marine Sand. Photographs of the location were taken, and samples labelled. The samples were prepared as detailed in **Chapter Three**, and the samples analysed for their amino acid content. After multiple analyses of these samples, it was clear that the state of molecular preservation (**figure 8.14**) indicated that considerable degradation of the molecules had taken place, and that the samples could not have been sourced from the Rapanui Marine Sand, but were from an older shellbed. Comparisons with the amino acid composition of other shellbeds indicated that the state of molecular preservation was similar to that of the Kupe Formation. The samples were therefore assigned to their correct bed mainly on the basis of the state of amino acid preservation.

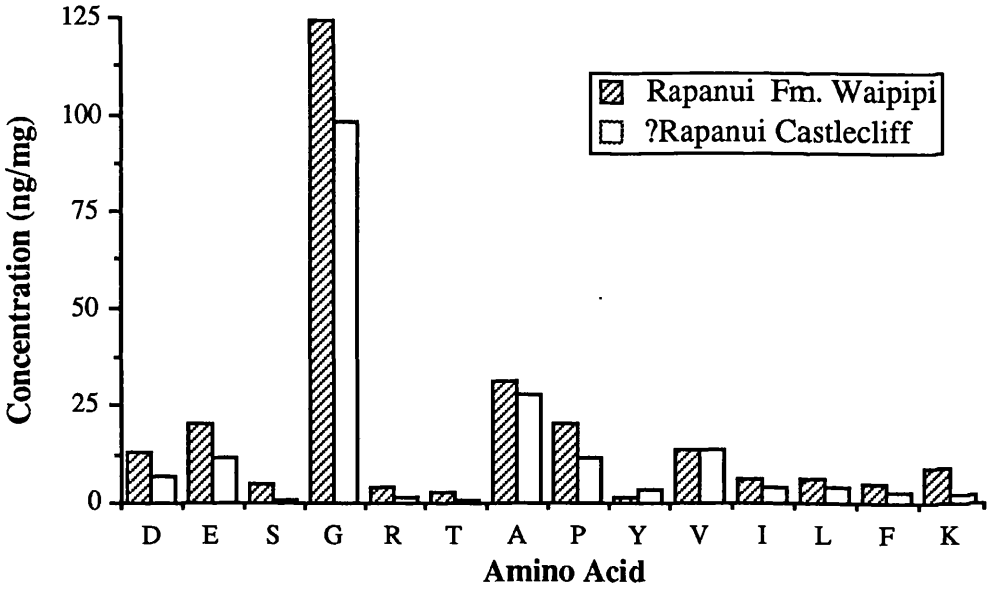


Figure 8.14: Bar chart of the concentration of amino acids from the type section of the Rapanui Fm at Waipipi Beach and the sample collected as ?Rapanui Fm, Castlecliff Beach. The concentration of the amino acids reveals that the sample from ?Rapanui Fm, Castlecliff Beach is much older than 0.2 Ma, and is therefore not from the Rapanui Fm.

(d) Problems with amino acid decomposition dating

The examples listed above use the amino acid decomposition reaction as a *relative* dating method. Other studies (e.g. Bada *et al.*, 1978; Bada and Man, 1980) have attempted to use these reactions as an *absolute* dating technique. However, dates produced from these attempts are acknowledged to be very different to the absolute dates derived by other methods of analysis. This is due to the changing rate of decomposition reactions following burial and associated temperature changes (**section 7.6**). For this reason, the method should not be used as an absolute dating method, although it is possible to use it as a relative dating technique within a single basin, where burial histories, and therefore the rates of degradative reactions, will be similar.

8.4.4 Other information derivable from intracrystalline molecules

Parallel research to that described in this study is investigating the way in which intracrystalline molecules may be involved in the process of biomineralisation (e.g. Cusack *et al.*, 1992). These molecules are in a position whereby they may influence the shape of biominerals or the location of biomineralisation (Weiner and Lowenstam, 1989). Further study of molecules from Recent samples should indicate the ways in which these molecules influence these processes. Research is also underway concerning the carbon and oxygen isotopes from shell macromolecules, and the way in which these may provide information regarding the diet or palaeoenvironments of the fossils examined. As these molecules are trapped within the shells, the original isotopic ratios will be preserved, rather than any diagenetic overprint.

8.5 Conclusions

This chapter has been concerned with the preservation of proteins and amino acids within the shells of fossil organisms. Several specific conclusions regarding the use of these molecules can be made from the data presented here. Despite the high levels of amino acid degradation, taxonomic information is preserved in intracrystalline molecules, which may be observed by the use of graphical presentation of multivariate statistical analysis of the relative proportion data of amino acids. The taxonomic level of this information is to at least the subordinal level when single horizons are studied, and to the ordinal level when all samples are studied together. This is a lower level of taxonomic value than was hoped at the start of this study, but still represents the preservation of characteristic amino acid signatures for orders of brachiopods which may be useful in the assignment of problematica.

Apart from taxonomy, the preservation of fossilised amino acids has a number of other applications, including relative dating and palaeoenvironmental reconstructions.

Chapter Nine: Amino acids from the acid insoluble fraction

Amino Acids from the acid insoluble fraction

9.1 Introduction

The acid decalcification of fossil and some Recent shell samples does not progress to completion. After the calcium carbonate of the shell has dissolved, a residue remains which is resistant to further dissolution with acid. This acid insoluble fraction has been the main method of analysing fossil proteins in previous work (e.g. Jope, 1967b; Akiyama, 1971; Armstrong *et al.*, 1983; Kolesnikov and Prosorovskaya, 1986), as it was this fraction which was thought to contain the true protein of the fossil sample. In the early part of this study, however, it was found that the intracrystalline protein from Recent brachiopods was mostly soluble in acid solution, and thus the main period of research was directed away from the insoluble fraction (which was routinely brought out of suspension by centrifugation and discarded).

The acid insoluble fraction has undergone a preliminary study with the following aims:

(i) To quantify the proportion of the shell which is insoluble in acid solution, and to examine the relationship between this proportion and the age of the sample.

(ii) To assess the amino acid content of the insoluble fraction, with particular emphasis on those molecules that are rapidly lost from the soluble fraction (**Chapter Seven**) and to compare the amino acid composition with that of the soluble fraction.

The nature and source of the insoluble compounds has been investigated in previous studies, and did not undergo investigation in this preliminary study. Humic acid indigenous to the shell, and which released amino acids on hydrolysis of the compound, was described by Hare and Hoering (1977), and identified both by amino acid enantomers and by isotopes. Earlier studies had identified the similarity between naturally occurring humic acids and melanoidins (Hoering, 1973), compounds which result from the reaction between carbohydrates and amino compounds (Maillard, 1913). It is therefore assumed, in this study, that the acid insoluble compound present within the shell is melanoidin, a humic acid-like compound formed by the reaction between the active groups of amino acids and carbohydrates. Free sugars have been described from within the shells of both Recent (Collins *et al.*, 1991a) and fossil (H. Clegg, pers. comm.) brachiopods, which are available for this reaction, although the concentration of the free sugar available for reaction is much less than that of the amino acid, suggesting that this reaction may not proceed to completion (i.e. not all amino acids react with carbohydrates).

9.2 Methods and Materials

Samples were prepared as described in **Chapter Three**. In summary, clean, bleached shell powders were decalcified in the presence of excess HCl (2N). Once the reaction was complete, the insoluble compounds were washed repeatedly with MilliQ™, centrifuged, lyophilised and weighed. The compounds were transferred to glass hydrolysis tubes and hydrolysed by 6N HCl vapour in an oven held at 165°C for 60 minutes. Released amino acids were solubilised with an aqueous solution (0.025% w/v) of K_3EDTA, and an aliquot of this added directly to the sample frit of the analysis system.

9.3 Results

The results of this analysis are considered in two ways in the following sections, in order to assess the proportion of insoluble compounds in the shell and also to examine the amino acid composition of this fraction.

9.3.1 Quantification of acid insoluble compounds

The results of this analysis are summarised in **tables 9.1-9.4**. All samples, including the Recent, contain some insoluble compounds which vary in colouration from colourless to brown to black. Several samples were lost during the extensive washing procedure, and these are omitted from the data shown in the tables. The results are

Bed	Age (Ma)	Sample wt (mg)	Insoluble wt (mg)	%insoluble
Recent	0	81.0	0.3	0.370
Landguard Sand	0.35	149.0	0.0	0.000
Upr Castlecliff Shellbed	0.38	93.1	0.5	0.537
Tainui Shellbed	0.40	94.6	0.2	0.211
Pinnacle Sand	0.42	121.9	0.5	0.410
Lr Castlecliff Shb (Waipuka Rd)	0.47	142.8	0.5	0.350
Lr Castlecliff Shb (Castlecliff)	0.47	125.0	0.4	0.320
Kupe Fm	0.50	69.6	0.1	0.144
Waipuru Shellbed	1.72	126.0	0.3	0.238
Undifferentiated Shb (4243)	1.77	74.2	0.4	0.539
Upr Okiwa Group	2.15	84.1	0.1	0.119
Hautawa Shellbed	2.20	105.8	0.3	0.284
Te Rama Shellbed	2.30	102.9	0.1	0.097
Upper Waipipi Shellbed	2.50	108.8	0.1	0.092
Middle Waipipi Shellbed	2.60	67.5	0.2	0.296

Table 9.1: Quantification of the intracrystalline acid insoluble compounds recovered from samples of *Neothyris*.

Bed	Age (Ma)	Sample wt (mg)	Insoluble wt (mg)	%insoluble
Recent	0.00	119.3	0.2	0.17
Rapanui Fm	0.20	051.7	0.5	0.97
Tainui Shellbed	0.40	065.6	0.7	1.07
Pinnacle Sand	0.40	234.5	0.9	2.61
Kupe Fm	0.50	58.4	0.8	1.37
Upr Okiwa Group	2.15	54.1	1.0	1.85
Hautawa Shellbed	2.20	65.7	1.5	2.28

Table 9.2: Quantification of the intracrystalline acid insoluble compounds recovered from samples of *Notosaria.*

Bed	Age (Ma)	Sample wt (mg)	Insoluble wt (mg)	%insoluble
Recent	0.00	113.9	0.3	0.26
Rapanui Fm	0.20	85.7	0.8	0.93
Tainui Shellbed	0.40	80.5	1.3	1.61
Pinnacle Sand	0.42	91.6	1.4	1.53
Lr Castlecliff Shb (Waipuka Rd)	0.47	66.7	1.0	1.50
Kupe Fm	0.50	77.5	0.3	0.39
Kaimatira Pumice Sand	0.73	72.7	0.6	0.83
Upper Waipipi Shellbed	2.50	90.2	0.7	0.78

Table 9.3: Quantification of the intracrystalline acid insoluble compounds recovered from samples of *T. sanguinea.*

expressed as %insoluble in order to normalise the effects of different sample size, and vary from 3.33% in *Waltonia* from the Pinnacle Sand to 0.09% from *Neothyris* from the Upper Waipipi Shellbed. When the %insoluble is graphically compared to the age of the samples (**figures 9.1-9.4**), it may be seen that the proportion of the shell mass which is due to the insolubles varies in a way which does not reflect the age of the sample. The data for the period 0.2-0.5 Ma shows some similarity to that of the soluble amino acid composition, with there being a spread of data, but differs in the relationship with further time. *Notosaria* samples show a general non-significant increase over time, although the other samples do not show any relationship of an increasing proportion of insolubles with increasing age.

9.3.2 Amino acid composition of acid insoluble compounds

The amino acid compositions of the samples are tabulated in **Appendix 1**. All insoluble compounds analysed contained amino acids, at a concentration higher than that present in blank sample tubes, analysed to measure the possibility of contamination during sample preparation. It is presumed that the samples are present in the combined state within the insolubles, and that any peptide or other HCl sensitive bonds are broken by the influence of hydrolysis. When the samples are normalised to ng of amino acid per mg of

Bed	Age (Ma)	Sample wt (mg)	Insoluble wt (mg)	%insoluble
Recent	0.00	78.2	0.5	0.64
Rapanui Fm	0.20	78.7	0.5	0.64
Upr Castlecliff Shellbed	0.38	123.7	1.8	1.46
Tainui Shellbed	0.40	96.6	1.4	1.45
Pinnacle Sand	0.42	69.0	2.3	3.33
Lr Castlecliff Shb (Waipuka Rd)	0.47	110.4	0.9	0.82
Lr Castlecliff Shb (Castlecliff)	0.47	116.3	1.1	0.95
Kupe Fm	0.50	117.0	0.8	0.68
Kaimatira Pumice Sand	0.73	93.4	0.5	0.54
Tewkesbury Fm	1.67	79.7	1.3	1.63
Waipuru Shellbed	1.72	101.9	1.1	1.08
Undifferentiated Shb (4243)	1.77	103.1	1.5	1.45
Nukumaru Brown Sand	1.85	133.5	1.0	0.75
Upr Okiwa Group	2.15	81.9	1.4	1.71
Hautawa Shellbed	2.20	54.6	1.4	2.56

Table 9.4: Quantification of the intracrystalline acid insoluble compounds recovered from samples of *Waltonia*.

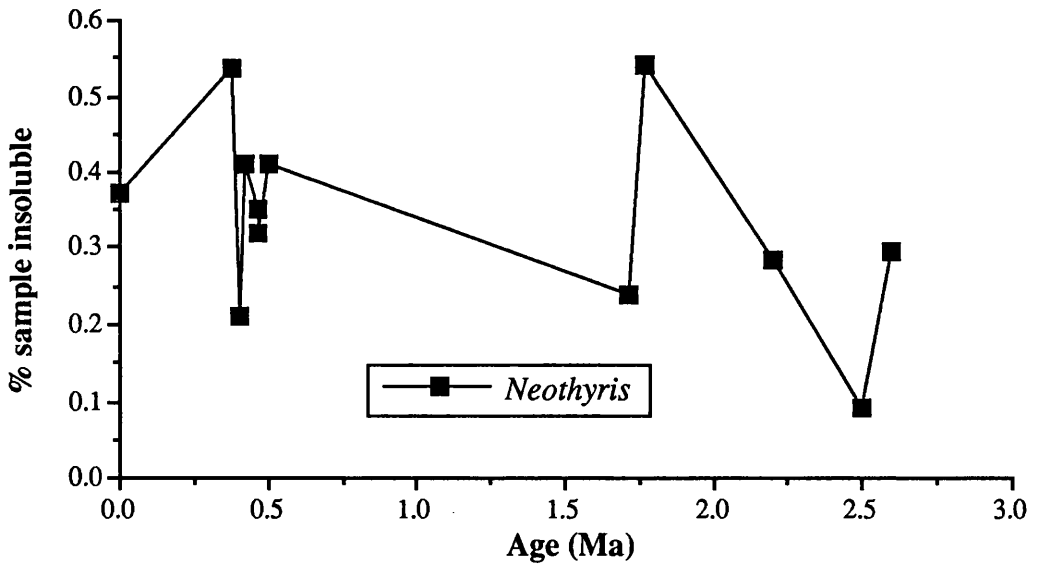


Figure 9.1: Proportion of the sample weight which is due to the acid insoluble compounds in samples of *Neothyris* compared to the age of the sample.

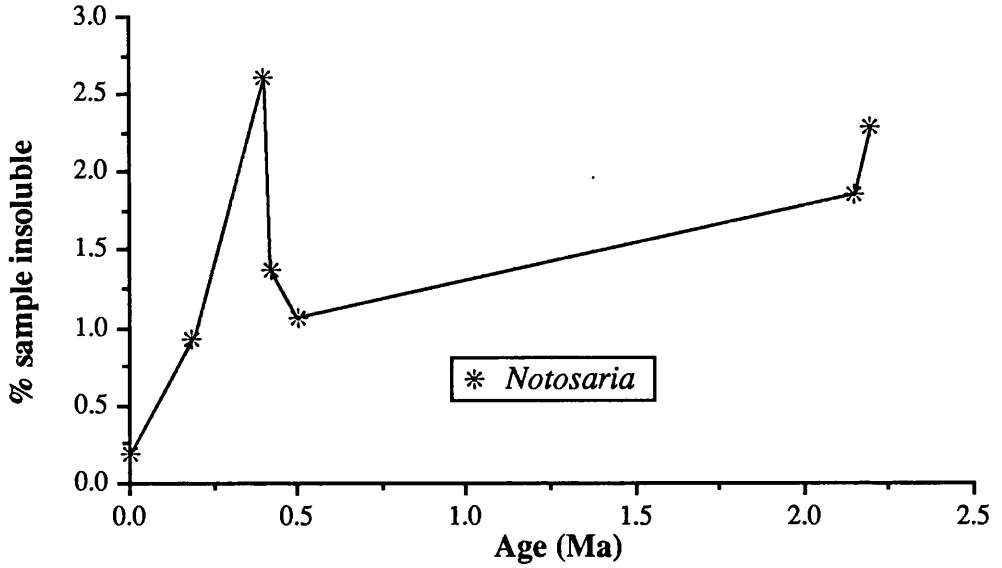


Figure 9.2: Proportion of the sample weight which is due to the acid insoluble compounds in samples of *Notosaria* compared to the age of the sample.

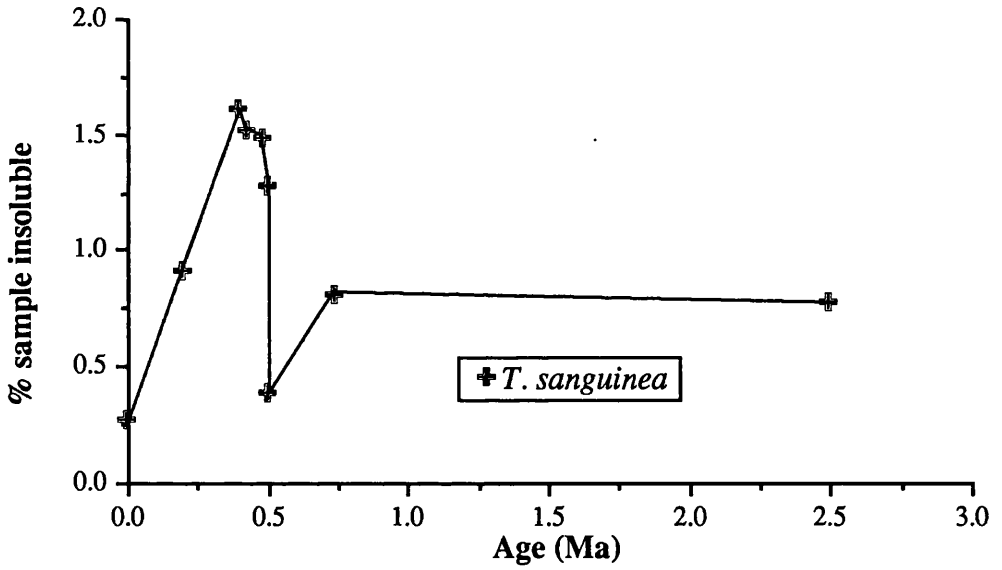


Figure 9.3: Proportion of the sample weight which is due to the acid insoluble compounds in samples of *T. sanguinea* compared to the age of the sample.

shell powder, it can be seen that the amino acids in the insoluble fraction are present at much lower concentrations than those in the soluble fraction, and that there is significant variation in the amino acid content over time, similar to that found in fossil scallops (Akiyama, 1971). At the level of sensitivity studied, there appears to be no correlation between the concentration of amino acid in a sample, and the amount of insoluble compound present. The proportion of the insoluble mass which is accounted for by the amino acids is highly variable (**table 9.5**). In Recent samples, the amino acid analyses indicate a greater mass than indicated by the insoluble weight. This is especially marked

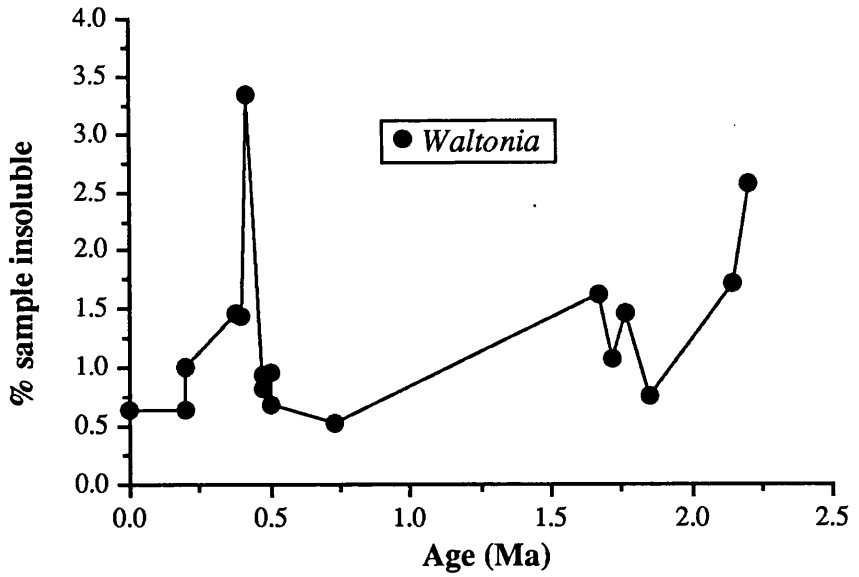


Figure 9.4: Proportion of the sample weight which is due to the acid insoluble compounds in samples of *Waltonia* compared to the age of the sample.

