PROBLEMS FOR TAXONOMIC ANALYSIS USING INTRACRYSTALLINE AMINO ACIDS: AN EXAMPLE USING BRACHIOPODS

by DEREK WALTON

ABSTRACT. Multivariate statistical analysis of the absolute abundance of amino acids extracted from the intracrystalline sites of brachiopods has the potential for constructing a molecular phylogeny. In all cases, separation of the brachiopods was possible to subordinal level and in some cases to subfamilial level. Older samples showed a merging of closely related genera, indicating the loss of specificity caused by the degradation of amino acids. Amino acid data alone are therefore not sufficient for molecular taxonomy in fossils; the degradative pathways should be sought to allow reconstruction of the original amino acid content.

THE use of proteins and amino acids to differentiate between Recent taxa is an established technique in taxonomic analysis (e.g. Dussart 1983). Mutations in the DNA may result in changes in the primary sequence of a protein and this is reflected in the relative abundance of the amino acids. Speciation is marked by a deviation of the amino acid composition. One of the stated long-term aims of molecular palaeontology is the establishment of a molecular phylogeny through the direct sequencing of fossil peptides and comparison with the sequence in Recent organisms (Curry 1988). Although this approach may have a great deal of value (Cohen 1994), the reality is, however, not straightforward. There have been very few reports of the sequencing of proteins from the shells of organisms (Sucov *et al.* 1987; Robbins and Donachy 1991; Cusack *et al.* 1992) and this paucity of sequence information for shell proteins makes comparisons with information from the fossil record difficult.

Consequently, the use of proteins from the fossil record as a taxonomic tool is restricted, even though their remains occur in the shells and bones of a wide range of organisms and their persistence is well documented (e.g. Abelson 1954; Jope 1967; Wyckoff 1972; Weiner *et al.* 1976; Collins *et al.* 1991; Kaufman *et al.* 1992). It has long been recognized that the original proteins are degraded over time through peptide bond degradation to form mixtures of smaller peptides which are so complex as to defy further purification in most circumstances (Abelson 1954, 1955; Akiyama 1971; Hare and Hoering 1977; Armstrong *et al.* 1983; Qian *et al.* 1995; Walton 1996, in press; cf. Robbins and Brew 1990). 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 and thus the molecular phylogeny could not be completed. As amino acid substitutions only affect relatively few sites in proteins (Cusack *et al.* 1992), it is likely that these changes would not be observed in fossil peptides.

Decomposition of proteins releases amino acids, and a number of studies have demonstrated that phylogenetic information is recoverable through statistical analysis of the amino acid composition of Recent (e.g. Degens *et al.* 1967; Cornish-Bowden 1979, 1983; MacFie *et al.* 1988; Robbins and Healy-Williams 1991; Walton *et al.* 1993) and fossil (King and Hare 1972; Haugen *et al.* 1989; Robbins and Brew 1990; Kaufman *et al.* 1992; Walton 1996) samples. However, the analysis of fossil proteinaceous remains is hindered as the amino acids undergo severe degradation with the loss of information from the shell, and a subsequent decrease in specificity in the analysis (e.g. Hare and Mitterer 1969; Hare 1974; Robbins and Donachy 1991; Kaufman *et al.* 1992; Walton in press).

Although intracrystalline proteins (*sensu* Sykes *et al.* 1995), are protected by the inorganic phase

[Palaeontology, Vol. 41, Part 4, 1998, pp. 753–770] © The Palaeontological Association

(Towe 1980; Collins *et al.* 1988) they are also highly degraded (Collins *et al.* 1991; Walton 1996), thus ensuring that it is unlikely that meaningful sequence data can be resolved from fossil organisms. However, intracrystalline amino acids retain phylogenetic information, as the carbonate of the shell approximates to a closed system (Collins *et al.* 1988; Albeck *et al.* 1993; Walton *et al.* 1993) and thus leaching should not occur. This is in contrast to the more open intercrystalline sites that are prone to leaching of material from the shell (Sykes *et al.* 1995). The residual amino acids and peptides recovered from intracrystalline sites are remnants of the original protein and may be examined in the same way as those extracted from Recent samples (Walton 1996). For amino acids to be of value in the taxonomy of fossils, it is essential that degradative patterns are recognized and that amino acids are extracted from the most protected sites.