Sample	*Neothyris*	*Notosaria*	*T. sanguinea*	*Waltonia*
Recent	82.737	-	205.640	95.278
Rapanui Fm, Waipipi	56.438	35.116	-	3.735
Rapanui Fm, Waitotara	-	-	92.150	6.028
Upr C'cliff Shb	-	-	-	9.307
Tainui Shb	79.950	19.392	18.811	12.446
Pinnacle Sand	43.310	22.416	6.660	5.022
Lr C'cliff Shb, Waipuka	20.070	-	9.556	4.961
Lr C'cliff Shb, Coast	-	-	-	2.948
Kupe Fm	63.980	-	-	1.604
Kaimatira Pumice Sand	-	-	26.217	2.630
Tewkesbury Fm	-	-	-	1.175
Waipuru Shellbed	123.483	-	-	4.865
Undifferentiated Nukumaru	-	-	12.660	7.738
Nukumaru Brown Sand	-	-	-	2.059
Upr Okiwa	12.986	-	-	5.214
Hautawa	17.950	7.845	-	1.750
Upr Waipipi Shb	-	-	-	20.586

Table 9.5: Calculation of the proportion of the acid insoluble compound which may be accounted for by the recoverable amino acids.

in *Notosaria*, which must be treated with caution as this could be indicative of contamination of the sample during the preparation of the insolubles for amino acid analysis. The figures for the fossil samples indicate that the amino acids often do not account for all of the mass of the insolubles. This shows that either other molecules, such as carbohydrates, are present within this mass, or the Maillard reaction has progressed so far that the amino acids may not be recovered by acid hydrolysis. If the latter case is true, then the insoluble

fraction may not be regarded as being the true fossilised protein (Weiner and Lowenstam, 1980; Akiyama, 1971).

Examination of the data for the individual amino acids in the insoluble fraction shows that all amino acids are present, although their concentration (both relative and absolute) is very much lower than that present in the soluble fraction, where the sample is dominated by the presence of Gly and Ala. The concentration of individual amino acids is high in the Recent sample of *Notosaria*, when compared to both other Recent samples and also to the samples of fossil *Notosaria*.

Amino acids that are unstable in the soluble fraction are also present within the insoluble fraction. When compared to the concentration of these amino acids in the soluble fraction, there initially appears to be selective preservation of Arg and Thr in the insoluble fraction. However, this is not selective preservation, it is a function of the threshold value of the analytical system. Soluble samples tend to lose their unstable amino acids due to their concentration dropping below that of the threshold. By using larger sample size in the insoluble investigation, the concentration is effectively higher. In all other cases of these amino acids in the insolubles, the concentration of these amino acids is lower than their counterparts in the soluble fraction. The concentration of these amino acids is also highly variable over time.

Recent samples, with the exception of *Notosaria* (**Chapter Five**), were not expected to contain a significant amount of insoluble material. However, all Recent samples contain some colourless insoluble material, which yielded significant amounts of amino acid on hydrolysis.

9.4 Discussion

This study has shown that the proportion of insoluble compounds present within the shell does not increase with increasing sample age, but is variable over the time period of the study. The proportion of the sample which is insoluble also varies between the genera of the study. As these compounds are sourced from within the shell of the fossil, it must be assumed that the presence of these compounds is due to reactions occurring within the shell. Proteins, lipids and carbohydrates have been extracted from within brachiopod shells (Cusack *et al.*, 1992; Curry *et al.*, 1991b; Collins *et al.*, 1991a), and it is between these compounds that the degradative reactions will occur. The Maillard Reaction (Maillard, 1913) involves an irreversible reaction between the reactive groups of carbohydrates and amino acids, leading to the formation of melanoidins (for a review of this process, see Furth, 1988). The only published measurement of the intracrystalline neutral sugar content of Recent brachiopods (Collins *et al.*, 1991a) has shown that *N. lenticularis* contains 0.12 ng/mg, and *W. inconspicua* contains 0.21 ng/mg. This is lower than the protein content from a sample of the fraction of the extract with a molecular weight greater than 10 kDa, which, in turn, is lower than that found in the complete protein and

amino acid extract in the present study (**table 5.1**). The limited amount of sugar available within the shell of the brachiopod must limit the production of melanoidins through the Maillard reaction. The reactive group of proteins and amino acids is the free amino group ($-NH_2$), present at the *N*-terminus of the protein/peptide and also in the side chains of the basic amino acids. If there were an excess of the neutral sugars within the shell, then it would be expected that the basic amino acids would preferentially react to form insoluble compounds and thus be completely incorporated into melanoidins. As these amino acids are preserved in the soluble fraction of the fossil record, there cannot be an excess of carbohydrate present.

This is secondarily illustrated by several other features of the insoluble fraction. The %insoluble of the sample does not show an increase with respect to time, as would be expected if there were a simple relationship between the insoluble compounds and the age of the sample, instead the level is variable and does not show this pattern. This would illustrate that once the available carbohydrate has reacted, then the production of the melanoidin ceases and the remaining amino acids and proteins follow different reaction pathways (discussed in **Chapter Seven**). The concentration of the amino acids within the insoluble fraction is low, possibly due to a limited amount of carbohydrate being available for reaction, thus not removing amino acids from the soluble fraction. As the concentration of amino acids within the insolubles does not correspond with the concentration lost from the soluble fraction, this reaction is probably not one of the major pathways for the degradation of brachiopod intracrystalline amino acids. In melanoidins, the amino acid content does not account for the entire weight of the compound.

All amino acids present in the soluble fraction are also represented in the insoluble fraction, showing that there is no preferential reaction between particular amino acids and carbohydrates. The distribution of the amino acids from insolubles also reveals that the most abundant, Gly and Ala, are the also the most abundant in the soluble fraction, although the concentrations are not as high as in that fraction. The samples of *Notosaria* show a concentration of Thr in the insoluble fraction which is consistent with that from other samples, which contrasts with that found in the study of the soluble fraction, where one of the degradation products of Asp, β-alanine, coeluted with Thr, leading to apparent increased concentrations of "Thr". As this is not the case in the insoluble fraction, it may be concluded that the β-alanine is present only in the soluble fraction, or in only minor quantities in the insoluble fraction. Melanoidins form almost immediately on fossilisation (e.g. by the Rapanui Fm (0.2 Ma)). β–alanine was formed mostly in the period 0.2-0.5 Ma (**Chapter Seven**), later than the apparent onset of melanoidin formation. This is consistent with there being a limited amount of carbohydrate present, which has mostly all reacted prior to β-alanine formation.

Recent materials contain significant amounts of amino acid present in HCl (2N) insoluble compounds, in all cases accounting for a high proportion of the mass of the insolubles. The insoluble compound is colourless, and is therefore not characteristic of

melanoidins. It is probable that these compounds are insoluble true proteins, which hydrolyse completely when exposed to 6N HCl (these compounds are also observed when large quantities of shells are processed ready for protein sequencing, M. Cusack, pers. comm.). This insoluble protein fraction from Recent samples has important implications for the study of fossil molecules in the soluble fraction. Natural hydrolysis progresses rapidly in the fossil record, releasing peptide bound amino acids into solution. If this insoluble *protein* fraction (rather than melanoidin) has undergone natural hydrolysis, then amino acids and short peptides will be released into solution by this action. This will effectively increase the concentration of the amino acid in the soluble fraction of younger fossil samples, a feature already recorded from fossil samples (**Chapter Seven**). This will effect the %change of the amino acids, which will explain some of the increases (e.g. in Val, Lys, Gly and Ala) which are higher than may be accounted for by the degradation of other amino acids. This figure should not affect the %free calculations to any great extent. The amino acids released from the insoluble protein fraction will decay in the same way as other amino acids in the soluble fraction, and will also be a potential distortion in the comparison of the Recent and fossil amino acid compositions.

Several authors (e.g. Jope, 1967b; Weiner and Lowenstam, 1980) have found increased concentrations of the unstable amino acids (e.g. Arg, Thr, Ser) in the insoluble fraction, and have suggested that this is due to selective reaction and preservation. This does not appear to have occurred in the samples of the present study, where most of the amino acids are present in similar quantities. Some samples maintain very low concentrations of Arg and Thr, in samples where this amino acid has disappeared from the soluble fraction. It is likely that this "preservation" of these amino acids will be due to larger amounts of shell material being initially utilised, allowing these compounds to be identified. In the soluble fraction, the lower initial sample weight will cause the concentration to fall below the threshold sooner; thus identifying none of the amino acid. The conclusion, however, of Akiyama (1971) that the peptide bond may act to stabilise these unstable amino acids was identified in the soluble fraction of the present study.

It is also interesting to note that fossil molecules extracted from mollusc shells by Weiner and Lowenstam (1980) were completely soluble, whereas those from extant representatives were insoluble in the acid solution. This contrasts with that found in Recent brachiopods in this study, where, while some have a fraction which is insoluble (**Chapter Five**), the majority of the included proteins are soluble, a feature indicated by the concentration of amino acid which they contain.

The variation of the amino acid content of these insoluble compounds makes it highly unlikely that these may be used in phylogenetic studies of brachiopods. Biochemical investigation of these compounds (Furth, 1988) appears to indicate that once the reaction has passed its first stage, then it is irreversible, and it is unlikely to yield amino acids on hydrolysis. Thus, the information contained within these amino acids will be lost.

9.5 Conclusions

This study has shown that fossil samples contain organic matter in the form of melanoidins (formed by the reaction between amino and carbohydrate compounds) within the shell crystallites, and that these compounds contain some amino acids which are recoverable by the use of vapour phase acid hydrolysis, although the mass of the insolubles is not entirely due to the amino acids. In contrast to other studies, compounds recovered from brachiopod shells are not enriched in the unstable amino acids such as Arg, Ser and Thr although these amino acids are present. The proportion of the weight of the sample which is due to the insoluble compounds is variable, and does not show an increase in concentration over time. This may be due to a limiting factor of the concentration of carbohydrate present in the shell of the brachiopod. The concentration of amino acid present in the insoluble fraction is lower than that in the soluble fraction, and also does not correspond with that lost from the soluble fraction.

Recent samples contain insoluble true protein, which solubilises on natural hydrolysis, and which will cause an increase in the concentration of amino acid present in fossil samples (**Chapter Seven**).

For these reasons, it is unlikely that the amino acid content of the insoluble fraction will yield any phylogenetic information regarding the Brachiopoda, and that the Maillard reaction is not a major decomposition pathway in these samples.

Chapter Ten: General discussion and suggestions for further work

General conclusions and suggestions for further work

10.1 Biogeochemistry of intracrystalline proteins

Shells of Recent brachiopods and molluscs contain intracrystalline proteins at a low, but significant, concentration and there are very low concentrations of free amino acids present in the shell (**Chapter Five**; Curry *et al.*, 1991a; 1991b; Addadi *et al*, 1991; Cusack *et al.*, 1992). Once the sample has been fossilised, however, the composition of the intracrystalline organic fraction has changed (**Chapters Six**; **Seven**; **Nine**; Abelson, 1954; 1955; Wyckoff, 1972; Weiner *et al.*, 1976; Weiner and Lowenstam, 1980). Analysis of the shell extract from these fossils reveals that there is little or no true protein remaining in the sample, and that the majority of the amino acids from these proteins are present in the free state in the acid soluble fraction of the organic extract. This indicates that the protein is severely degraded in less than 0.2 Ma, a result that was not expected at the start of this study, as it was hoped that the proteins would be better preserved when present within the protected microenvironment of the shell.

The question remains as to whether the proteins are degrading through the decomposition of the constituent amino acids (whilst remaining peptide bound), or whether the compound is undergoing natural hydrolysis reactions, both of which will release amino acids into the free state. Comparison of the concentration of amino acids in the Recent and young fossil indicates that there is an initial overall increase, although this rapidly drops. The concentrations of the amino acids however, do not appear to fluctuate greatly in the sample. This indicates that the sample is undergoing rapid natural hydrolysis, with the peptide bonds being rapidly cleaved, rather than individual molecules undergoing peptide bound decomposition. This is further supported in the literature by studies of younger material than those examined here (Abelson, 1955; van Kleef *et al.*, 1975; Zeven *et al.*, 1975; Shewry *et al.*, 1982) which have shown that the first stage in protein decomposition is the natural hydrolysis of peptide bonds.

The study of the incarcerated amino acids from the shells of a well dated succession of strata will also elucidate the state of preservation of the individual molecules over time. The concentration of molecules in the fossils shows an overall decrease in time (following the initial increase), which, in most cases, may be described by either a logarithmic or an exponential curve (**Chapter Seven**), leading to possible predictions of the age of samples or the source of derived specimens in shellbeds (**Chapter Eight**). The majority of the amino acids undergo rapid degradation, with c. 80% of the molecules being degraded over the 2.5 Ma period under study. The rate of loss of molecules dramatically slows in older samples, indicating the formation of a relatively stable residue in the soluble fraction.

Some amino acids are less stable than others, a fact which has been highlighted

in pyrolysis experiments of aqueous solutions of pure amino acids (Abelson, 1954; Vallentyne, 1964; 1968) and of shell powders (Jones and Vallentyne, 1960; Totten *et al.*, 1972), and which has been reinforced by the study of natural systems in the present study. The amino acids Arg, Thr, Ser and Tyr are unstable in the fossil record, and rapidly degrade to other compounds. This leads to an overall decrease in the amino acid composition of the sample. The amino acid composition of the acid insoluble fraction (**Chapter Nine**), in contrast to previous studies (e.g. Jope, 1967b; Weiner and Lowenstam, 1980) shows no enrichment of these amino acids, nor of the basic amino acids, which contain a second reactive group, which would be preferentially destroyed in reaction with carbohydrates.

The degradation of the molecules is considered in detail in **Chapter Seven**. The pathways of decay are not obvious for all amino acids, but there are general ways in which decomposition may proceed. The amino acids rapidly undergo decarboxylation and deamidation reactions, which render that amino acid unrecognisable on the amino acid analyser, unless a specific search is made for the reaction product. These and other reactions may result in the formation of non-standard amino acids (e.g. ornithine from Arg), a proteinogenic amino acid (e.g. Ala from Ser) or a non-amino acid (e.g. methylamine from Gly). Any compound which contains an amino group will react with the PITC used in the analysis procedure and will be eluted from the column in the same way as amino acids. These other molecules appear as non-identified peaks on the chromatograms from the analysis system. It is not possible to differentiate these diagenetic molecules from the original molecules, and hence some will show a rapid and large increase over time (e.g. Ala). Some decomposition compounds may co-elute with proteinogenic amino acids (e.g. β-alanine with Thr), and hence increase the apparent concentration of the original molecule.

Pyrolysis experiments have been used as a guide the rate and products of decomposition (e.g. Bada and Man, 1980). The rate of decomposition of amino acids found in the present study is much greater than that found in pyrolysis of pure aqueous solutions of the amino acids (Abelson, 1954; Vallentyne, 1964; 1968), a fact which was noted during the pyrolysis of shell powders (Jones and Vallentyne, 1960; Totten *et al*, 1972). Clearly other factors, such as oxygen (Conway and Libby, 1958), the mixture of the amino acids (Vallentyne, 1964), the presence of metal ions (Ikawa and Snell, 1954) and carbohydrates (Vallentyne, 1964) will also affect the decomposition of the molecules. Pyrolysis of aqueous solutions of the pure compounds should therefore only be used as a guide for decomposition, and more analyses are required to study the decomposition of amino acids in shells, either through the study of natural systems, or through pyrolysis of shell compounds.

The rate of reaction of the molecules increase by an order of magnitude when the temperature is increased from 20° to 40°C (calculated from the Arrhenius equations of Vallentyne, 1964). This indicates that any burial of fossils (and corresponding increase in temperature) will cause a rapid increase in the degradation of the incarcerated

molecules. Thus it is not possible to use an average temperature for molecular decomposition in samples, and unless a detailed burial history for the samples is known, the molecular state of preservation cannot be utilised in dating or other geological uses.

The peptide bond acts to stabilise some of the less stable amino acids. This is demonstrated by the preservation of Ser in the fossils in this study. Older samples do still contain Ser, but this is not present in the free state and remains peptide bound. The relationship of Ser with the peptide bond must be determined by the other residue attached to the bond. The rate of peptide bond decomposition varies between pairs of amino acids (van Kleef *et al.*, 1975; Kahne and Still, 1988), although the rates for all possible pairs have not been studied. The peptide fragments which remain after natural hydrolysis are of unknown length, although this may be constrained in the following way. As the amino acid concentration of the greater than 10 kDa fraction is relatively low when compared to the whole intracrystalline extract, then it is unlikely that there is a high concentration of peptides of a size above that of the cutoff of the filtration system (c. more than 100 amino acids in lengths; **Chapter Six**). In the majority of samples, the proportion of the amino acids present in the free state is c. 70% (compared to c. less than 10% in Recent samples), indicating that most of the peptides have completely degraded. The immune response of antigens from similar samples to those considered in this study (Collins *et al.*, 1991b) is very much reduced, but is still present, indicating that some peptides which include determinants are preserved within the shell. Determinants are usually 4-8 amino acids of a particular protein, and it is these sites (or very similar ones if there is no direct match) to which an antibody, produced in response to the protein, attaches.

Intracrystalline proteins, present within the shells of Recent samples, undergo very rapid natural hydrolysis reactions, releasing amino acids. These molecules then undergo various degradative reactions, which reduce the concentration of that particular amino acid, but may lead to an increase in others, or the production of non-amino acid compounds. Some small peptides are preserved, although these have a range of both molecular weight and chemical activities, and may include immunological determinants. Pyrolysis experiments on aqueous solutions of pure amino acids may only act as a guide for predicting the state of preservation of the fossil molecules, as other factors also need to be taken into consideration.

10.2 Characterisation of fossil peptides

Proteins and amino acids from fossil samples are highly degraded (**Chapters Six and Seven**). As a result, the separation of these molecules by conventional biochemical studies is not possible due to the range of molecular weights and chemical activities of the remaining peptides. A similar result was obtained independently by Collins *et al.* (1991a) from similar samples to those analysed in the present study. Endo (1992) met with some success in fractionating peptides from fossil brachiopods, but also found that the peaks in

these samples were wide, and of variable height when compared to those of the Recent samples.

As it has not proved possible to fractionate the crude organic extract from the shells, it is not likely that a peptide will be purified to homogeneity and characterised by peptide sequencing. This was the primary goal of molecular palaeontology (Curry, 1988), although this study has shown that it is not likely to be possible, even for molecules from the relatively protected sites of the intracrystalline molecules. The only realistic method for characterisation of the protein fragments is therefore via immunology. In previous studies (Westbroek, 1979; Collins *et al.*, 1991b; Endo, 1992), antibodies are produced against the crude shell extract, which contain a range of proteins and also carbohydrates, both of which may be antigenic. This is a crude, indirect, method of analysis, as the antibody may be specific against the carbohydrate or any one of the proteins. It is also highly likely that antibodies will be produces against determinants on the surface of a protein, rather than in adjacent site in the primary sequence of the protein. If the proteins are separated from each other and from the carbohydrates, and denatured prior to producing antibodies, then better characterised antibodies will be produced against a molecule with a higher degree of similarity to that of the fossil peptide. This will therefore allow a level of characterisation of the fossil protein.

10.3 Amino acid taxonomy

The underlying aim of all palaeontology is to elucidate the maximum amount of information regarding the phylogenetic relationships of fossil organisms. This is also one of the aims of molecular palaeontology. This study has shown that it is possible to group Recent species in terms of their amino acid composition alone (**Chapter Five**), extracted from very small samples of shell powder, in contrast to other molecular studies which require large amounts of material (**Chapter Three**). The new method used in this study utilises statistical manipulation of the amino acid data in order to summarise the vast amount of information produced by the fourteen amino acid variables in the sample. Graphical representation of the data produced in this way groups samples in terms of the variability of the datasets, hence it would be expected that samples which have a similar amino acid composition will be taxonomically closely related. This is indeed the case, with samples of the same genera or family plotting closely together, and those of different orders plotting very far apart. This is a direct measurement of the difference between the amino acids of the samples and hence between the genetics of the organisms (**Chapters Two and Five**).

Recent samples were analysed in order to be certain that the method could be used to discriminate between the organic content of the samples, and that the conclusions were consistent both with that of morphological studies (Williams *et al.*, 1965) and of immunological investigations (Collins *et al.*, 1988; Endo, 1992). As this was the case, it

was assumed that the conclusions reached by this method were valid, provided that the stipulated level of variation (> 75%, Sneath and Sokal, 1973) was present in the latent vectors. Statistical analysis of the normalised concentrations of amino acid present in the samples (**Chapter Seven**) have shown that fossil samples may also be grouped in the same way as the Recent samples. However, in the case of the fossil molecules, the sensitivity of the method is reduced in older samples. In the Rapanui Fm (0.2 Ma) the samples are well separated to the subordinal level. By the Lower Castlecliff Shellbed (0.45 Ma), the samples have begun to merge, with samples of the same order plotting closely together, and overlapping. This is mirrored in the cluster diagrams, and represents the decay of the unstable amino acids to form a very closely related residue of the stable amino acids. Samples which had very different distribution of original amino acid concentrations (i.e. different order or phylum) have a residue which is still very different to that of the other orders.

The present study, although considering a different fossil group, has enlarged previous work into amino acid taxonomy of foraminifera (King and Hare, 1972; Haugen *et al.*, 1989) and the immunological investigation of Collins *et al.* (1991a). The fossil samples have a sufficiently different amino acid signature allowing separation to be completed to at least the ordinal level in all cases. This will be especially important in the consideration of problematica, provided that a large enough sample size (in terms of range of other species for comparison) could be used to provide the amino acids for study. Evolutionary studies are possible by following the amino acid signature over time, although care must be taken to recognise differences due to amino acid degradation.

10.4 Geological applications of amino acids

The term 'Molecular Palaeontology' (Calvin, 1968) was originally coined to describe the study of degraded molecules present in oils, a field which has undergone rapid growth and has made major contributions to geology through organic geochemistry. The term was redefined by Runnegar (1986), and more emphasis placed on the taxonomic applications of the organic molecules from fossils and Recent samples, although the study of molecules to reveal geological information was still considered. In the present study, the main aims were to study the taxonomic applications of the molecules, although others were considered (**Chapter Eight**).

As discussed above, specific and generic classifications are unlikely to be possible using the amino acid composition alone. Immunological investigations using antibodies prepared against Recent proteins, or using a combination of several different techniques (Endo and Curry, 1990; Endo, 1992) may reveal specific information. Information concerning taxonomy, phylogeny and evolution from the direct study of proteins and other molecules is likely to come from molecules extracted from Recent shells (Curry *et al.*, 1991a; 1991b; Cusack *et al.*, 1992), rather than from fossils.

Fossil molecules are likely to have a role in other geological applications. Some of these, such as dating using the amino acid racemisation reaction have already been well documented (Schroeder and Bada, 1976), whereas dating by amino acid decomposition (**Chapter Eight**) has not been considered prior to this study. The existence of amino acid molecules within sediments has been recognised by organic geochemists, but has not received much investigation from sedimentologists. Amino acids are present in the core of oolites (Mitterer, 1968) and also affect the crystallisation of polymorphic calcium carbonate (Kitano and Hood, 1965), and it is likely that the influence of these molecules on the sediments could be of prime importance, discussed further below.

10.5 Suggestions for further work

Proteins within the shells of fossil are rapidly degraded to smaller peptides and free amino acids, and these degraded molecules cannot be separated by conventional biochemical techniques (**Chapters Six, Seven and Eight**). It is unlikely that these fossil peptides will therefore be purified to a sufficient degree to allow sequencing. The present study has suggested that such attempts would be expensive and would not give high levels of information regarding the taxonomy of the organism. Sequencing of the proteins or DNA of an organism should therefore be confined to those from Recent samples, and to infer rates of evolution and divergence from these molecules (the molecular fossils of Runnegar, 1986).

Taxonomy of fossil organisms is likely to be accomplished by the application of immunology or from amino acid analysis. Recent attempts to purify a fossil protein to homogeneity for immunisation of rabbits (Endo, 1992) did not prove successful, and it was concluded that this method would be time consuming and difficult to achieve. Improvements in preparative techniques have made it possible to fractionate Recent intracrystalline proteins from brachiopods (Curry *et al.*, 1991a; Cusack *et al.*, 1992), although immunisation using these purified proteins has not yet been attempted. Antibodies prepared against the range of proteins within the shell will give a better resolution of the immunological method. If proteins are unravelled before immunisation, antibodies would be prepared against the primary sequence, and this is the form that peptides are likely to be preserved in the fossil record.

Amino acid analysis has revealed the extent of decomposition of the proteins and amino acids within the shell. The molecular state of preservation is not as high as was hoped, with degradation proceeding unhindered. However, use of the intracrystalline fraction alone ensures that the degradation products remain within the shell, whereas intercrystalline molecules will be successively leached from the shell by the percolating effects of groundwater. This indicates that most of the information preserved in the shell will be sampled, in a way not possible in immunology or peptide sequencing. The methods outlined above (**Chapters Three, Five, Seven and Eight**) could be adapted to use higher

concentrations of HCl to dissolve the shell (e.g. 6N or higher) which will increase the effective loading of the sample onto the analysis system, allowing quantification of molecules which were at low concentrations in the present study. As the proteins have undergone extensive natural hydrolysis, the effect of the acid on the remaining peptide bonds is not important. As the amino acid composition reveals a 'signature' for organisms which is characteristic to at least the ordinal level, this method will be useful in the ordinal classification of problematica by amino acid composition.

Molecular palaeontology should be concerned with other areas in addition to taxonomy. This study has shown that fossil molecules may be used for a range of dating techniques and also for correlation of samples. This role should be further explored and samples relevant to basinal correlation further examined. Perhaps one of the most exciting of the uses of intracrystalline fossil amino acids currently under investigation is the assessment of the environment of formation of the molecules by the use of isotopes. As the molecules have been entombed within the shell since biomineralization, the isotopic signal of the carbon and oxygen in the samples will be sealed, and hence uncontaminated by percolating groundwaters etc.

It is likely that the emphasis of molecular palaeontology will move from the fossil to Recent samples, where the information rich molecules are better preserved. Sequencing of the proteins from within the shells will allow evolutionary assessments, but will also direct investigation of the role of organics into the biomineralization of the shell. Sequencing of the genomic DNA will give a direct measure of the genetic variability of samples of the same species from different localities, which will prove useful as a guide to examine how the ecophenotypic variation in samples reflects genetics, i.e. how much genetic variation there is in samples which appear to be morphometrically very different. This will be of immeasurable value as an aid to recognising which fossil samples are likely to be different species, or which are the same species with higher ecophenotipic variation.

Although much of this thesis has shown how poorly preserved the amino acids in fossil samples are, it has also been able to show where problems lie. It is highly unlikely that it will be possible to routinely sequence intracrystalline peptides or DNA from fossils, as the molecules are too highly degraded, and the quantity of shell material required too high. However, molecular palaeontology has a vital role to play via the roles outlined above and to some extent from the general conclusions of this thesis. Increased emphasis should be placed on molecules from Recent samples and in the interaction of organic molecules with sediments. In this way, problems pertaining to biomineralization, evolution and ecophenotypic variability in organisms , and the organic component of sediments may be examined in a new light.

"Now I am free, enfranchised and at large"

William Wordsworth, *The Prelude*

Appendix One: Amino acid tables

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	6.76	8.35	3.54	56.06	3.15	2.69	5.18	7.74	1.39	6.16	1.78	2.92	1.70	1.47
Rapanui Formation, Waipipi	0.78	2.24	0.48	21.28	0.00	0.06	6.23	1.34	0.21	2.14	2.41	0.77	0.39	0.44
?Rapanui formation, Castlecliff	3.40	6.22	0.45	42.33	0.71	0.36	12.45	4.18	2.09	5.07	1.82	1.85	0.89	0.57
Landguard Sand	5.17	9.54	0.83	49.89	1.49	1.11	19.89	7.19	3.17	8.00	2.33	3.09	1.44	1.29
Upper Castlecliff Shellbed	4.21	6.82	0.52	57.58	1.12	0.73	15.95	6.13	2.59	6.78	1.45	2.28	1.54	0.98
Tainui Shellbed	4.46	7.88	0.39	52.19	1.01	0.48	17.44	7.1	3.88	8.13	2.22	2.75	1.32	1.16
Pinnacle Sand	3.83	6.3	0.44	45.89	1.17	0.49	16.35	7.13	2.24	8.21	2.64	3.08	1.55	2.07
Lr Castlecliff Shellbed,Castlecliff	2.64	5.01	0.40	30.98	0.68	0.40	10.96	4.41	2.20	4.54	1.27	1.69	1.01	0.63
Lr Castlecliff Shellbed, Waipuka	3.20	6.49	0.45	37.39	0.63	0.27	11.64	5.14	1.70	5.40	1.51	1.86	1.40	0.85
Kupe Formation	3.07	5.48	0.71	39.38	0.67	0.00	11.09	4.07	2.22	4.32	1.03	1.61	0.94	0.56
Waipuru Shellbed	1.20	3.86	1.40	34.78	0.00	0.00	10.41	2.71	0.74	4.04	0.88	0.93	0.37	0.75
Upper Nukumaru group (4243)	0.97	3.15	0.59	32.38	0.00	0.00	9.80	1.48	0.00	4.20	0.61	1.10	0.55	0.36
Upper Okiwa Group (4357)	0.87	2.34	0.69	16.24	0.00	0.00	6.68	1.66	0.00	2.59	0.59	0.58	0.00	0.00
Hautawa Shellbed	1.32	1.49	0.79	11.69	0.00	0.00	4.15	1.10	0.00	1.89	0.44	0.00	0.00	0.35
Te Rama Shellbed	1.09	3.42	0.75	27.82	0.00	0.00	9.25	3.21	0.98	3.71	0.70	1.21	0.83	0.55
Upper Waipipi Shellbed	1.03	2.56	0.00	24.52	0.00	0.00	6.86	1.80	0.00	2.71	0.72	0.56	0.00	0.00
Middle Waipipi Shellbed	0.96	3.67	0.67	40.16	0.00	0.00	12.85	3.30	1.11	4.80	0.67	1.22	0.74	0.53

Appendix 1.1: *Neothyris* free and combined amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	214.52	14.35	11.81	218.20	9.25	5.89	16.08	14.96	43.22	7.90	2.81	3.96	13.45	4.39
Rapanui Marine Sand (Waipipi)	109.54	25.12	5.10	165.49	9.87	3.92	36.28	19.51	34.44	16.17	5.78	10.16	15.96	7.31
?Rapanui Formation (Castlecliff)	72.12	17.15	0.72	134.75	1.74	69.21	59.80	9.88	10.45	8.68	3.29	3.94	6.76	3.08
Tainui Shellbed	42.58	12.79	0.55	107.52	2.44	53.78	49.68	10.60	6.69	12.28	3.43	4.14	7.88	4.12
Pinnacle Sand	41.72	12.73	0.58	96.27	2.29	50.91	47.57	10.64	12.78	12.72	2.99	4.12	11.78	4.30
Upper Okiwa Group (4357)	1.98	2.96	0.75	40.09	0.00	22.29	23.38	1.17	0.81	3.36	1.25	0.67	1.15	0.86
Hautawa Shellbed	1.61	2.52	0.67	29.03	0.00	23.27	23.96	1.29	0.79	3.11	0.87	0.78	0.88	0.49

Appendix 1.2: *Notosaria* free and combined amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	12.12	13.51	7.66	69.70	6.28	6.58	12.24	17.32	5.42	12.82	7.39	7.69	4.15	7.42
Rapanui Marine Sand (Waitotara)	8.52	13.41	2.60	101.54	2.20	0.56	29.34	13.50	5.58	14.21	5.48	6.21	3.26	2.99
?Rapanui Formation (Castlecliff)	6.41	10.49	1.14	102.97	1.53	0.42	30.99	13.33	3.62	14.20	5.12	4.77	3.54	7.30
Tainui Shellbed	8.06	10.33	1.09	81.32	1.44	0.98	25.08	14.69	2.95	14.95	5.26	4.78	2.87	2.86
Pinnacle Sand	9.58	15.25	1.68	103.44	2.60	1.21	35.94	17.69	6.42	19.49	6.26	6.93	3.70	6.35
Lr Castlecliff Shellbed (Castlecliff)	9.67	12.56	0.84	97.46	0.92	0.55	27.17	12.52	3.55	13.91	4.20	4.49	2.59	2.69
Lr Castlecliff Shellbed (Waipuka)	7.78	13.02	0.74	103.31	1.04	0.77	34.09	18.07	3.94	18.97	5.34	6.52	3.43	4.41
Kupe Formation	9.15	12.52	0.80	115.37	1.13	0.60	32.26	14.42	3.82	13.42	3.26	4.22	2.66	1.72
Kaimatira Pumice Sand	5.11	7.40	0.70	88.67	0.44	0.84	19.09	8.64	1.26	8.56	1.97	2.00	1.62	0.83
Okehu Shell Grit	7.59	9.96	1.79	94.68	1.18	0.98	24.95	11.30	1.43	11.17	2.61	2.85	2.58	1.36
Upper Waipipi Shellbed	1.66	4.31	0.83	57.10	0.99	0.92	15.08	4.95	0.66	6.04	1.60	0.96	0.87	0.62

Appendix 1.3: *T. sanguinea* free and combined amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	8.23	9.82	6.94	76.76	6.70	3.74	11.66	12.85	2.94	10.74	4.31	3.96	2.22	3.58
Rapanui Marine Sand (Waipipi)	13.06	20.67	4.78	124.02	4.22	2.79	31.47	20.45	3.10	13.67	6.47	5.98	5.03	8.74
Rapanui Marine Sand (Waitotara)	9.17	12.68	2.02	103.99	1.64	0.61	26.17	10.16	5.67	12.76	4.70	5.37	2.71	2.44
?Rapanui Formation (Castlecliff)	6.87	11.62	0.48	98.11	1.66	0.70	28.32	11.69	3.71	13.70	3.83	4.23	2.97	1.84
Upper Castlecliff Shellbed	9.22	15.42	1.56	127.24	1.76	0.89	37.65	14.93	5.82	17.74	5.10	6.02	2.88	5.18
Tainui Shellbed	9.17	14.96	1.13	109.61	2.73	1.20	34.43	15.51	3.99	19.35	5.29	6.08	3.06	4.79
Pinnacle Sand	6.27	10.91	0.58	81.83	1.56	0.64	24.00	10.57	4.03	13.76	3.75	3.90	1.82	1.94
Lr Castlecliff Shellbed (Castlecliff)	8.68	15.10	1.06	116.67	1.70	0.53	29.32	14.39	4.11	14.43	4.06	4.74	2.74	2.55
Lr Castlecliff Shellbed (Waipuka)	6.07	11.24	0.70	89.05	0.96	0.65	27.61	11.68	4.17	14.41	3.82	4.62	1.90	3.23
Kupe Formation	10.64	18.34	1.01	130.96	1.59	0.26	40.52	15.98	5.43	18.02	3.71	6.70	3.46	3.53
Kaimatira Pumice Sand	4.88	6.85	0.74	81.13	0.48	1.25	18.37	5.58	1.70	7.37	1.52	2.03	1.81	0.75
Tewksbury Formation	0.99	5.49	0.65	58.25	0.75	0.63	17.20	5.06	0.92	7.65	1.44	1.62	0.56	0.71
Waipuru Shellbed	1.42	6.07	0.68	53.88	0.00	0.84	17.81	5.83	0.91	8.34	1.94	1.88	0.92	0.96
Upper Nukumaru Group (4243)	1.99	6.84	0.66	65.97	0.00	0.57	23.04	5.83	1.03	9.94	1.99	2.61	0.98	1.65
Nukumaru Brown Sand	3.39	7.56	0.72	91.67	1.41	0.00	24.70	7.28	0.97	10.82	2.39	2.19	0.76	1.08
Upper Okiwa Group (4357)	0.81	3.70	0.84	41.44	0.00	0.00	11.84	3.12	0.00	4.53	1.24	0.93	0.89	0.59
Hautawa Shellbed	1.05	2.85	0.83	30.56	0.00	0.00	9.23	2.93	0.00	4.30	1.24	0.68	0.60	0.50