The aim of this study was threefold: (1), to undertake multivariate statistical analysis of the amino acid composition of intracrystalline molecules extracted from fossil brachiopods; (2), to demonstrate that taxonomically relevant information can be retrieved despite the degradation of the proteins and amino acids; (3), to highlight potential problems in taxonomic analysis using amino acids and to suggest ways in which such analyses might be refined. The amino acid compositions of these brachiopods and their degradative pathways will be discussed elsewhere (Walton in press) and are not considered in great detail here. This study is concerned with the application of the data to palaeontological analysis.

MATERIALS AND METHODS

Sample collection

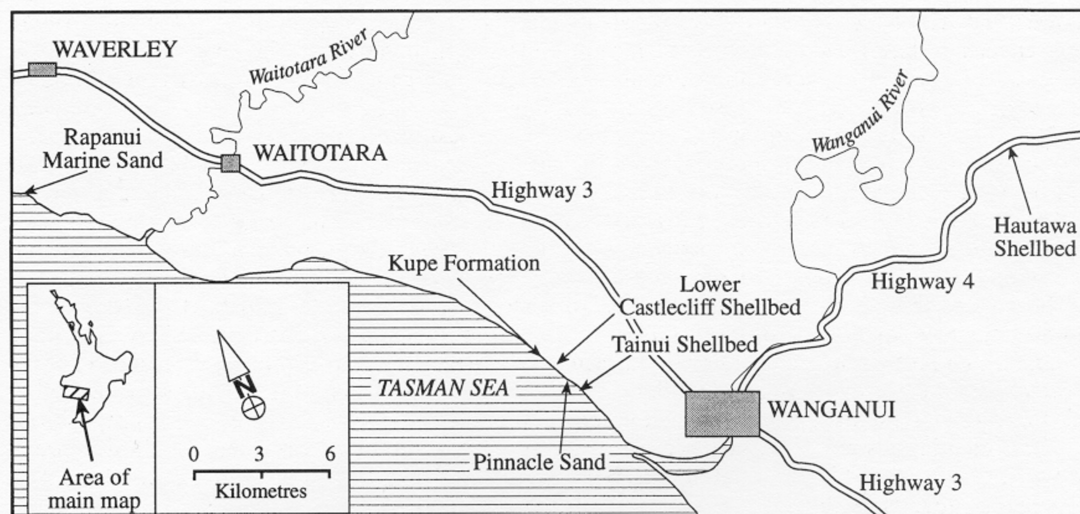
Samples of brachiopods (*Neothyris lenticularis, Calloria inconspicua, Terebratella sanguinea* and *Notosaria nigricans*) and molluscs (turratellids and pectenids) were collected from the rich and diverse fauna of the South Wanganui Basin, North Island, New Zealand (Text-fig. 1; Table 1). These samples contain intracrystalline proteins and amino acids which have been partially characterized (Cusack *et al.* 1992; Walton *et al.* 1993; Walton and Curry 1994; Walton 1996, in press), and have proved to be near-ideal for the investigation of fossil macromolecules as their shells are composed of diagenetically stable low-Mg calcite. Molluscs were collected from the shell beds to act as outgroups in the analysis and to ensure that similarities in the data were due to taxonomic similarities, rather than the homogenization of the amino acid content through the shell bed.

The tectonic setting of the South Wanganui Basin (a back-arc basin) has allowed rapid subsidence and the accumulation of up to 4 km of sediments, most deposited in shallow marine conditions (Anderton 1981), although estuarine and terrestrial facies are recorded (Fleming 1953). Interspersed throughout the sedimentary sequence are a number of richly fossiliferous shell beds containing abundant macrofossils, ranging in age from 120 Ka to *c.* 2·6 Ma.