Appendix 1.4: *Waltonia* free and combined amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	0.66	0.00	0.00	2.83	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.47	0.00	0.32
Rapanui Formation, Waipipi	1.03	0.70	0.00	17.67	0.00	0.00	5.69	1.42	0.40	2.24	0.85	0.63	0.32	0.00
?Rapanui formation, Castlecliff	2.77	1.48	0.26	37.48	0.62	0.22	10.12	4.35	1.93	4.42	1.11	1.42	0.80	0.37
Landguard Sand	4.42	2.32	0.00	44.95	1.01	0.00	14.87	6.65	3.00	6.54	1.72	2.63	1.14	1.15
Upper Castlecliff Shellbed	4.10	2.01	0.00	49.05	0.94	0.45	12.68	5.47	2.79	5.53	1.29	2.10	1.11	1.04
Tainui Shellbed	3.50	1.69	0.00	39.57	0.57	0.38	12.26	5.54	2.41	6.06	1.63	2.01	1.73	0.63
Pinnacle Sand	3.73	1.81	0.00	47.28	0.73	0.39	15.67	6.73	3.45	7.61	2.29	3.04	1.08	1.70
Lr Castlecliff Shellbed,Castlecliff	2.84	1.49	0.00	39.34	0.66	0.46	11.34	5.09	1.87	5.03	1.03	1.60	1.46	0.60
Lr Castlecliff Shellbed, Waipuka	2.89	1.43	0.00	33.90	0.52	0.29	9.42	4.43	0.30	4.46	1.06	1.81	0.87	0.40
Kupe Formation	4.06	1.63	0.00	43.02	0.79	0.29	10.70	3.87	2.03	4.20	0.85	1.54	1.04	0.36
Waipuru Shellbed	0.66	0.83	0.00	24.79	0.00	1.23	8.31	1.83	0.00	2.95	0.45	0.64	0.00	0.46
Upper Nukumaru group (4243)	0.57	1.16	0.00	36.65	0.00	0.00	11.39	2.38	0.48	3.90	0.77	0.84	0.43	0.00
Upper Okiwa Group (4357)	0.00	0.75	0.00	14.58	0.00	0.00	7.83	1.54	0.00	2.37	0.51	0.59	0.00	0.43
Hautawa Shellbed	0.00	0.00	0.00	9.67	0.00	0.00	3.71	0.95	0.00	1.50	0.47	0.61	0.00	0.30
Te Rama Shellbed	1.32	0.94	0.00	25.86	0.00	0.00	10.09	2.66	0.61	3.25	0.00	0.91	0.61	0.47
Upper Waipipi Shellbed	0.60	0.00	0.00	18.02	0.00	0.00	5.93	1.46	0.00	2.22	0.51	0.00	0.00	0.48
Middle Waipipi Shellbed	1.04	1.06	0.00	36.06	0.00	0.00	11.92	2.69	0.72	4.08	0.50	0.97	0.64	0.44

Appendix 1.5: *Neothyris* free amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	0.88	0.00	0.00	2.26	0.00	0.00	0.36	0.41	32.37	0.25	0.00	0.39	0.54	0.00
Rapanui Marine Sand (Waipipi)	121.94	5.59	1.82	167.43	5.26	8.52	35.85	16.84	48.90	11.63	3.96	6.60	11.28	4.03
?Rapanui Formation (Castlecliff)	72.26	3.45	0.00	134.22	1.62	81.00	61.44	9.58	12.60	8.69	1.89	3.14	4.74	1.77
Tainui Shellbed	46.80	3.68	0.70	108.92	1.44	57.05	52.54	9.87	9.70	11.29	4.14	3.96	9.17	3.24
Pinnacle Sand	41.53	3.50	0.00	95.42	1.10	53.56	45.37	9.27	14.68	10.41	2.75	3.45	8.28	2.72
Upper Okiwa Group (4357)	1.83	0.89	0.00	32.96	0.00	20.07	20.98	0.98	0.94	3.05	0.94	0.00	0.79	0.51
Hautawa Shellbed	1.51	0.70	0.00	22.25	0.00	21.55	21.38	1.27	1.34	2.72	0.83	0.66	0.62	0.34

Appendix 1.6: *Notosaria* free amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	0.66	0.00	0.65	3.65	0.00	0.00	0.52	0.77	0.00	0.00	0.00	0.86	0.63	0.00
Rapanui Marine Sand (Waitotara)	5.05	2.11	0.30	66.31	0.83	0.48	20.21	9.37	2.73	9.16	2.22	2.41	1.19	0.76
?Rapanui Formation (Castlecliff)	5.24	2.56	0.45	80.09	0.90	0.48	22.28	9.84	3.06	9.78	2.39	2.63	1.70	0.89
Tainui Shellbed	5.88	2.49	0.35	68.70	0.81	0.65	22.08	13.56	3.57	13.02	3.67	4.03	2.47	2.03
Pinnacle Sand	7.95	3.51	0.93	87.23	1.38	0.78	29.85	14.89	5.29	15.55	4.44	6.07	2.43	4.07
Lr Castlecliff Shellbed (Castlecliff)	4.78	2.02	0.00	62.46	0.73	0.60	18.92	8.75	2.10	9.06	2.28	3.00	1.52	1.28
Lr Castlecliff Shellbed (Waipuka)	7.70	3.33	0.34	98.57	1.02	1.08	32.83	15.77	4.61	16.57	4.45	6.09	2.65	3.08
Kupe Formation	7.77	3.20	0.00	106.06	1.21	0.91	29.68	11.97	3.43	11.18	2.53	3.49	2.48	1.21
Okehu Shell Grit	6.77	2.89	0.00	93.69	0.82	0.80	23.07	11.03	2.16	10.32	2.13	2.77	1.77	1.10
Upper Waipipi Shellbed	1.61	1.05	0.00	48.12	0.00	0.82	13.87	4.18	0.00	4.93	1.33	0.84	0.94	0.62

Appendix 1.7: *T. sanguinea* free amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	0.00	0.00	0.00	1.43	0.00	0.00	0.00	0.00	0.00	0.31	0.00	0.30	0.00	0.00
Rapanui Marine Sand (Waipipi)	8.42	4.45	2.80	106.27	3.15	2.14	25.77	16.11	1.98	11.67	3.30	6.32	3.67	6.03
Rapanui Marine Sand (Waitotara)	6.16	2.75	0.53	94.57	0.90	0.68	23.91	9.36	3.91	10.65	2.06	2.40	1.80	0.64
?Rapanui Formation (Castlecliff)	8.28	3.47	0.40	99.76	1.30	0.81	27.98	11.24	4.47	12.83	2.91	3.95	2.46	1.59
Upper Castlecliff Shellbed	7.61	3.74	0.70	100.06	1.76	1.33	31.72	12.42	6.81	14.64	3.71	5.09	2.83	3.55
Tainui Shellbed	6.05	2.51	0.43	69.43	0.99	0.70	21.55	9.83	4.58	12.65	4.21	4.61	3.30	2.23
Pinnacle Sand	5.18	2.72	0.27	82.34	1.07	0.50	22.32	9.95	5.37	11.64	3.32	3.37	1.29	1.38
Lr Castlecliff Shellbed (Castlecliff)	6.31	3.03	0.00	90.15	1.43	0.73	24.20	10.54	4.88	12.06	2.68	3.93	2.33	1.59
Lr Castlecliff Shellbed (Waipuka)	7.16	3.69	0.27	107.14	1.26	1.23	31.08	11.96	4.84	14.75	3.06	4.80	2.33	2.50
Kupe Formation	9.79	5.09	0.40	118.44	1.32	0.72	34.47	14.14	5.93	15.84	3.35	5.75	3.24	2.70
Tewksbury Formation	1.37	1.68	0.00	55.04	0.00	0.00	16.99	4.88	1.03	7.01	1.43	2.21	0.54	0.77
Waipuru Shellbed	1.58	1.92	0.00	53.58	0.00	0.73	18.31	6.27	1.13	6.76	1.55	1.53	1.02	0.89
Upper Nukumaru Group (4243)	1.34	1.85	0.00	57.17	0.00	0.00	20.66	5.41	0.86	8.86	1.87	2.15	0.80	1.36
Nukumaru Brown Sand	3.45	2.25	0.00	80.35	1.10	0.00	21.65	6.15	1.58	9.23	1.76	2.14	2.54	1.14
Upper Okiwa Group (4357)	0.00	0.88	0.00	30.70	0.00	0.00	9.47	2.45	0.00	3.41	0.86	0.59	0.63	0.45
Hautawa Shellbed	0.00	0.74	0.00	24.57	0.00	0.00	8.19	2.60	0.00	3.39	0.74	0.66	0.50	0.52

Appendix 1.8: *Waltonia* free amino acid concentration (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	0.20	0.35	0.23	0.38	0.20	0.13	0.26	0.28	0.04	0.22	0.15	0.28	0.11	0.22
Rapanui Formation, Waipipi Beach	0.13	0.26	0.21	0.36	0.11	0.10	0.21	0.03	0.19	0.07	0.14	0.05	0.06	0.12
?Rapanui formation, Castlecliff Beach	0.14	0.29	0.21	0.38	0.16	0.13	0.47	0.17	0.00	0.45	0.29	0.36	0.18	0.21
Upper Castlecliff Shellbed	0.29	0.57	0.38	0.85	0.34	0.35	0.94	0.27	0.03	0.68	0.43	0.57	0.22	0.47
Tainui Shellbed	0.05	0.09	0.10	0.21	0.12	0.08	0.25	0.09	0.03	0.18	0.11	0.18	0.08	0.12
Pinnacle Sand	0.05	0.07	0.05	0.17	0.07	0.17	0.26	0.19	0.07	0.23	0.15	0.20	0.04	0.06
Lower Castlecliff Shellbed,Castlecliff	0.23	0.34	0.22	0.82	0.26	0.23	0.76	0.26	0.00	0.49	0.26	0.34	0.16	0.29
Lower Castlecliff Shellbed, Waipuka	0.04	0.07	0.07	0.12	0.03	0.03	0.07	0.10	0.00	0.05	0.03	0.05	0.02	0.04
Waipuru Shellbed	0.21	0.36	0.29	0.64	0.17	0.14	0.30	0.08	0.00	0.19	0.12	0.18	0.09	0.18
Upper Nukumaru Group (4243)	0.04	0.07	0.10	0.18	0.04	0.03	0.04	0.00	0.03	0.03	0.02	0.05	0.02	0.04
Hautawa Shellbed	0.02	0.05	0.03	0.09	0.02	0.02	0.05	0.02	0.00	0.04	0.04	0.06	0.03	0.03
Te Rama Shellbed	0.05	0.11	0.12	0.22	0.06	0.04	0.07	0.02	0.00	0.05	0.04	0.08	0.04	0.05
Upper Waipipi Shellbed	1.12	1.03	0.55	2.52	0.72	0.53	1.08	0.57	0.10	0.63	0.35	0.60	0.71	0.49
Middle Waipipi Shellbed	0.66	1.46	1.19	2.26	0.99	0.60	1.04	0.39	0.04	0.93	0.80	1.32	0.59	0.94

Appendix 1.9: Amino acid concentration of insolubles from *Neothyris* (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	23.05	34.26	24.53	6.39	32.84	21.54	6.94	8.57	16.67	19.61	13.22	19.10	10.18	0.00
Rapanui formation, Waipipi Beach	0.32	0.34	0.20	0.80	0.20	0.15	0.38	0.08	0.00	0.27	0.14	0.20	0.12	0.19
?Rapanui formation, Castlecliff Beach	0.09	0.18	0.12	0.46	0.17	0.08	0.31	0.40	0.00	0.22	0.16	0.26	0.13	0.13
Tainui Shellbed	0.16	0.32	0.23	0.81	0.20	0.25	0.71	0.51	0.07	0.75	0.45	0.58	0.26	0.25
Pinnacle Sand	0.16	0.24	0.19	0.65	0.15	0.09	0.33	0.18	0.05	0.27	0.19	0.27	0.14	0.17
Upper Okiwa Group (4357)	0.09	0.19	0.15	0.55	0.10	0.10	0.22	0.20	0.06	0.23	0.15	0.19	0.06	0.11
Hautawa Shellbed	0.06	0.13	0.10	0.27	0.08	0.17	0.16	0.20	0.11	0.17	0.12	0.15	0.03	0.05

Appendix 1.10: Amino acid concentration of insolubles from *Notosaria* (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	0.61	0.63	0.37	0.84	0.34	0.34	0.53	0.29	0.05	0.34	0.22	0.32	0.19	0.35
Rapanui Formation, Waipipi Beach	0.29	0.92	0.50	1.02	0.41	0.42	1.15	0.44	0.00	1.02	0.60	0.85	0.27	0.73
?Rapanui Formation, Castlecliff	0.13	0.28	0.13	0.46	0.14	0.30	0.54	0.47	0.06	0.49	0.30	0.35	0.12	0.15
Tainui Shellbed	0.08	0.17	0.11	0.34	0.16	0.16	0.27	0.60	0.00	0.32	0.23	0.34	0.16	0.09
Pinnacle Sand	0.04	0.07	0.04	0.13	0.04	0.07	0.08	0.13	0.10	0.11	0.08	0.09	0.00	0.03
Lower Castlecliff Shellbed,Castlecliff	0.15	0.28	0.18	0.51	0.26	0.55	0.58	0.86	0.03	0.52	0.35	0.49	0.29	0.16
Lower Castlecliff Shellbed, Waipuka	0.05	0.09	0.07	0.27	0.05	0.11	0.12	0.10	0.03	0.16	0.12	0.14	0.06	0.05
Kupe Formation	0.07	0.18	0.18	0.39	0.13	0.08	0.26	0.24	0.04	0.30	0.17	0.27	0.10	0.08
Kaimatira Pumice Sand	0.08	0.15	0.08	0.32	0.07	0.10	0.26	0.21	0.02	0.22	0.20	0.23	0.17	0.05
Upper Waipipi Shellbed	0.03	0.08	0.07	0.17	0.10	0.22	0.13	0.23	0.02	0.13	0.12	0.17	0.08	0.05

Appendix 1.11: Amino acid concentration of insolubles from *T. sanguinea* (ng/mg)

Sample	D/N	E/Q	S	G	R	T	A	P	Y	V	I	L	F	K
Recent	3.40	3.33	2.95	17.11	3.10	2.30	7.88	3.42	0.11	4.05	2.20	2.15	1.35	2.10
Rapanui Formation, Waipipi Beach	0.22	0.29	0.14	0.35	0.13	0.11	0.23	0.05	0.00	0.21	0.10	0.15	0.07	0.11
Rapanui Formation, Waitotara	0.37	0.63	0.28	0.58	0.34	0.26	0.54	0.25	0.00	0.36	0.21	0.36	0.17	0.28
?Rapanui formation, Castlecliff Beach	0.25	0.57	0.31	1.12	0.29	0.40	1.11	0.67	0.05	1.00	0.62	0.77	0.32	0.37
Upper Castlecliff Shellbed	0.05	0.11	0.08	0.47	0.10	0.27	0.56	0.55	0.05	0.51	0.26	0.25	0.10	0.08
Tainui Shellbed	0.14	0.28	0.22	0.77	0.30	0.33	0.80	0.71	0.00	0.77	0.46	0.60	0.25	0.22
Pinnacle Sand	0.10	0.14	0.10	0.39	0.09	0.17	0.26	0.19	0.05	0.25	0.19	0.24	0.11	0.08
Lower Castlecliff Shellbed,Castlecliff	0.05	0.06	0.04	0.25	0.06	0.06	0.11	0.11	0.02	0.11	0.00	0.10	0.06	0.04
Lower Castlecliff Shellbed, Waipuka	0.09	0.10	0.06	0.25	0.05	0.16	0.24	0.20	0.04	0.32	0.19	0.19	0.10	0.04
Kupe Formation	0.02	0.04	0.03	0.12	0.02	0.02	0.08	0.04	0.00	0.08	0.05	0.06	0.03	0.02
Kaimatira Pumice Sand	0.05	0.09	0.08	0.30	0.04	0.02	0.22	0.07	0.02	0.13	0.08	0.10	0.04	0.05
Tewkwsbury Formation	0.03	0.03	0.03	0.15	0.03	0.05	0.05	0.08	0.03	0.05	0.04	0.05	0.03	0.01
Waipuru Shellbed	0.05	0.07	0.07	0.21	0.12	0.20	0.23	0.34	0.02	0.25	0.18	0.24	0.12	0.07
Upper Nukumaru Group (4243)	0.05	0.09	0.08	0.33	0.07	0.14	0.35	0.26	0.07	0.40	0.24	0.24	1.02	0.06
Nukumaru Brown Sand	0.02	0.03	0.03	0.16	0.03	0.05	0.06	0.06	0.02	0.07	0.05	0.06	0.05	0.02
Upper Okiwa Group (4357)	0.16	0.27	0.18	0.33	0.14	0.16	0.39	0.17	0.02	0.30	0.18	0.29	0.12	0.20
Hautawa Shellbed	0.05	0.11	0.10	0.29	0.07	0.10	0.11	0.12	0.00	0.13	0.10	0.15	0.07	0.06

Appendix 1.12: Amino acid concentration of insolubles from *Waltonia* (ng/mg)

Appendix Two: Papers

Biogeochemistry of brachiopod intracrystalline molecules.

Curry, G.B.[1], Cusack, M.[1], Walton, D.[1], Endo, K.[1], Clegg, H.[1,2], Abbott, G.[2], Armstrong, H.[3]

[1] Department of Geology & Applied Geology, University of Glasgow, Glasgow, G12 8QQ, Scotland

[2]Newcastle Research Group in Fossil Fuels and Environmental Geochemistry, University of Newcastle Upon Tyne, Newcastle, NE1 7RU, England.

[3]Department of Geological Sciences, University of Durham, South Road, Durham, DH1 3LE, England.

CONTENTS

SUMMARY

Brachiopods contain both proteins and lipids within the biocrystals of their shell. One intracrystalline chromoprotein causes red shell coloration, while the other molecules may be involved with biomineralization, may strengthen the biocrystal, or may simply have been inadvertently engulfed by calcite during shell growth. Evidence is presented which indicates that intracrystalline molecules function as a closed system in fossil brachiopods. Investigations of the remains of these intracrystalline molecules in fossils have geological application in fields such as environmental reconstruction, isotopic determinations, taxonomy, and the interpretation of colour in extinct organisms.

1. INTRODUCTION

This paper deals with molecules that occur within the calcite of the brachiopod shell. These *intra*crystalline molecules are quite distinct from the *inter*crystalline organic membranes which surround the calcite biocrystals, and also from the 'metabolising' evaginations of

mantle tissue which penetrate many brachiopod shells (the caeca). This is not to say that the molecules recovered from within brachiopod calcite are uniquely distributed within the shell; indeed it is certain that they are synthesised within the secretory outer epithelium, and it is also possible that they have a function within the soft tissues of the organism.

The geological interest in intracrystalline molecules stems from the possibility that in vivo incorporation within resistant biocrystals may provide long-term protection from the various physical, chemical, and biological factors which would otherwise bring about the destruction of these molecules. The implication is that the inevitable decay of these molecules over geological timespans may not only proceed at much slower rates within the enclosing biocrystal, but that it may occur essentially *in situ*. Intracrystalline molecules may therefore provide unique access to fossilized remains of the original organic constituents of an organism which have not been subjected to significant post-depositional ingress by the abundant organic debris of various sources and ages which occurs within, and migrates through, enclosing sediments. Brachiopod shells are composed of robust low magnesium calcite which remains stable and in its original configuration for hundreds of millions of years (Williams 1984). Intercrystalline molecules situated between, rather that within, brachiopod biocrystals do, by contrast, decay rapidly and are prone to bacterial infection (Collins 1986).

Intracrystalline molecular fossils will undoubtedly be greatly altered over geological time irrespective of the stability of their enclosing biocrystal. Factors which will affect fossilisation potential include the range of molecules present inside biocrystals, their location within the shell, their interactions with the inorganic phase, their chemical stability, the environmental conditions (temperature, pressure etc), and the extent to which they can react with other intracrystalline molecules. As some of these factors are likely to have changed over geological time, there is no doubt that the breakdown of intracrystalline molecules will be extremely complicated. In this paper evidence for the *in situ* degradation of intracrystalline proteins and lipids in brachiopods is presented from the investigation of living and closely related fossil shells. The ultimate aim of this research is to determine the biological function of these molecules during biomineralization, and the geological applications of their preserved remains in fossils.

2. METHODS

Surface contaminant molecules and remnants of body tissue were removed by thoroughly cleaning the shells and incubating them for 2 h at 22°C in an aqueous solution of bleach (5% v/v). The shells were then powdered in a ceramic pestle and mortar before incubating overnight at 4°C in an aqueous solution of bleach (1% v/v). The bleach was removed by repeated washes with MilliQ™ water followed by centrifugation (8 g.h). The precipitate was washed until no bleach could be detected (typically ten 2 litre washes) and the precipitate was then lyophilised.

The $CaCO_3$ shell was dissolved using EDTA (20% w/v) at a ratio of 23 ml to 1 g shell.

The entire mixture was agitated at 4°C for 72 h or until the inorganic phase had dissolved. Following centrifugation (20 kg.h) the supernatant was concentrated, washed, and the EDTA removed using the Minitan™ tangential flow system from Millipore. The preparation was further concentrated in a Minicon™ concentrator (Amicon) to obtain sufficient quantities of intracrystalline proteins for biochemical analysis.

EDTA is very difficult to remove and various investigators have demonstrated spurious spectrophometer readings, non-reproducible gel electrophoresis patterns and other distortions which have been attributed to the incomplete removal of EDTA (Weiner 1984, Benson, Benson & Wilt 1986). During the present study the success in removing EDTA from preparations was monitored using an amino acid analyser, as EDTA produces a characteristic pattern of peaks on the reverse phase chromatography in this system. The results confirmed that EDTA cannot be entirely removed by dialysis or simple ultrafiltration. However the Minitan™ system, which has a much greater filtration area and tangential flow preventing the blocking of filters, allowed the removal of EDTA down to a level undetectable on the amino acid analyser (ie. less that 10^{-12} M). Typically the final solution represented a 10,000 fold concentration. Samples intended for bulk amino acid analysis were prepared by dissolving the shells in 2M HCl.

Samples for electrophoresis in 15% acrylamide gels were heated at 100°C for 4 min in an equal volume of sample buffer containing final concentrations of 0.15 M Tris/HCl, pH 6.8, 0.2 M 2-mercaptoethanol, 0.1% (w/v) SDS, 30% (v/v) glycerol and 0.0002% (w/v) of the tracking dye, bromophenol blue. Electrophoresis of samples in the small gel system (9cm x 7cm) required a constant voltage of 100 V for 2 h. Following electrophoresis, a constant voltage of 50V was applied to transfer the proteins onto ProBlott™ membrane in CAPS buffer (10 mM, pH 11) with methanol (10 % v/v). The membrane was then washed briefly with water and then methanol before staining with Coomassie Brilliant Blue-R in destain (0.1 % w/v) for 1 minute. Background staining was reduced with an aqueous dilution of methanol (50 % v/v). Hplc analysis of intracrystalline molecules was carried out using a WATERS 650 HPLC fitted with a multiwavelength detector. Amino acid analysis was carried out using an ABI 420H amino acid analyser fitted with an automatic hydrolysis head. Amino acid sequencing was carried out on the stained protein bands immobilised on ProBlott™ membrane using a pulsed liquid protein sequencer (Applied Biosystems 477A).

For lipid analysis, the shells were cleaned as above, and then etched in cold dilute HCl until 20% of their weight was removed. The etched shells were further cleaned by repeated washes in dichloromethane, and then powdered in a ceramic pestle and mortar. Initially the quantities of lipids extracted from Recent and fossil brachiopod shells using a 93/7 v/v mixture of dichloromethane/methanol were insignificant. Similarly demineralization of the shell powders did not yield significant levels of lipids even with repeated extractions with dichloromethane. Successful extraction of intracrystalline lipids was, however, achieved by a modified method in which 2N methanolic HCl was

added once the shell had been dissolved in cold 6N HCl, and the resulting aqueous mixture refluxed for 24 h. The solution was extracted with dichloromethane and the dichloromethane layer evaporated to dryness. The lipid fraction thus obtained was separated into two fractions using a short column of silica gel, eluted with 50/50 petroleum ether/ dichloromethane (2 ml) and methanol (2 ml).

The individual fractions thus collected were evaporated to dryness and analysed via combined gas chromatography/mass spectrometry (GC/MS). Gas chromatography was carried out using a Carlo Erba Mega series 5360 gas chromatograph fitted with a DB-5 coated, 30m fused silica capillary column (0.25μm film thickness and 0.32mm i.d.). The column was temperature programmed from 50-300°C at 4°C/minute with an isothermal temperature of 300°C for 20 minutes. Cold on column injection of the sample was employed and a detector temperature of 310°C was maintained. GC/MS analyses were carried out on a VG TS250 mass spectrometer (electron energy 70eV; filament current 4000mA; acceleration voltage 4kV; source temperature 200°C) interfaced with a Hewlett Packard 4890 GC fitted with a DB-5 coated, 30m fused silica capillary column (0.25μm film thickness and 0.32mm i.d.).

The possibility of contamination was monitored at all stages of preparation by analysing reagents. In the text and figures the standard three letter amino acid abbreviations have been used.

3. INTRACRYSTALLINE PROTEINS IN RECENT BRACHIOPODS

Polyacrylamide gel electrophoresis of intracrystalline molecules reveals a variable number of Coomassie-stained bands in different brachiopod groups. For example *Neothyris* has three main bands with estimated molecular weights of 47 kDa, 16 kDa and 6.5 kDa (figure 1). In contrast *Terebratulina* contains only a single band of molecular weight 30 kDa. These bands are sharp and readily stained by Coomassie Brilliant Blue suggesting that these are proteins rather than glycoproteins. This interpretation has been supported by amino acid analysis which revealed no sign of galactosamine or glucosamine.

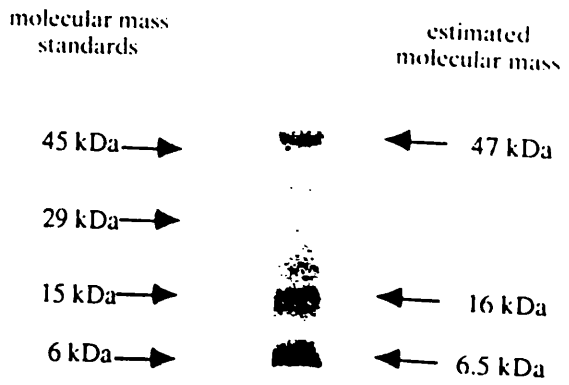


Figure 1: SDS PAGE of intracrystalline proteins in the Recent brachiopod *Neothyris* from New Zealand

Brachiopod intracrystalline proteins appear to be very different from the sparse data available from other phyla. Weiner (1983) isolated two proteins of similar molecular weight to the two smaller proteins in *Neothyris* from the calcite component of the bivalve *Mytilus*, but the amino acid analyses are completely different (the *Mytilus* proteins have over 50% Asp and Glu, compared with 20% for the brachiopod). Similarly phosphophryns, the major non-collageous proteins of dentin, have an distinctive amino acid composition (Veis 1989) unlike that of any brachiopod intracrystalline proteins. In terms of electrophoretic behaviour and amino acid analysis, brachiopod intracrystalline proteins are again different from the glycoproteins extracted from sea urchins, which could not be stained by Coomassie (Benson, Benson & Wilt 1986).

The sequence of a 50 kDa sea urchin intracrystalline glycoprotein has been determined by nucleotide sequencing of the coding gene (Benson *et al* 1987), but the development of a protocol which allowed purification of brachiopod intracrystalline proteins to homogeneity has permitted direct amino acid sequencing (Curry *et al* in press; Cusack *et al* submitted). As yet no similarity has been detected between the amino acid sequence of brachiopod intracrystalline proteins and published amino acid or nucleotide sequences (Curry *et al* in press; Cusack *et al* submitted).

The function of intracrystalline proteins is poorly understood. It has recently been suggested that they strengthen the shell (Berman *et al* 1990), although it is also possible that some have been inadvertently incorporated into the shell during growth. As discussed below, the investigation of brachiopod shell proteins has revealed that one is responsible for shell colour (Cusack *et al* submitted). The indications are that there is considerable variation in the nature of intracrystalline proteins both within different groups of brachiopods and between different phyla.

4. INTRACRYSTALLINE AMINO ACIDS IN FOSSIL BRACHIOPODS

There are several lines of evidence that suggest that intracrystalline molecules remain a closed system over geological time. Firstly amino acid analyses of progressively older fossil representatives of a single genus reveal a pattern of decay that is predictable

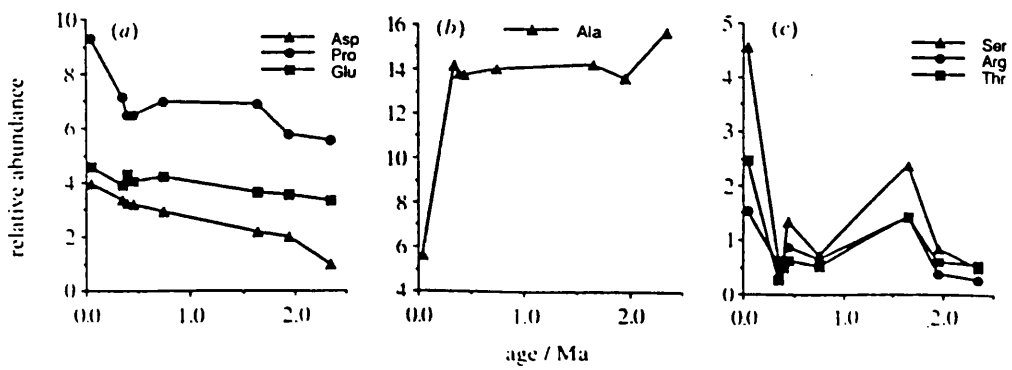


Figure 2: Relative abundance of intracrystalline amino acids in Recent and fossil shells of *Waltonia* from New Zealand.

from the chemical characteristics of individual amino acids. Within the shell of *Waltonia* for example, relatively unstable amino acids (such as Asp, Glu, Pro & Ser) show a steady decline in relative abundance with increasing age (figure 2b). Alanine, one of the breakdown products of serine (Hoering 1980), shows a corresponding increase (figure 2c). This suggests that the parent proteins have progressively decayed *in situ* and have not been contaminated by the abundant amino acids which have been shown to occur in the surrounding sediments (Walton & Curry in press).

Immunological techniques have also demonstrated that original antigenic determinants, including peptides, have survived in these brachiopods, and such techniques offer an important method of extracting original molecular information from Pleistocene shells (see Lowenstein & Scheuenstuhl 1991). The immunological approach has proved less useful with older fossils, possibly due to interference from other intracrystalline compounds, and possibly EDTA, as well as the decay of determinants on the target molecules. However, amino acid analysis has suggested that intracrystalline molecules do remain uncontaminated for much greater periods of geological time. For example 27 million year old (Ma) specimens of *Terebratulina* have a similar amino acid composition to that of the single protein found in Recent shells of this genus. The major differences, the increase in Gly and Ala (both products of the decay of Ser), and decrease of Asp, Ser, Thr and Pro, are again predictable consequences of the known chemical properties of these amino acids (figure 3). The absolute abundance of soluble amino acids in 27 Ma shells of *Terebratulina*

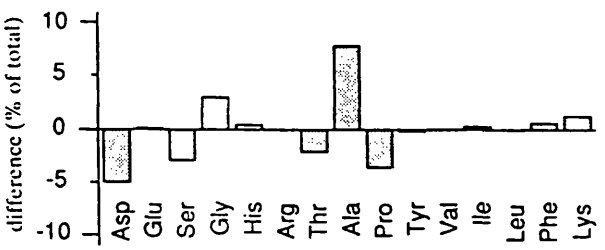


Figure 3: Comparison of the intracrystalline amino acid compositions of Recent *Terebratulina* and 27 Ma fossil shells of the same genus from New Zealand. The data for each analysis have been expressed in mole percentages, and the vertical axis indicates the difference detected in relative abundance for each amino acid between the fossil and Recent shells.

has dropped to about 50% of that in Recent shells. Some amino acids decay completely, but the most significant reduction is probably caused by incorporation of other amino acids into insoluble molecular agglutinations (melanoidins of Hoering 1980) which increase in abundance in fossils and are excluded from amino acid analysis of the soluble fraction.

A third line of evidence comes from principal component analyses (PCA) of amino acid analyses of intracrystalline proteins. In Recent shells, differences in homologous amino acid sequences and the variation in the number and possibly also the proportions of proteins present, has produced differences in amino acid composition which can be revealed by PCA (figure 4). For example four New Zealand brachiopod genera are readily

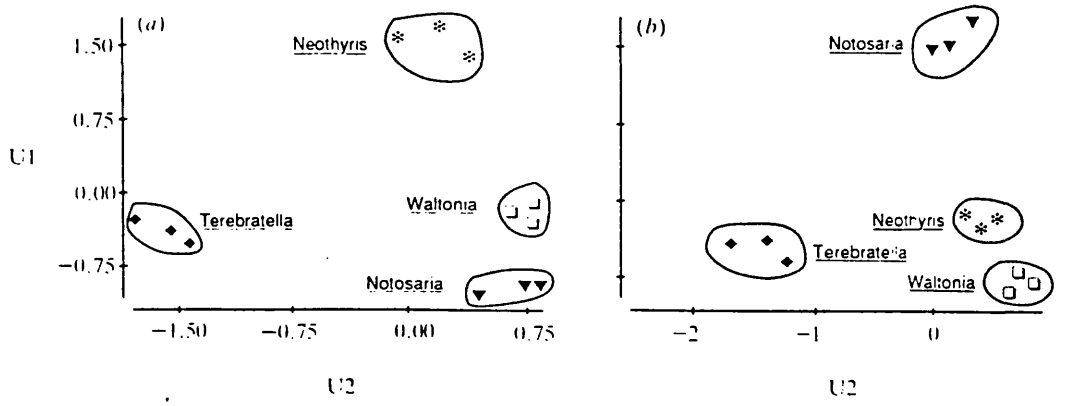


Figure 4: Plot of first (U1) and second (U2) principal component axes of amino acids in the shells showing the differentiation (a) between four Recent brachiopod genera and (b) between fossil representatives of the same taxa, collected from the Shakespeare Group, near Wanganui, N. Island, New Zealand. In (a) Gly has a high positive scoring on the first principal component axis, and Ser, Ala, Ile, Leu and Lys have high negative scores, whereas for the second principal component axis, Asp and Glu have high positive scores and His, Arg, Thr and Pro have high negative scores. In (b) Thr has a high positive scoring on the first principal component axis, and Gly, Arg Valine and Leucine have high negative scores, whereas for the second principal component axis, Asp and Glu have high positive scores and Pro, Ile and Ala have high negative scores.

distinguished on the basis of amino acid composition of their shells (figure 4a). The raw data from amino acid analyses is difficult to compare, and PCA not only has the advantage of presenting a clear graphical summary of the dispersion of the data but the eigenvector values provide a clear indication of which amino acids are important in distinguishing between different taxa (Walton & Curry in press). The same four genera were collected as fossils from within a single horizon of the Shakespeare Group in New Zealand (approximately 350,000 years BP), and were again distinguished by PCA analysis of their intracrystalline amino acid composition (figure 4b). The shells of the four genera have not homogenised in the fossil record, as would be the case if they had absorbed amino acids from the surrounding sediments. Amino acids are abundant in sediments, but in all cases investigated their compositions are readily distinguished from brachiopod shell amino acids when subject to PCA (Walton & Curry in press). This implies that the shells have remained closed systems, and that differences detected by PCA reflect back to the original compositional variability detected in their living representatives. Although shell amino acid composition may be affected by environmental conditions during the life of the organism, in this case these effects have been identical because the specimens lived together and certainly experienced identical diagenetic histories.

5. LIPIDS IN RECENT BRACHIOPODS

GC and GC/MS analysis of intracrystalline molecules from Recent brachiopod shell extracts reveals the presence of an homologous series of saturated acyclic carboxylic acids (derivatised as the methyl ester) as well as a number of unsaturated components. The

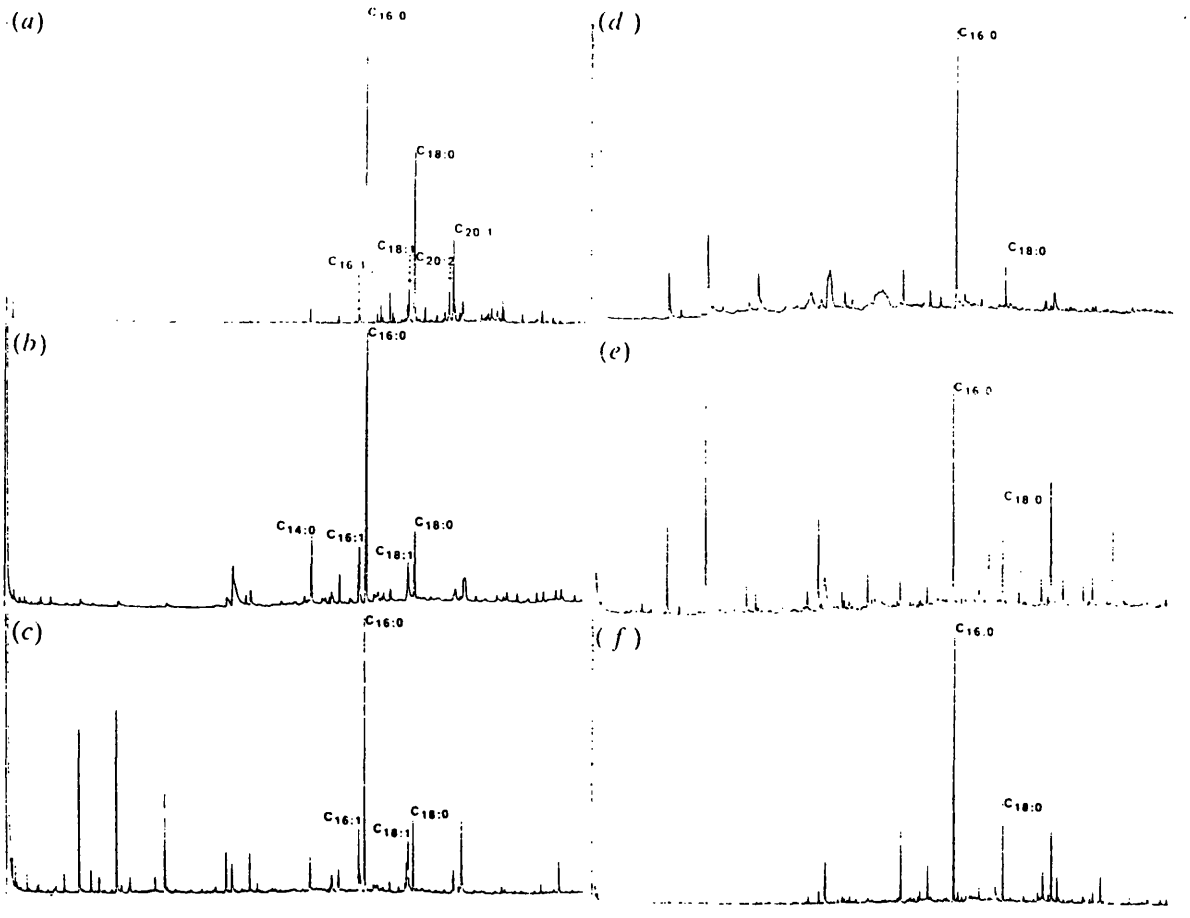


Figure 5: Gas chromatograms of petroleum ether; dichloromethane (50:50 by volume). Carbon numbers of the *N* fatty acids are indicated. (a) Recent *Neothyris*; (b) Shakespeare Group *Neothyris*; (c), Tewkesbury Fm *Neothyris*; (d) Castlepoint Fm *Neothyris*; (e) Gee Greensand Fm *Pachymagus*; (f), McDonald Limestone Fm *Liothyrella*. In (e) the major peak on right of chromatogram (retention time = 51 min) has been assigned to contamination.

dominant lipids detected in *Neothyris* in order of decreasing abundance are:- n-$C_{16:0}$, n-$C_{18:0}$, n-$C_{20:1}$, n-$C_{17:0}$, n-$C_{18:1}$, n-$C_{20:2}$ fatty acids (figure 5a).