Sample preparation

Samples were prepared according to the methods of Walton and Curry (1994), in which shells that were excessively bored or fractured were excluded from further study. Adhering sediment was scrubbed from the sample and encrusting epifauna removed by scraping. Articulated shells were disarticulated and body tissue (only present in Recent samples) removed before being incubated in an aqueous solution of bleach (10 per cent. v/v) for 2 hours at room temperature, washed extensively with Milli RO® water (Millipore) and air dried. Samples were ground using a ceramic pestle and mortar, and the powder incubated in an aqueous solution of bleach (10 per cent. v/v) under constant motion for 24 hours at room temperature, then washed by repeated agitation with MilliQ® water (Millipore) and centrifugation (typically ten washes) and lyophilized.

An aqueous solution of HCl (2M) at a ratio of 11 μl/mg was used to dissolve the shell powder and release the incarcerated biomolecules. Once demineralization was complete, insoluble particles were removed by centrifugation (20 g.h.). All samples were hydrolysed by vapour-phase HCl (6N) automated hydrolysis (Applied Biosystems 420A; Dupont *et al.* 1989). Standard proteins and peptides were used during every analysis to ensure that hydrolysis proceeded to completion. Blank



TEXT-FIG. 1. Locations of the horizons from which samples were collected (adapted from Fleming 1953).

TABLE 1. Locations of samples utilized in this study. Grid references correspond to the maps accompanying Fleming (1953).

Horizon	Location	Grid reference
Rapanui Marine Sand	Waipipi Beach	N137/168 993
Tainui Shellbed	Castlecliff Beach	N137/485 888
Pinnacle Sand	Castlecliff Beach	N137/479 895
Lower Castlecliff Shellbed	Castlecliff Beach	N137/470 902
Kupe Formation	Castlecliff Beach	N137/459 908
Hautawa Shellbed	Parapara Road	N138/803 029

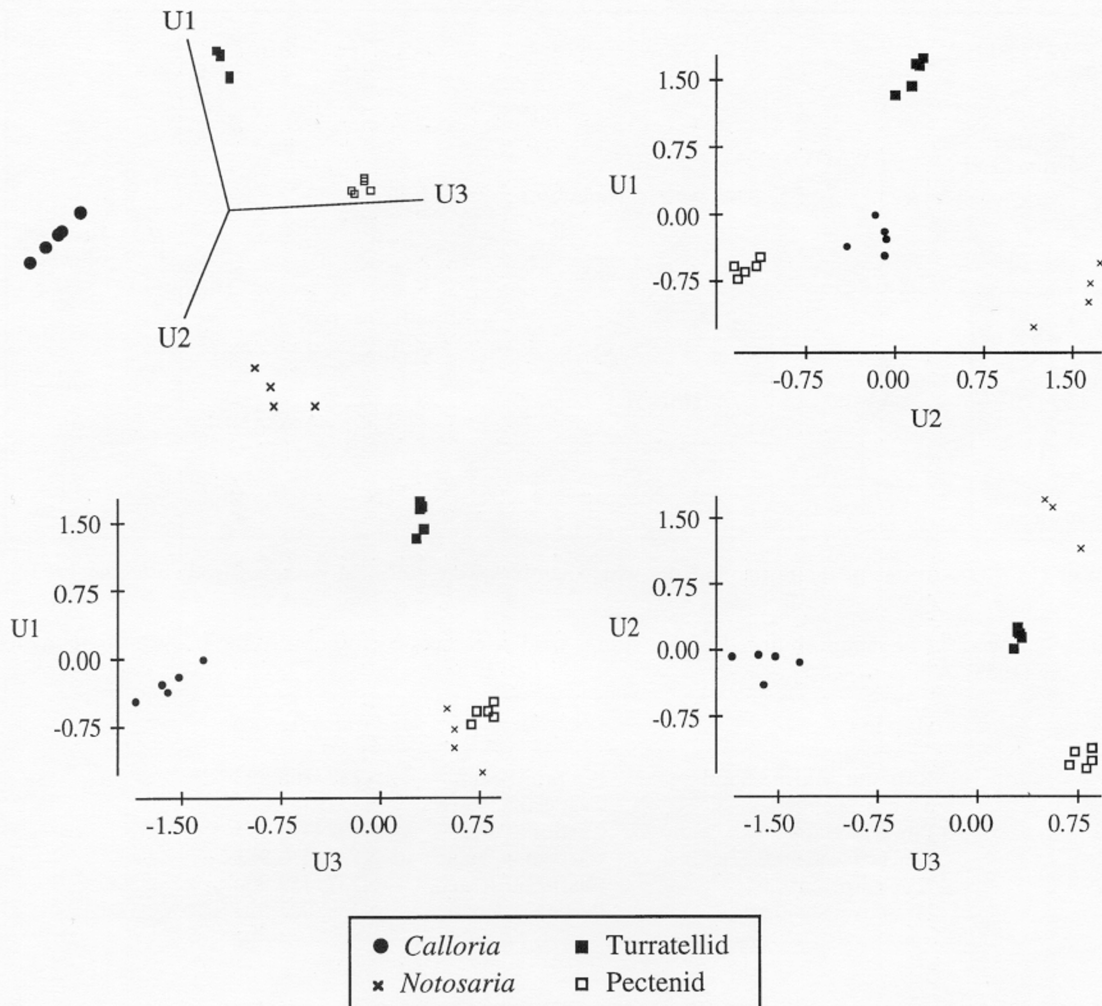
analyses were included to check for background levels of contamination. Individual amino acids were derivatized using phenylisothiocyanate (Heinrikson and Meredith 1984), and transferred to a dedicated narrowbore hplc system for separation and quantification. Analyses were repeated at least three times. The data were subjected to principal components analysis (PCA; Davis 1986) using the statistical analysis program DATADESK™.