6. LIPIDS IN FOSSIL BRACHIOPODS

A range of *Neothyris* fossil extracts were analysed using the method described above. *Neothyris* shells from the Shakespeare Group (350,000 years BP) contained a range of saturated and unsaturated carboxylic acids, namely n-$C_{16:0}$, n-$C_{18:0}$, n-$C_{14:0}$, n-$C_{16:1}$, n-C_{18-1} in order of decreasing abundance (figure 5b). With increasing age low molecular weight lipids begin to increase in relative abundant to the above acids, but as yet their assignments are equivocal. *Neothyris* from the Tewkesbury Formation (Fm) (1.6 Ma) contained n-$C_{16:0}$, n-$C_{18:0}$, n-$C_{16:1}$, n-$C_{18:1}$ fatty acids (figure 5c), while the same genus from the older Castlepoint Fm (1.85 Ma) contained n-$C_{16:0}$ and n-$C_{18:0}$, although there is a notable absence of unsaturated fatty acids (figure 5d).

Analyses of older fossil brachiopods of a different genus (*Pachymagas*) from the Gee Greensand Fm (South Island, New Zealand; 23Ma) revealed n-$C_{16:0}$ and n-$C_{18:0}$ fatty acids (figure 5e), and these same compounds were also recovered from the shells of *Liothyrella* from the McDonald Limestone Fm (30 Ma). These fossil extracts did not yield unsaturated fatty acids (figure 5f).

The presence of lipids within brachiopod shells is well known (Jope 1965; Stoyanova 1984) but until now there has been no attempt to distinguish between intercrystalline and intracrystalline lipids. The distinction is important as the intercrystalline lipids are more prone to post-depositional ingress. In this study intercrystalline lipids have been detected in extractions of the shell powders, and these are quite distinct from the intracrystalline lipids released from the subsequent demineralization of the residues. These results have therefore demonstrated for the first time the presence of lipids within brachiopod biocrystals.

All Recent and fossil brachiopod shells examined contained normal saturated carboxylic acids, with $C_{16:0}$ (palmitic) and $C_{18:0}$ (stearic) fatty acids the dominant compounds present. With increasing age there is a decrease in the levels of lipids extracted from the shells, and a decrease in the levels of unsaturated fatty acids. In the samples studied the unsaturated carboxylic acids disappear in shells older than approximately 2 Ma. Low molecular weight compounds increased in relative abundance with increasing age, with the exception of *Liothyrella* from the McDonald Limestone Fm (figure 5b-f); these compounds are not present in the shells of Recent brachiopods.

The function of these intracrystalline lipids is poorly understood, although these compounds may be responsible for establishing some of the essential conditions for calcification, as has been suggested for the lipid fraction of ostrich egg shells (Kriesten *et al,* 1979). The distribution and abundance of intracrystalline lipids may well be amenable to the kind of multivariate analysis applied to amino acids, and could augment the molecular discrimination between Recent and fossil taxa.

7. APPLICATIONS OF INTRACRYSTALLINE MOLECULES

The demonstration that brachiopod intracrystalline molecules do not appear to be contaminated by extraneous amino acids or lipids over time suggests several geological applications; their taxonomic uses have been discussed above and in immunological papers (Lowenstein and Scheuenstuhl 1991) and will not be expanded here. The short amino acid sequences recoverable from fossils have little geological application, but extensive research has already highlighted important applications for amino acids in absolute dating, stratigraphic correlation (including correlation from land to sea (Bowen *et al* 1989)), environmental reconstruction, and nitrogen and carbon stable isotope determinations to investigate the diets of ancient organisms. It has been suggested that stable isotope ratios from individual amino acid enantiomers will also allow the indigenity

of the amino acid to be demonstrated (Engel & Macko 1986), especially from biominerals (such as the apatite in bone) which has a more porous structure than brachiopod calcite.

Previous isotopic determination from shells must inevitably have included carbon and nitrogen from molecules as well as from calcite, but as the absolute quantities of molecules involved is probably of the order of 1% of shell weight (amino acids alone account for about 0.04% of shell weight in *Terebratulina*, not including the substantial quantities of lipids, carbohydrates, and other molecules present) the difference is likely to be comparatively insignificant and within the range of experimental error. However if there was any preferential release of carbon and nitrogen from intracrystalline molecules during the preparation of shells for isotopic investigation then the significance of these molecules would become much more significant.

There are indications that the decay rates of proteins in brachiopod shells provide useful environmental information. For example several relatively unstable amino acids appear to be much better preserved in the *Waltonia* specimens from the Tewkesbury Fm than in most other Plio-Pleistocene horizons investigated (ie the peak at approximately 1.6Ma in figure 2c). The Tewkesbury Fm represents a shallow marine environment, as compared with the fully marine faunas of the other formations investigated (Fleming 1953). The improved preservation of serine and other usually fragile amino acids in the Tewkesbury Fm may reflect superior preservation conditions or a relationship between environmental conditions and the types of amino acids incorporated. These possibilities are being investigated given the widespread distribution of Recent brachiopods and their fossil ancestors in New Zealand ecosystems and Plio-Pleistocene successions. Other applications may appear once the function of intracrystalline molecules is revealed; for example if any were toxic they would discourage predation.

8. SHELL COLOUR

During the biochemical investigation of Recent brachiopod intracrystalline molecules, it became clear that red shells retained their colour throughout the preparation protocol which stripped off all intercrystalline molecules. The solution produced by dissolving the calcium carbonate of the shell was also red, indicating that shell coloration is caused by a soluble intracrystalline molecule. Reverse phase hplc of these samples allowed the isolation of the coloured fractions, which were then re-analysed by SDS PAGE. In all six red-shelled brachiopod genera analyzed, the molecule responsible for shell colour is an intracrystalline protein (ICP 1) with an apparent molecular weight of 6.5 kDa (Cusack *et al* submitted). All other intracrystalline molecules are colourless, and non-coloured shells do not have ICP1.

Detailed analyses of ICP1 have confirmed that this chromoprotein has a very similar N-terminal amino acid sequence in the three red brachiopod genera *Waltonia*, *Neothyris*, *Terebratella* (Cusack *et al* submitted). The proteins are clearly homologous (there is no need to insert gaps to align them) and searching of recent versions of the protein and DNA

sequence databases (EMBL, NBRF, GenBank & SwissProt using the GCG package) suggests that this molecule has not been sequenced previously.

UV spectrophotometry and standard colour tests (Dunning 1963) have indicated that the prosthetic compound may be a carotenoid, which has a similar role in crustacean chromoproteins (Zagalsky 1976). ICP1, like other chromoproteins, is soluble in water, but the prosthetic chromophore becomes insoluble when separated which is characteristic of a lipid-protein association. This emphasises the importance of preparation technique, because the chromophore could not be isolated for analysis in organic solvents until it had been stripped from the protein segment of the molecule.

The importance of demonstrating that the red colour of brachiopods is caused by an intracrystalline protein, stems from the fact that fossil brachiopods with preserved colour patterns have been documented from many geological horizons. During the investigation of New Zealand Pleistocene brachiopods a number of specimens with red coloured shell corresponding to that of Recent con-generic species have been recovered from sediments up to 400,000 years old. Records of colour pattern do extend much further back in time, indeed as far back as the Palaeozoic. Surprising perhaps, some of the best recorded examples of patterned brachiopod shells occur in Devonian and Carboniferous rocks, with ages ranging from 350 to over 400 Ma (Blodgett, Boucot & Koch 1988).

At these ages the colour has usually faded or altered, although there are some reports of original colours being preserved which fade rapidly on exposure to light and/or oxygen. Most fossil 'coloured' brachiopods simply show colour patterns which are reddish brown or black, suggesting that the original prosthetic molecule has altered over geological time. The majority of fossil brachiopods do not show preserved colour patterns, although shells which were originally uniformly coloured, rather than patterned, are easily overlooked. If the shell does indeed remain a closed system for long periods of geological time then it may be possible to recover the alteration or breakdown products of the chromoprotein from shells which have lost their colour as a result of the light or oxygen sensitivity of the prosthetic group. This may open up the possibility of reconstructing the colour of fossil shells from an intracrystalline biomarker, an important new ability given that colour may be a crucial factor in evolutionary history (radiations into shallow water and onto land) because it provides protection from predators and radation. Blumer (1965) has shown that this is possible by determining the breakdown pathway of a chromophore found in the shell of fossil crinoids.

9. DISCUSSION

Apart from lipids and proteins it is thought that other intracrystalline compounds are present inside biocrystals such as carbohydrates. The full exploitation of this phenomenon will only be realised when these compounds have been localized, characterized, and their function determined. It is possible to recover amino acids, lipids and probably other types of molecules from early Mesozoic or even Palaeozoic brachiopods (ie up to 550 Ma) but

undoubtedly it becomes progressively more difficult to utilize these molecules in older specimens. Potential problems which must be addressed include the possibility of regeneration (eg. amino acids decay and then are produced *de novo* from simple building blocks), coelution and hence misidentification of decay products with amino acids due to identical chromatographic behaviour, and the generation of entirely new molecular species. The precise location of molecules within the shell, and their relationship with the calcite matrix, are also crucial factors which have considerable bearing on the extent to which intracrystalline molecules survive in the fossil record, are free to interact with other components, and can be utilised in palaeobiological studies.

Intracrystalline Molecules from Brachiopod Shells

Curry, G.B., Cusack, M., Endo, K., Walton, D. and Quinn, R. Department of Geology and Applied Geology. University of Glasgow. Lilybank Gardens. Glasgow. G12 8QQ. Scotland. U.K.

Key words: Brachiopod, intracrystalline, protein primary sequence, amino acid analysis.

INTRODUCTION

Shells are composed of both organic and inorganic constituents. It is believed that the organic compounds have important functions at several stages during the formation of biominerals. In brachiopod shells the disposition of inorganic biominerals and their enclosing organic sheaths have been thoroughly investigated using both scanning and transmission electron microscopy but little is known about the biochemistry of the intracrystalline molecules i.e. those enclosed within the inorganic portion. Such information is crucial for an understanding of biominerals if, as has been suggested, these compounds (i) induce crystal nucleation by providing a surface for precipitation, (ii) form compartments that determine the shape and volume of the biocrystal and (iii) determine the pattern of growth in the mineral phase in what is termed 'matrix mediated mineralisation' [1].

This study presents the first details of the organic intracrystalline components from the shell of the articulate brachiopod, *Neothyris lenticularis* (Deshayes). The shell of *N. lenticularis* is composed of numerous long calcite fibres from which the proteins studied here have been extracted. Although this protocol has disadvantages in that it is very time-consuming and the great proportion of the shell is discarded, the major advantage of such a strategy is that it avoids the possibility of including extraneous molecules both from contaminating organisms, such as bacteria, which may infest the organic sheaths of shell calcite fibres and from human finger tips during collection and preparation.

Partial N-terminal sequence and amino acid analyses of two shell proteins are presented here alongside SDS PAGE and hplc analyses of the intracrystalline molecules of *N. lenticularis*.

MATERIALS AND METHODS

Extraction of Shell Proteins

Shells of living *N. lenticularis* were collected from Stewart Island, New Zealand and killed by dehydration. The shells were cleaned thoroughly and incubated for 2 h at 22°C in an aqueous solution of bleach (5% v/v) to destroy the organic sheath and any possible bacterial contamination. The shells were then powdered in a ceramic pestle and mortar before incubating overnight at 4°C in an aqueous solution of bleach (1% v/v). The bleach was removed by repeated washes with Milli Q™ water followed by centrifugation (8 g.h). The precipitate was washed until no bleach could be detected (typically ten 2 l washes) and then lyophilised and EDTA (20% w/v), pH 11 added in the ratio of 23 ml to 1 g shell. The entire mixture was agitated at 4°C for 72 h or until the inorganic phase had dissolved. Following centrifugation (20 g.h) the supernatant was concentrated and the EDTA removed using the Millipore Minitan™ tangential flow system. The preparation was further concentrated in a minicon static concentrator (Amicon) with a 10 kDa cut-off membrane.

Separation of Proteins by hplc

An aliquot of concentrated shell extract was applied to a reverse-phase Aquapore™ RP-300 narrow bore (2.1 mm diameter x 30 mm length) column in trifluroacetic acid (0.1% v/v) at a flow rate of 0.1 ml / min. After 5 min, a 40 min linear gradient of 0 to 70 % (v/v) acetonitrile, in 0.1% (v/v) trifluroacetic acid, was applied to fractionate the shell proteins. The eluate was monitored at 280 and 214 nm.

Separation of Proteins by SDS PAGE

Small gels (9 cm x 7 cm) of 0.75 mm thickness containing 15% polyacrylamide were prepared according to the method of Schagger and Van Jagow [2]. Glycine, which is used in most SDS PAGE systems is here replaced by tricine. Samples for electrophoresis were heated at 100°C for 4 min in an equal volume of sample buffer containing final concentrations of 0.15 M Tris / HCl, pH 6.8, 0.2 M 2-mercaptoethanol, 0.1% (w/v) SDS, 30% (v/v) glycerol and 0.0002% (w/v) of the tracking dye, bromophenol blue. Molecular weight standards; bovine albumin (66 kDa), egg albumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase (36 kDa), carbonic anhydrase (29 kDa), trypsinogen (24 kDa), trypsin inhibitor (20.1 kDa) and a-lactalbumin (14.2 kDa) were included on every gel. Electrophoresis of samples in the small gel system required a constant voltage of 100 V for 2 h. Following electrophoresis, proteins were either fixed in the gel and visualised

using Coomassie Brilliant Blue-R or electroblotted onto ProBlott™ membrane (Applied Biosystems).

Electroblotting of Proteins

Following SDS PAGE, the proteins in the gel were transferred to ProBlott™ membrane. The transfer was performed in transfer buffer (10 mM CAPS buffer, pH 11, 10% (v/v) methanol) in a Bio-Rad Trans Blot cell. A constant voltage of 50 V for 0.5 h moved the proteins from the gel towards the membrane. Coomassie Blue staining was used to reveal the protein bands on the ProBlott™ membrane.

Amino-Terminal Sequence Determination and Amino Acid Analysis

Automatic Edman degradation was carried out on the stained bands using a pulsed liquid protein sequencer (Applied Biosystems 477A). Bands were also loaded onto the 420-H amino acid analyser with automatic hydrolysis (Applied Biosystems) to determine the overall amino acid composition. Amino acid analysis of the hplc fractions was employed to identify the major proteins by comparing with the analyses from the homogeneous electroblotted proteins. Stained ProBlott™ membrane with no protein attached and all buffer solutions employed were analysed to determine the background level of amino acids present.

RESULTS AND DISCUSSION

The shell of *N. lenticularis* contains a mixture of proteins of different molecular weight, as determined by SDS PAGE (Figure 1). The major proteins are of molecular weight 10.5, 13 and 47 kDa. An identical copy of the the protein pattern in an SDS gel is obtained on a membrane, such as ProBlott™, using electroblotting. This process, first described by Towbin *et al* in 1979 [3], is ideal for the isolation of single proteins from crude mixtures for N-terminal sequence analysis.

The elution profile of the shell proteins from the hplc system is presented in Figure 2. The light pink colour of Recent *N. lenticularis* becomes deep red as the shell extract is concentrated. The red pigment elutes from the hplc system in the elution volume 42 to 50 ml.(Figure 2). This corresponds with the elution of the 10.5 kDa protein. One possible explanation for the elution of the 10.5 kDa protein as several different peaks is that the molecule may in fact be a glycoprotein and the various peaks represent various degrees of glycosylation. The harsh conditions employed in SDS PAGE causes the protein(s) and pigment to dissociate.

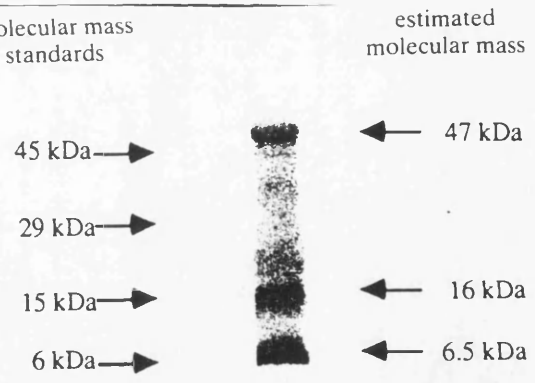


Figure 1 Intracrystalline Proteins of N. lenticularis shell

The proteins were separated by SDS PAGE (see Materials and Methods for details) and then revealed using Coomassie Brilliant Blue staining.

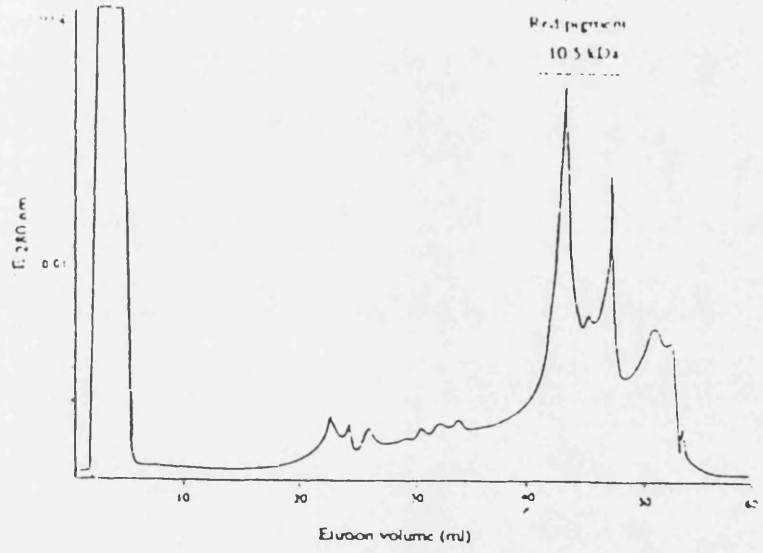


Figure 2 Elution profile of the intracrystalline molecules of N. lenticularis from reverse-phase hplc

An aliquot of the intracrystalline extract was applied to a reverse-phase Aquapore™ RP-300 narrow bore column and a linear gradient of 0 to 70 % (v/v) acetonitrile was applied to fractionate the shell proteins (See Materials and Methods for details).

The amino acid composition of the 10.5 and 47 kDa proteins are listed in Figure 3.

	47 kDa	10.5 kDa
Residue	mole %	
D/N	8.53	10.33
E/Q	8.18	9.06
S	5.57	5.65
G	8.89	11.50
H	1.10	0.76
R	0.10	3.21
T	6.12	6.21
A	13.02	9.21
P	6.47	16.23
Y	4.27	1.58
V	3.97	7.02
C	1.10	0.66
M	1.81	1.58
I	8.13	2.39
L	10.49	10.28
F	3.21	2.29
K	9.04	2.03

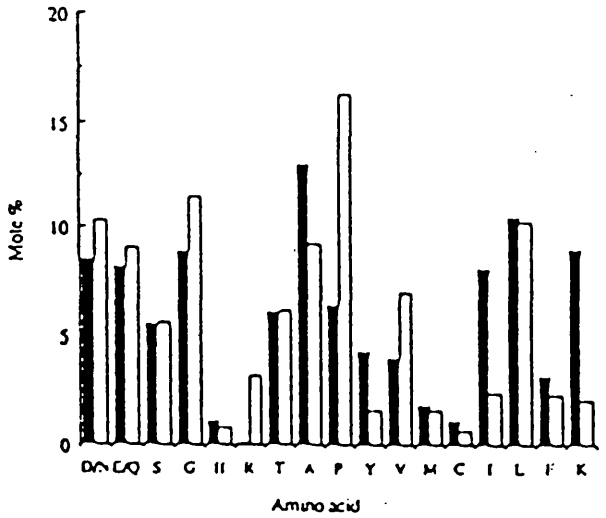


Figure 3 ***Amino acid composition of two of the major intracrystalline proteins of N. lenticularis shell.***

Proteins were extracted and purified to homogeneity using SDS PAGE and electroblotted onto ProBlott™ and the amino acid composition determined using the 420-H amino acid analyser (Applied Biosystems). Values are presented for both the 10.5 ( ) and 47 kDa ( ) proteins.

Amino acid concentrations are presented as mole percentages to enable direct comparison of the two proteins. Actual yields of intracrystalline amino acids from the bulk mixture extracted from *N. lenticularis* indicate that intracrystalline amino acids occur at about 80 nmole/g shell material.

The conditions employed for the hydrolysis of the peptide bonds also destroys 100 and up to 80% of tryptophan and cysteine residues respectively. However,the hydrolysis reaction is automated and highly reproducible and thus each protein should lose the same proportion of these vulnerable residues allowing direct comparison of the compositions.

The N-terminal amino acid sequence of the 10.5 and 47 kDa proteins are presented in Figure 4.

	1	6	11	16
10.5 kDa	GPEQL	PYATM	ISKTS	NATKP
	1	6		
47 kDa	ANLVL	AGRGD		

Figure 4 ***Amino-terminal sequence of 10.5 and 47 kDa protein from Neothyris lenticularis shell.***

Proteins were extracted from N lenticularis and then fractionated by SDS PAGE before being transferred to ProBlott™ for sequence determination (see Materials and Methods).

These protein sequences show no significant similarity to any protein sequence listed in the NBFR or EMBL data base or to any implied peptide sequence in the EMBL or GenBank DNA sequence data base on the basis of searches conducted by the 'FastA' and 'TFastA' programs in the GCG sequence analysis package (Version 6) [4]. Elucidation of the entire primary sequence of these intracrystalline proteins should make sequence comparisons more meaningful and may enhance our understanding of the possible role these proteins play in the process of biomineralisation.

REFERENCES

[1] Lowenstam, H.A. (1981) Science **211** 1126 - 1131

[2] Schagger, H. and Von Jagow, G. (1987) Anal. Biochem. **166** 368 - 379

[3] Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. **76** 4350 - 4354

[4] Devereux, J., Haeberli, P. and Smthies, O. (1984) Nucleic Acids Research **12**(1) 387 - 395

Appendix Three: Addresses of suppliers

Addresses of Suppliers

Amicon Division,
W.R. Grace and Co.,
Conn.,
Beverly,
MA01915, USA.

Applied Biosystems Ltd.,
Kelvin Close,
Birchwood Science Park North,
Warrington,
Cheshire,
WA3 7PB.

BDH Chemicals Ltd.,
Merck Ltd.,
Burnfield Avenue,
Thornliebank,
Glasgow,
G46 7TP.

Bio-Rad,
Caxton Way,
Watford,
Hertfordshire.

The Boots Company PLC,
Nottingham,

Heraeus,
Unit 9, Wates Way,
Brentwood,
Essex,
CM15 9TB.

V.A. Howe and Company,
12-14 St. Ann's Crescent,
London,
SW18 2LS.

Millipore UK Ltd.,
Research Avenue South,
Research Park,
Herriot-Watt University,
Riccarton,
Edinburgh,
EH14 4AB.

Pharmacia,
Pharmacia House,
Midsummer Boulevard,
Central Milton Keynes,
Bucks.,
MK9 3HP.

Pierce Chemicals,
44, Upper Northgate Street,
Chester,
CH1 4EF.

Scotlab,
Kirkshaws Road,
Coatbridge,
Strathclyde,
ML5 8AD.

Sigma Chemicals,
Fancy Road,
Poole,
Dorset,
BH17 7NH.

The following companies are marketed through Scotlab:

Cel-cult
Eppendorf
Falcon
Luckman
Rainin
Sterilin

References cited

References Cited

Abelson, P.H. (1954). Organic constituents of fossils. *Carnegie Inst. Wash. Yearb.*, **53**, 97-101.

Abelson, P.H. (1955). Organic constituents of fossils. *Carnegie Inst. Wash. Yearb.*, **54**, 107-109.

Abelson, P.H. (1956). Paleobiochemistry. *Sci. Am.*, **195**, 83-92.

Addadi, L., Berman, A. and Weiner, S. (1991). Intracrystalline proteins from a sea urchin and a mollusc: a comparison. *In* Suga, S. and Nakahara, H. (eds.): *Mechanisms and phylogeny of mineralization in biological systems*, **29-33**. Springer-Verlag, Tokyo. 517pp.

Aitken, A., Geisow, M.J., Findlay, J.B.C., Holmes, C. and Yarwood, A. (1989). Peptide preparation and characterization. *In* Findlay, J.B.C. and Geisow, M.J. (eds.): *Protein sequencing: a practical approach,* **43-68**, (Rickwood, D. and Hames, B.D., series eds). IRL Press, Oxford. 199pp.

Akiyama, M. (1971). The amino acid composition of fossil scallop shell proteins and non-proteins. *Biomineralisation*, **3**, 65-70.

Akiyama, M. (1980). Diagenetic decomposition of peptide linked serine residues in the fossil scallop shells. *In* Hare, P.E., Hoering, T.C. and King, K. (eds.): *Biogeochemistry of the amino acids*, **115-120**. John Wiley and Sons, New York. 558pp.

Allen, G. (1989). *Sequencing of proteins and peptides: Laboratory techniques in biochemistry and molecular biology* (Burdon, R.H. and Von Knippenberg, P.H. series eds.). Elseiver, Amsterdam. 425pp.

Al-Rikabi, I. (1992). A molecular approach to palaeontology: Biochemical method applications of brachiopod proteins. Unpublished MSc thesis, University of Glasgow, U.K.

Anderton, P.W. (1981). Structure and evolution of the South Wanganui Basin, New Zealand. *NZ J. Geol. Geoph*, **24**, 39-83.

Armstrong, W.G., Halstead, L.B., Reed, F.B. and Wood, L. (1983). Fossil proteins in vertebrate calcified tissues. *Phil. Trans. Royal. Soc. Lond.*, **301B**, 301-343.

Bada, J.L. (1971). Kinetics of the non-biological decomposition and racemization of amino acids in natural waters. *Adv. Chem. Ser.*, **106**, 309-331.

Bada, J.L and Man, E H. (1980). Amino acid analysis in Deep Sea Drilling Project Cores: kinetics and mechanisms of some reactions and their applications in geochronology and in paleotemperature and heat flow determinations. *Ear. Sci. Rev.*, **16**, 21-55.

Bada, J.L and Miller, S.L. (1968). Ammonium ion concentration in the primitive ocean. *Science*, **159**, 423-425.

Bada, J.L and Miller, S.L. (1969). Kinetics and mechanism of the non enzymatic reversible deamination of aspartic acid. *J. Am. Chem. Soc.*, **91**, 3946-3947.

Bada, J.L and Miller, S.L. (1970). Kinetics and mechanism of the reversible non enzymatic deamination of aspartic acid. *J. Am. Chem. Soc.*, **92**, 2774-2780.

Bada, J.L., Shou, M.-Y., Man, E.H. and Schroeder, R.A. (1978). Decomposition of hydroxy amino acids in foraminifera tests; kinetics, mechanism and geochronological implications. *Earth. Planet. Sci. Lett.*, **41**, 67-76.

Benner, R., Maccubbin, A.E. and Hodson, R.E. (1984). Anaerobic biodegradation of the lignin and polysaccharide components of lignocellulose and synthetic lignin by sediment microflora. *Appl. Envir. Microbiol.*, **47**, 998-1004.

Benson, S.C., Benson, N.C. and Wilt, F. (1986). The organic matrix of the skeletal spicule of sea urchin embryos. *J. Cell. Biol.*, **102**, 1878-1886.

Benson, S., Sucov, H., Stephens, L., Davidson, E. and Wilt, F. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. *Devel. Biol.*, **120**, 499-506.

Berman, A., Addadi, L., Kvick, A., Leiserowitz, L., Nelson, M. and Weiner, S. (1990). Intercalation of sea urchin proteins in calcite: study of a crystalline composite material. *Science*, **250**, 664-667.

Berman, A., Addadi, L. and Weiner, S. (1988). Interactions of sea-urchin skeleton macromolecules with growing calcite crystals - a study of intracrystalline proteins. *Nature*, **331**, 546-548.

Beu, A.G. and Edwards, A.R. (1984). New Zealand Pleistocene and late Pliocene glacio-

eustatic cycles. *Palaeogeog. Palaeoclim. Palaeoecol.*, **46**, 119-142.

Blodgett, R.B., Bocot, A.J. and Koch, W.F. (1988). New occurrences of color patterns in Devonian articulate brachiopods. *J. Paleont.*, **62**, 46-51.
Blumer, M. (1965). Organic pigments: their long-term fate. *Science*, **149**, 722-726.

Bowen, D.Q., Hughes, S., Sykes, G.A. and Miller, G.H. (1989). Land-sea correlation in the Pleistocene based on isoleucine epimerization in non-marine molluscs. *Nature*, **340**, 49-51.

Brenner, S., Jacob, F. and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, **190**, 576-581.

Briggs, M.H. (1961). Amino acids and peptides from some New Zealand fossils. *NZ. J. Geol. Geoph.*, **4**, 387-91.

Calvin, M. (1968). Molecular Palaeontology. *Trans. Leicester. Literary. Phil. Soc.*, **62**, 45-69.

Collins, M.J. (1986). Post mortality strength loss in shells of the Recent articulate brachiopod Terebratulina retusa from the West Coast of Scotland. *In* Racheboeuf, P.R. and Emig, C.C. (eds.): *Les brachiopodes fossiles et actuels*, **209-219.** Biostratigraphie du Peléozoique 4, 500pp.

Collins, M.J., Curry, G.B., Muyzer, G., Quinn, R., Xu, S., Westbroek, P. and Ewing, S. (1991c). Immunological investigations within the Terebratulid brachiopods. *Palaeontology*, **34**, 785-796.

Collins, M.J., Curry, G.B., Quinn, R., Muyzer, G., Zomerdijk, T., and Westbroek, P. (1988). Sero-taxonomy of skeletal macromolecules in living terebratulid brachiopods. *Hist. Biol.*, **1**, 207-224.

Collins, M.J., Muyzer, G., Curry, G.B., Sandberg, P. and Westbroek, P. (1991a). Macromolecules in brachiopod shells: characterization and diagenesis. *Lethaia*, **24**, 387-397.

Collins, M.J., Muyzer, G., Westbroek, P., Curry, G.B., Sandberg, P., Xu, S.J., Quinn, R. and MacKinnon, D. (1991b). Preservation of fossil biopolymeric structures: conclusive immunological evidence. *Geochim. Cosmochim. Acta.*, **55**, 2253-2257.

Conway, D. and Libby, W.F. (1958). The measurement of very slow reaction rates; decarboxylation of alanine. *J. Am. Chem. Soc.*, **80**, 1077-1084.

Cornish-Bowden, A. (1979). How reliably do amino acid composition comparisons predict sequence similarities between proteins? *J. Theor. Biol.*, **76**, 369-386.

Cornish-Bowden, A. (1983). Relating proteins by amino acid composition. *Meth. enzymology.*, **91**, 60-75.

Connor, S. (1988). Genetic fingers in the forensic pie. *New Scientist,* **117**, no. 1597, 31-32.

Crick, F.H.C., Barnet, H.L., Brenner, S. and Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. *Nature*, **192**, 1227-1232.

Cronin, J.R., Gandy, W.E. and Pizzarello, S. (1980). Amino acids of the Murchison meteorite. *In* Hare, P.E., Hoering, T.C. and King, K. Jr. (eds.): *Biogeochemistry of the Amino Acids*, **153-168**. John Wiley and Sons, New York. 558pp.

Curry, G.B. (1987). Molecular palaeontology: new life for old molecules. *Trend. Ecol. Evol.*, **2**, 161-165.

Curry, G.B. (1988). Amino acids and proteins from fossils. *In* Runnegar, B. and Schopf, J.W. (eds.): *Molecular evolution and the fossil record*, **20-33.** Short courses in Paleontology, 1. Paleontological Society. 167pp.

Curry, G.B. and Ansell, A.D. (1986). Tissue mass in living brachiopods. *In* Racheboeuf, P.R. and Emig, C.C. (eds.): *Les brachiopodes fossiles et actuels*, **231-242.** Biostratigraphie du Pelézoique 4, 500pp.

Curry, G.B., Ansell, A.D., James, M. and Peck, L. (1989). Physiological constraints on living and fossil brachiopods. *Trans. R. Soc. Edin.: Ear. Sci.*, **80**, 255-262

Curry, G.B., Cusack, M., Endo, K., Walton, D. and Quinn, R. (1991a). Intracrystalline molecules from brachiopod shells. *In* Suga, S. and Nakahara, H. (eds.): *Mechanisms and phylogeny of mineralization in biological systems*, **35-40.** Springer-Verlag, Tokyo. 517pp.

Curry, G.B., Cusack, M., Walton, D., Endo, K., Clegg, H., Abbott, G. and Armstrong, H. (1991b). Biogeochemistry of brachiopod intracrystalline molecules. *Phil. Trans. R. Soc.*

Lond., **333B**, 359-366.

Curry, G.B., Quinn, R., Collins, M.J., Endo, K., Ewing, S., Muyzer, G. and Westbroek, P. (1991c). Immunological responses from brachiopod skeletal macromolecules; a new technique for assessing taxonomic relationships using shells. *Lethaia*, **24**, 399-407.

Cusack, M., Curry, G.B., Clegg, H. and Abbott, G. (1992). An intracrystalline chromoprotein from red brachiopod shells: implications for the process of biomineralisation. *Comp. Biochem. and Physiol.*, **102B**, 93-95.

Davis, J.C. (1973). *Statistics and data analysis in geology.* John Wiley and Sons, New York. 550pp.

Degens, E.T., Spencer, D.W. and Parker, R.H. (1967). Paleobiology of molluscan shell proteins. *Comp. Biochem. Physiol.*, **20**, 553-579.

De Jong, E.W., Westbroek, P., Westbroek, J.F., and Bruning, J.W. (1974). Preservation of antigenic properties of macromolecules over 70 Myr. *Nature*, **252**, 63-64.

Devereux, J.R., Haeberli, P. and Smithies, O. (1984). A comprhensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387-395.

Dupont, D.R., Keim, P.S., Chui, A., Bello, R., Bozzini, M., and Wilson, K.J. (1989). A comprehensive approach to amino acid analysis. *In* Hugli, T.E. (ed.) *Techniques in protein chemistry*, **284-294**, Academic Press, New York, 612pp.

Dunning, H.N. (1963). Geochemistry of organic pigments. *In* Breger, I.A. (ed.) *Organic geochemistry*, **9**, 367-430.

Eglinton, G. and Logan, G.A. (1991). Molecular preservation. *Phil. Trans. R. Soc. Lond.*, **333B**, 315-328.

Endo, K. and Curry, G.B. (1991). Molecular and morphological taxonomy of a Recent brachiopod genus *Terebratulina*. *In* MacKinnon, D.I., Lee, D.E. and Campbell, J.D. (eds). *Brachiopods through time.* A.A. Belkema, Rotterdam, 447pp.

Endo, K. (1992). *Molecular systematics of Recent and Pleistocene brachiopods.* Unpublished PhD thesis, University of Glasgow, U.K., 207pp.

Ekstrom, A., Fookes, C.J.R., Hambley, T., Loeh, H.J., Miller, S.A. and Taylor, J.C. (1983). Determination of the crystal structure of a petroporphyrin isolated from oil shale.

Nature, **306**, 173-174.

Fleming, C.A. (1953). The geology of the Wanganui Subdivision. *New Zealand Geological Survey, Bulletin 52*. Department of Scientific and Industrial Research, Wellington. 362pp.

Furth, A. (1988). Sweet peril for proteins. *New Scientist*, 3rd March 1988, 58-62.

Gage, M (1945). The Tertiary and Quaternary geology of Ross, Westland. *Trans. R. Soc. NZ.*, **75**, 138-159.

Geiger, R. and Fritz, H. (1983). Trypsin. *In* Bergmeyer, H.U. (ed.) *Methods of enzymatic analysis. V. Enzymes 3: Peptidases, proteinases and their inhibitors*, **119-129**, Weinheim, Verlag Chemie.

Golenberg, E.M., Giannasi, D.E., Clegg, M.T., Smiley, C.J., Durbin, M., Henderson, D. and Zurawski, G. (1990). Chloroplast DNA sequences from a Miocene *Magnolia* species. *Nature*, **344**, 656-658.

Gross, E. and Witkop, B. (1961). Selective cleavage of the methionyl peptides bonds in ribonuclease with cyanogen bromide. *J. Am. Chem. Soc.*, **83**, 1510-1511

Hamilton, P.B. (1965). Amino acids on hands. *Nature*, **205**, 284-285.

Hare, P.E. and Abelson, P.H. (1968). Racemization of amino acids in shells.*Carnegie. Inst. Wash. Yb.*, **66**, 526-528.

Hare, P.E. and Mitterer, R.M. (1967). Non protein amino acids in fossil shells. *Carnegie Inst. Wash. Yb.*, **65**, 362-364.

Hare, P.E and Mitterer, R.M. (1969). Laboratory simulation of amino acid diagenesis in fossils. *Carnegie Inst. Wash .Yb.*, **67**, 205-208.

Harris, E.L.V. and Angal, S. (1989). *Protein purification method: a practical approach* (Rickwood, D. and Hames, B.D. series eds). IRL Press, Oxford. 317pp.

Haugen, J.-E., Sejrup, H.-P. and Vogt, N.B. (1989). Chemotaxonomy of Quaternary benthic foraminifera using amino acids. *J. Foram. Res.*, **19**, 38-51.

Hedges, J.I., Cowie, G.L., Ertel, J.R., Barbour, R.J. and Hatcher, P.G. (1985). Degrada-

tion of carbohydrates and lignins in buried woods. *Geochim. Cosmochim. Acta.*, **49**, 701-711.

Heinrikson, R.L. and Meredith, S.C. (1984). Amino acid analysis by reverse-phase high-performance liquid chromatography: precolumn derivatization with phenylisothiocyanate. *Anal. Biochem.*, **136**, 65-74.