It is usual 'to extract only enough eigenvectors to remove the majority, say 75 per cent., of the total variance of the data matrix' Sneath and Sokal (1973, p. 246). From computer calculations, it can be seen that the majority of the variance within the samples can be defined by the first three eigenvectors. This representation of the amino acids in PCA form in three dimensional space is a useful method of comparing multivariate distributions of a larger sample size.

RESULTS

The state of molecular preservation of the intracrystalline proteins and amino acids in these fossils is reported elsewhere (Walton 1996, in press). Proteins are almost completely hydrolysed by 120 Ka and the amino acids have degraded relatively rapidly (although at different rates and by different pathways) over the 2·2 Ma of the study. This degradation of amino acids will lead to changing concentrations of the molecules, therefore changing the data for the PCA (Walton in press). As a consequence, the resolution of the PCA should decrease as samples of increasing age are analysed.

Interpretation was made in two ways, within and between individual horizons, in order to



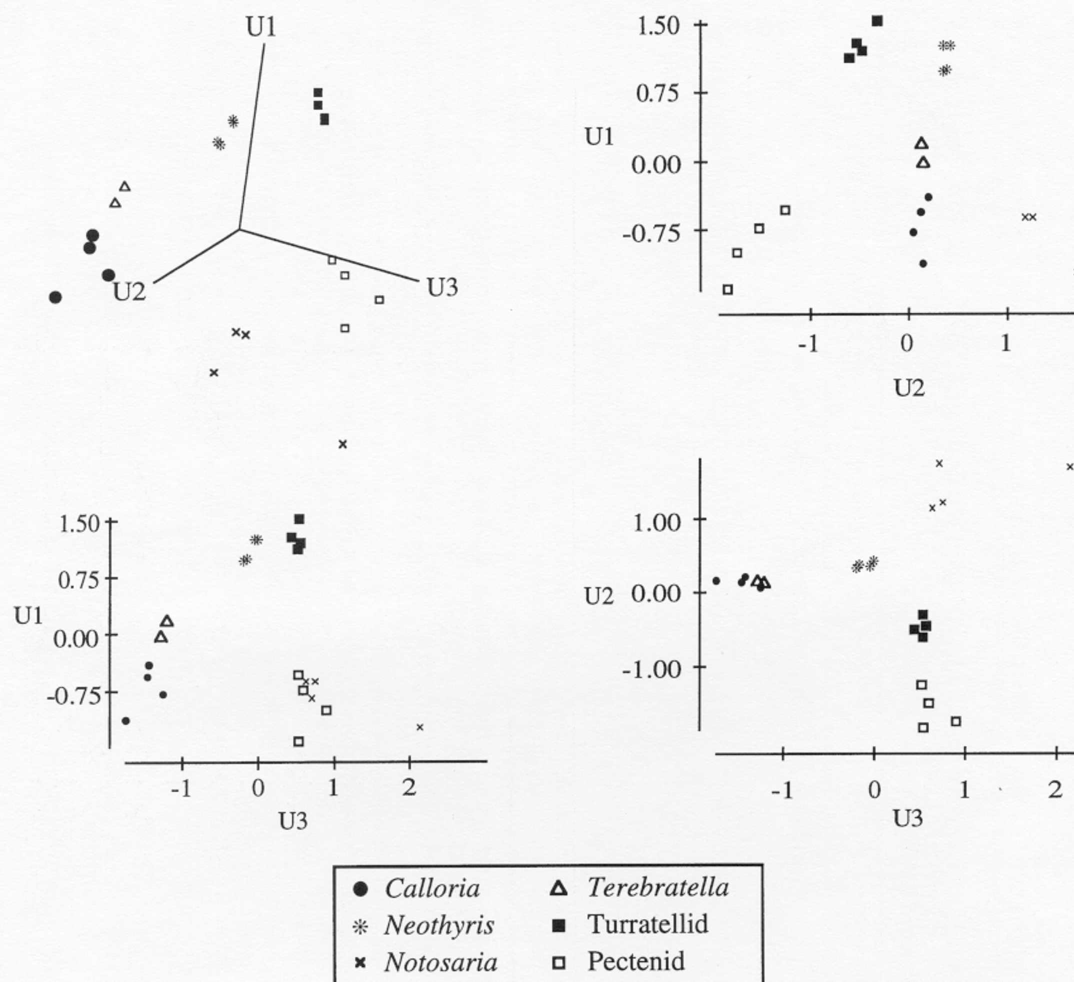
TEXT-FIG. 2. Plots of the first three principal components for the concentration of amino acids from samples collected from the Rapanui Marine Sand. Scatterplots are shown in this and subsequent figures to allow better interpretation of the 3D plot to the left, in which the axes are at 90°. Note the good separation of all data points.

determine how time will affect the separation of groupings identified in Walton *et al.* (1993). As the PCA is derived from a specific dataset (i.e. the amino acid content of fossils from a horizon), graphical representations from each horizon cannot be compared directly (as the information in each diagram is sourced from different data). To compare data from different horizons it is therefore necessary to complete a new PCA including all of the data simultaneously rather than individually.

Samples collected from the same horizon should be of approximately the same age, and will have been subjected to approximately 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, although see Norris and Grant-Taylor (1989) and Wehmiller *et al.* (1995) for discussion of homogeneity in shell beds). Changes in the amino acid content due to diagenetic alteration 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 oversimplification of possible relationships, and the amino acid composition of the

TABLE 2. Principal component analysis calculated from the absolute abundance of amino acids in the sample. Only the first three eigenvectors and eigenvalues are given in each case. NI = data not included for PCA.