Ho, T-Y. (1966) Stratigraphic and paleographic applications of water-insoluble fraction of residual shell proteins in fossil shells. *Geol. Soc. Am. Bull.*, **77**, 375-392.

Hoering, T.C. (1973). A comparison of melanoidin and humic acid. *Carnegie. Inst. Wash. Yearb.*, **73**, 590-595.

Hoering, T.C. (1980). The organic constituents of fossil mollusc shells. *In* Hare, P.E., Hoering, T.C. and King, K. (eds.). *Biogeochemistry of the amino acids.* **193-201.** John Wiley and Sons, New York. 558pp.

Horai, S., Kando, R., Murayama, K., Hayashi, S. Koike, H. and Nakai, N. (1991). Phylogenetic affiliation of ancient and contemporary humans inferred from mitochondrial DNA. *Phil. Trans. R. Soc. Lond*, **333B**, 409-417.

Hudson, J.D. (1967). The elemental composition of the organic fraction, and the water content of some Recent and fossil mollusc shells. *Geochim. Cosmochim. Acta.*, **31**, 2361-2378.

Ikawa, M. and Snell, E.E. (1954). Oxidative deamination of amino acids by pyidoxal and metal salts. *J. Am. Chem. Soc.*, **76**, 4900-4902.

Johnson, S.C. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**, 241-254.

Jones, J.D. and Vallentyne, J.R. (1960). Biogeochemistry of organic matter-I. *Geochim. Cosmochim. Acta.*, **21**, 1-34.

Jope, H.M. (1965). Composition of brachiopod shell. *In* Moore, R.C. (ed.) *Treatise on invertebrate palaeontology, Part H, Brachiopoda.* **H156-164.** Geol. Soc. Am. and Univ. Kansas Press. 927pp.

Jope, M. (1967a). The protein of brachiopod shell-I. Amino acid composition and implied protein taxonomy. *Comp. Biochem. Physiol.*, **20**, 593-600.

Jope, M. (1967b). The protein of brachiopod shell-II. Shell protein from fossil articulates: amino acid composition. *Comp. Biochem. Physiol.* **20**, 601-605.

Jope, M. (1973). The protein of brachiopod shell-V. N-terminal end groups. *Comp. Biochem. Physiol.*, **45B**, 17-24.

Jope, M. (1977). Brachiopod shell proteins: Their functions and taxonomic significance. *Amer. Zool.*, **17**, 133-140.

Jope, M. (1979). The protein of brachiopod shell-VI. C-terminal end groups and sodium dodecyl sulphate-polyacrylamide gel electrophoresis: molecular constitution and structure of the protein. *Comp. Biochem. Physiol.*, **63B**, 163-173.

Jope, M. (1980). Phylogenetic information derivable from fossil brachiopods. *In* Hare, P.E., Hoering, T.C. and King, K. (eds.). *Biogeochemistry of the amino acids.* **83-94.** John Wiley and Sons, New York. 558pp.

Kahne, D. and Still, W.C. (1988). Hydrolysis of a peptide band in neutral water. *J. Am. Chem. Soc.*, **110**, 7529-7534.

King, K. Jr. and Hare, P.E. (1972). Amino acid composition of the test as a taxonomic character for living and fossil planktonic foraminifera. *Micropalaeontology*, **18**, 285-293.

Kitano, Y. and Hood, D.W. (1965). The influence of organic material on the polymorphic crystallization of calcium carbonate. *Geochim. Cosmochim. Acta.*, **29**, 29-41.

Klapper, M.H. (1977). The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochem. Biophys. Res. Comm.*, **78**, 1018-1024.

Koch, G.S. and Link, R.F. (1971). *Statistical analysis of geological data. Volume II.* John Wiley and Sons, New York. 438p.

Kolesnikov, C.M. and Prosorovskaya, E.L. (1986). Biochemical investigations of Jurassic and Recent brachiopod shells. *In* Racheboeuf, P.R. and Emig, C.C. (eds.) Les brachiopodes fossiles et actuels. *Biostratigraphie du Pelézoique* **4**, 113-119. 500pp.

Kriesten, K., Egge, H. and Faust, R. (1979). Lipids in the egg shell of the ostrich (*Struthio camelus*). *Experimentia*, **35**, 1032-1033.

Lance, G.N. and Williams, W.T. (1968). New information - statistic classificatory

program. *Computer J.*, **11**, 195.

Liu, T.-Y. and Chang, Y.H. (1971). Hydrolysis of proteins and peptides with ρ-toluene sulfonic acid. Determination of tryptophan. *J. Biol. Chem.*, **246**, 2842-2848.

Logan, G.A., Collins, M.J. and Eglinton, G. (1991). Preservation of organic biomolecules. *In* Allison, P.A. and Briggs, D.E.G. (eds). *Taphonomy: releasing the data locked in the fossil record*, **1-24**. Plenum Press, New York. 553pp.

Lowenstam, H.A. (1981). Minerals formed by organisms. *Science*, **211**, 1126-1131.

Lowenstam, H.A. and Weiner, S. (1989). *On Biomineralization.* Oxford University Press, New York. 324pp.

Lowenstein, J.M. (1980). Species specific proteins in fossils. *Naturwissenschaften.*, **67**, 343-346.

Lowenstein, J.M. and Scheuenstuhl, G. (1991). Immunological methods in molecular palaeontology. *Phil. Trans. R. Soc. Lond.*, **333B**, 375-380.

MacKinnon, D.I. and Williams, A. (1974). Shell structure of Terebratulid brachiopods. *Palaeontology*, **17**, 179-202.

Maillard, L.C. (1913). Formation de matieres humiques par action de polypeptides sur sucres. *C. r. Acad. Sci.*, **156**, 148-159.

Mann, S. (1983). Mineralization in biological systems. *Struct. Bonding.*, **54**, 125-1

Matsubara, H and Sasaki, R.M. (1969). High recovery of tryptophan from acid hydrolysates of proteins. *Biochem. Biophys. Res. Comm.*, **35**, 175-181.

Meisler, A. (1965). *Biochemistry of the amino acids:* Academic Press, New York.

Meyers, P.A., Keenheer, M.J., Erstfeld, K.M. and Bourbonniere, R.A. (1980). Changes in spruce composition following burial in lake sediments for 10 000 years. *Nature*, **287**, 534-535.

Mitterer, R.M. (1968). Amino acid composition of organic matrix in calcareous oolites. *Science*, **162**, 1498-1499.

Mobil Oil Corporation (1971). *Reflection seismic and magnetic survey, licence 688, offshore North Island, New Zealand.* Unpublished, held as open file petroleum report 588, by New Zealand Geological Survey, Lower Hutt.

Moore, S. and Stein, W.H. (1948). Photometric ninhydrin method for use in the chromatography of amino acids. *J. Biol. Chem.*, **176**, 388.

Moore, S. and Stein, W.H. (1963). Chromatographic determination of amino acids by the use of automatic recording equipment. *Methods in Enzymology*, **6**, 819-831.

Murray, K., Rasmussen, P.S., Neustadter, J. and Luck, J.M. (1965). The hydrolysis of arginine. *J. Biol. Chem.*, **240**, 705-709.

Muyzer, G., Westbroek, P. and Wehmiller, J.F. (1988). Phylogenetic implications and diagenetic stability of macromolecules from Pleistocene and Recent shells of Mercenaria mercenaria (Mollusca, Bivalvia). *Hist Biol.*, **1**, 135-144.

Nickl, H.J. and Henisch, H.K. (1969). Growth of calcite crystals in gels. *J. Electrochem. Soc.*, **116**, 1258-1260.

Norris, R.M. and Grant-Taylor, T.L. (1989). Late Quaternary shellbeds, Western Shelf, New Zealand. *NZ J Geol. Geophys.*, **32**, 343-356.

Oró, J. and Skewes, H.B. (1965). Free amino acids on Human fingers: the question of contamination in microanalysis. *Nature*, **207**, 1042-1045.

Ourisson, G., Albrecht,P. and Rohmer, M. (1979). The hopanoids. Palaeochemistry and biochemistry of a group of natural product. *Pure. Appl. Chem.*, **51**, 709-729.

Ourisson, G., Albrecht,P. and Rohmer, M. (1984). The microbial origin of fossil fuels. *Sci. Am.*, **251**, 34-41.

Patterson, C. (1987). *Molecules and morphology in evolution: conflict or compromise?* Cambridge University Press, Cambridge, 229pp.

Penke, B., Ferenczi, R. and Kovács, K. (1974). A new acid hydrolysis method for determining tryptophan in peptides and proteins. *Anal. Biochem.*, **60**, 45-50.

Pilaar, W.F.H. and Wakefield, L.L. (1978). Structural and stratigraphic evolution of the Taranaki Basin, offshore North Island, New Zealand. *APEA J.*, **18**, 93-101.

Pillans, B. (1983). Upper Quaternary marine terrace chronology and deformation, South Taranaki, New Zealand. *Geology*, **11**, 292-297.

Popp, B.N., Podosek, F.A., Brannon, J.C., Anderson, T.F., and Pier, J. (1986). $^{87}Sr/^{86}Sr$ ratios in Permo-Carboniferous sea water from the analyses of well preserved brachiopod shells. *Geochim. Cosmochim. Acta.*, **50**, 1321-1328.

Prager, E.M. and Wilson, A.C. (1971). The dependence of immunological cross reactivity upon sequence resemblance among Lysozymes. I microcomplement fixation studies. *J.Biol. Chem.*, **246**, 5978-5989.

Quinn, R. (1990). *Brachiopod serotaxonomy and the application of immunology to geology*. Unpublished PhD thesis, University of Glasgow, U.K. 145pp.

Robinson, A.B. and Rudd, C.J. (1974) Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. *Cur. Top. Cell. Reguln.*, **8**, 247-294.

Robbins, L.L. and Brew, K. (1991). Proteins from the organic matrix of core-top and fossil planktonic foraminifera. *Geochim. Cosmochim. Acta.*, **54**, 2285-2292.

Rohlf, F.J. (1968). Stereograms in numerical taxonomy. *Systematic Zool.*, **17**, 246-255.

Rudwick, M.J.S. (1970). *Living and fossil brachiopods*. Hutchinson and Co., London, 199pp.

Runnegar, B. (1986). Molecular palaeontology. *Palaeontology*, **29**, 1-24.

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

Schroeder, R.A. (1975). Absence of β-alanine and γ-aminobutyic acid in cleaned foraminiferal shells : implications for use as a chemical criterion to indicate removal of non - indigenous amino acid contaminants. *Earth. Planet. Sci. Lett.*, **25**, 274-278.

Schroeder, R.A. and Bada, J.L. (1976). A review of the geochemical applications of the amino acid racemization reaction. *Ear. Sci. Rev.*, **12**, 347-391.

Seward, D. (1974). Age of New Zealand Pleistocene substages by fission-track dating of

glass shards from tephra horizons. *Ear. Planet. Sci. Lett.*, **24**, 242-248.

Shackleton, N.J. and Updyke, N.D. (1973). Oxygen isotope and palaeomagnetic stratigraphy of equatorial paccific core V-28-238: oxygen isotope temperatures and ice volumes on a 10^5 year and 10^6 year scale. *Quat. Res.*, **3**, 39-55.

Shewry, P.R., Kirkman, M.A., Burgess, S.R., Festenstein, G.N. and Miflin, B.J. (1982). A comparison of the protein and amino acid composition of old and recent barley grain. *New Phytologist.*, **90**, 455-466.

Sneath, P.H.A. and Sokal, R.R. (1973). *Numerical Taxonomy*. W.H. Freeman and Company, San Francisco. 573p.
Spackman, D.H., Stein, W.H. and Moore, S. (1958). *Anal. Chem.*, **30**, 1190-1206.

Steinberg, S.M. and Bada, J.L. (1983). Peptide decomposition in the neutral pH region via the formation of diketopiperazines. *J. Org. Chem.*, **48**, 2295-2298.

Stern, T.A. and Davey, F.J. (1990). Deep seismic expression of a foreland basin: Taranaki basin, New Zealand. *Geology*, **18**, 979-982.

Stoyanova, R. Zh. (1984). Alkane and fatty acid content and composition in Palaeozoic brachiopods. *Dokl. Bolg. Akad. Nauk.*, **37**, 771-774.

Stryer, L. (1988). *The molecular design of life*. W.H. Freeman and Company, New York. 215pp.

Sucov, H.M., Benson, S., Robinson, J.J., Britten, R.J., Wilt, F. and Davidson, E.H. (1987). A lineage specific gene encoding a major matrix protein of the sea urchin embryo spicule. II Structure of the gene and derived sequence of the protein. *Devel. Biol.*, **120**, 507-519.

Summons, R.E. (1988). Biomarkers: molecular fossils. *In* Runnegar, B. and Schopf, J.W. (eds.) *Molecular evolution and the fossil record*, **98-113.** Short courses in Paleontology, 1. Paleontological Society. 167pp.

Swadesh, J.K., Thannhauser, T.W. and Scheraga, H.A. (1984). Sodium sulfite as an antioxidant in the acid hydrolysis of bovine pancreatic ribonuclease A. *Anal. Biochem.*, **14**, 397-401.

Tan-Wilson, A.L. (1983). Alanine biosynthesis and the general transaminases. *In* Hermann, K.M. and Somerville, S.L. (eds.). *Amino acids: Biosythesis and genetic regulation.*

Addison-Wesley Publishing Company, Reading, Massachusetts. 453pp.

Tsugita, A. and Scheffler, J.J. (1982). A rapid method for acid hydrolysis of protein with a mixture of trifluoroacetic acid and hydrochloric acid. *Eur. J. Biochem.*, **124**, 585-588.

Totten, D.K. Jr., Davidson, F.D. and Wyckoff, R.W.G. (1972). Amino acid composition of heated oyster shells. *Proc. Nat. Acad. Sci. USA.*, **69**, 784-785.

Towbin, H., Stahelin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Nat. Acad. Sci. USA*, **76**, 4350-4354.

Towe, K.M. (1980). Preserved organic ultrastructure: an unreliable indicator for Palaeozoic amino acid biogeochemistry. *In* Hare, P.E., Hoering, T.C. and King, K. (eds.). *Biogeochemistry of the amino acids*, **65-74**. John Wiley and Sons, New York. 558pp.

Towe, K.M. and Thompson, G.R. (1972). The structure of some bivalve shell carbonates prepared by ion-beam thinning. *Calc. Tiss. Res.*, **10**, 38-42.

Treibs, A. (1934). Organic mineral substances II. Occurrence of chlorophyll derivatives in an oil shale of the Upper Triassic. *Ann. Chem.*, **509**, 103-114.

Van Kleef, F.S.M., de Jong, W.W., and Hoenders, H.J. (1975). Stepwise degredations and deamidation of the eye lens protein α-crystallin in ageing. *Nature*. **258**, 264-266.

Vallentyne, J.R. (1964). Biogeochemistry of organic matter II: Thermal reaction kinetics and transformation products of amino compounds. *Geochim. Cosmochim. Acta.*, **28**, 157-188.

Vallentyne, J.R. (1968). Pyrolysis of proline, leucine, arginine and lysine in aqueous solution. *Geochim. Cosmochim. Acta.*, **32**, 1353-1356.

Veis, A. (1989). Biochemical studies of vertebrate tooth mineralization. *In* Weiss, S., Webb, J. and Williams, R.J.P. (eds.) *Biomineralization*, **189-222**, VCH Publishers, Weinheim.

Walton, D. and Curry, G.B. (1991). Amino acids from fossils, facies and fingers. *Palaeontology*, **34**, 851-858.

Watabe, N. (1963). Decalcification of thin sections for electron microscope studies of

crystal-matrix relationships in mollusc shells. *J. Cell. Biol.*, **18**, 701-703.

Watson, J.D. and Crick, F.H.C. (1953). Molecular structure of nucleic acid. A structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.

Weiner, S. (1983). Mollusk shell formation: isolation of two organic matrix proteins associated with calcite deposition in the bivalve *Mytilus californianus*. *Biochemistry*, **22**, 4139-4145.

Weiner, S. (1984). Organization of organic matrix components in mineralized tissues. *Am. Zool.*, **24**, 945-951.

Weiner, S. and Lowenstam, H.A. (1980). Well preserved mollusk shells: characterization of mild diagenetic processes. *In* Hare, P.E., Hoering, T.C. and King, K. (eds.): *Biogeochemistry of the amino acids*, **95-114** John Wiley and Sons, New York. 558pp.

Weiner, S., Lowenstam, H.A., and Hood, L. (1976). Characterisation of 80-million-year-old mollusk shell proteins. *Pro. Nat. Acad. Sci. U.S.A.*, **73**, 2541-2545.

Westbroek, P., van der Meide, P.H., van der Wey-Kloppers, van der Sluis, R.J., de Leeuw, J.W. and de Jong, E.W. (1979). Fossil macromolecules from cephalopod shells, characterization, immunological response and diagenesis. *Paleobiology*, **5**, 151-167.

Westall, F. and Hesser, H. (1974). Fifteen minute hydrolysis of peptides. *Anal. Biochem.*, **61**, 610-613.

Williams, A. *et al* (1965). Brachiopoda. *In* Moore, R.C. (ed.) *Treatise on invertebrate palaeontology, Part H, Brachiopoda.* Geol. Soc. Am. and Univ. Kansas Press. 927pp.

Williams, A and Rowell, A.J. (1965a). Evolution and phylogeny. *In* Moore, R.C. (ed.) *Treatise on invertebrate palaeontology, Part H, Brachiopoda.* **H164-199**. Geol. Soc. Am. and Univ. Kansas Press. 927pp.

Williams, A and Rowell, A.J. (1965b). Morphology. *In* Moore, R.C. (ed.) *Treatise on invertebrate palaeontology, Part H, Brachiopoda.* **H57-155**. Geol. Soc. Am. and Univ. Kansas Press. 927pp.

Williams, A. (1968). *Evolution of the shell structure of articulate brachiopods. Special Papers in Palaeontology, No. 2.* Palaeontological Association, London. 55pp.

Williams, A. (1984). Lophophorates. *In* Bereiter-Hahn, J, Matolsty, A.G. and Richards, K.S. (eds.) *Biology of the integument. I Lophophorates*, **728-745**, Springer-Verlag, Berlin.

Williams, D.F., Moore, W.S. and Fillon, R.H. (1981). Role of glacial Arctic Ocean ice sheets in Pleistocene oxygen isotope and sea level records. *Earth. Planet. Sci. Lett.*, **56**, 157-166.

Wilson, H. and Cannon, R.K. (1937). The glutamic acid-pyrrolidone carboxylic acid system. *J. Biol. Chem.*, **119**, 309-331.

Worms, D. and Weiner, S. (1986). Mollusk shell organic matrix: fourier transform infrared study of the acid macromolecules. *J. Exp. Zool.*, **237**, 11-20.

Wyckoff, R.W.G. (1972). *The biochemistry of animal fossils*. Scientechnia, Bristol. 151pp.

Yokote, Y. Arai, K.M. and Akahane, K. (1986). Recovery of tryptophan from 25 minute acid hydrolysates of protein. *Anal. Biochem.*, **152**, 245-249.
Zagalsky, P.F. (1976). Carotenoid-protein complexes. *Pure appl. Chem.*, **47**, 103-120.

Zeven, A.C., Doukes, G.J. and Kislev, M. (1975). Proteins in old Grains of *Triticum sp.* *J. Arch. Sci.*, **2**, 209-213.