	Sample location																	
	Rapanui Marine Sand			Tainui Shellbed			Pinnacle Sand			Lower Castlecliff Shellbed			Kupe Formation			Hautawa Shellbed		
Eigenvalue	**e1**	**e2**	**e3**	**e1**	**e2**	**e3**	**e1**	**e2**	**e3**	**e1**	**e2**	**e1**	**e1**	**e2**	**e3**	**e1**	**e2**	**e3**
Values	7·908	3·824	1·355	6·588	3·931	2·158	6·841	3·515	1·902	8·437	2·901	1·501	10·047	2·894	0·601	5·992	3·423	1·954
Variance Proportion	56·5	27·3	9·7	47·1	28·1	15·4	48·9	25·1	133·6	60·3	20·7	10·7	71·8	20·7	4·3	46·1	26·3	15
Total variance		93·5			90·6			87·6			91·7			96·8			87·4	
Eigenvectors	**V1**	**V2**	**V3**	**V1**	**V2**	**V3**	**V1**	**V2**	**V3**	**V1**	**V2**	**V3**	**V1**	**V2**	**V3**	**V1**	**V2**	**V3**
D	−0·248	0·260	0·407	−0·272	0·209	0·381	−0·327	0·162	0·269	−0·321	−0·106	−0·208	−0·304	0·119	−0·071	−0·329	0·249	0·166
E	**−0·334**	−0·099	0·069	−0·352	−0·172	−0·018	−0·341	−0·143	−0·020	**−0·329**	0·001	0·135	**−0·311**	−0·030	0·129	−0·125	**−0·453**	0·200
S	−0·241	−0·346	0·233	−0·236	−0·374	0·070	−0·333	−0·191	−0·054	−0·313	−0·137	−0·231	−0·233	0·254	0·124	−0·218	−0·205	0·137
G	−0·294	0·241	−0·238	−0·251	0·327	−0·241	−0·071	0·387	−0·406	0·030	**0·549**	−0·196	−0·275	0·258	−0·250	−0·328	−0·144	−0·376
R	−0·271	0·229	0·080	**−0·362**	0·067	−0·115	**−0·363**	−0·085	−0·040	−0·296	0·015	−0·208	−0·299	−0·005	0·236	NI	NI	NI
T	−0·295	−0·041	0·056	−0·167	0·350	0·356	−0·144	**0·427**	0·318	−0·314	−0·008	−0·231	−0·002	**−0·517**	**−0·579**	**−0·356**	0·248	0·090
A	**−0·342**	−0·082	−0·005	**−0·363**	0·121	0·122	**−0·361**	0·164	0·018	−0·326	0·073	−0·185	**−0·309**	0·091	−0·109	**−0·390**	0·089	−0·126
P	−0·251	0·176	**−0·530**	−0·213	0·032	**−0·473**	−0·039	0·046	**−0·444**	−0·139	0·356	**0·493**	−0·301	−0·101	−0·315	0·084	**−0·511**	−0·125
Y	−0·187	**0·404**	0·229	−0·198	**0·417**	0·076	−0·143	**0·488**	0·019	−0·051	**0·518**	−0·167	−0·293	0·182	0·190	−0·329	0·267	0·105
V	−0·300	−0·012	**−0·444**	−0·277	0·110	−0·440	−0·177	0·170	**−0·594**	−0·138	0·489	0·069	−0·305	0·119	−0·077	−0·212	−0·209	**−0·459**
I	−0·228	−0·345	−0·181	−0·298	−0·275	−0·141	−0·335	−0·211	−0·079	−0·321	0·008	−0·001	−0·300	0·029	**−0·386**	−0·218	−0·331	−0·068
L	−0·171	**−0·409**	0·287	−0·142	**−0·433**	0·150	−0·254	−0·303	0·144	−0·305	−0·146	0·293	−0·168	**−0·489**	0·198	0·134	−0·215	**0·594**
F	−0·249	0·317	0·253	−0·183	−0·028	**0·388**	−0·141	0·342	0·282	−0·227	−0·023	**0·595**	−0·212	−0·427	0·181	−0·313	−0·055	0·322
K	−0·273	−0·320	0·022	−0·299	−0·295	0·140	−0·358	−0·168	−0·007	**−0·329**	−0·084	−0·030	−0·250	−0·314	0·368	−0·344	−0·211	0·222

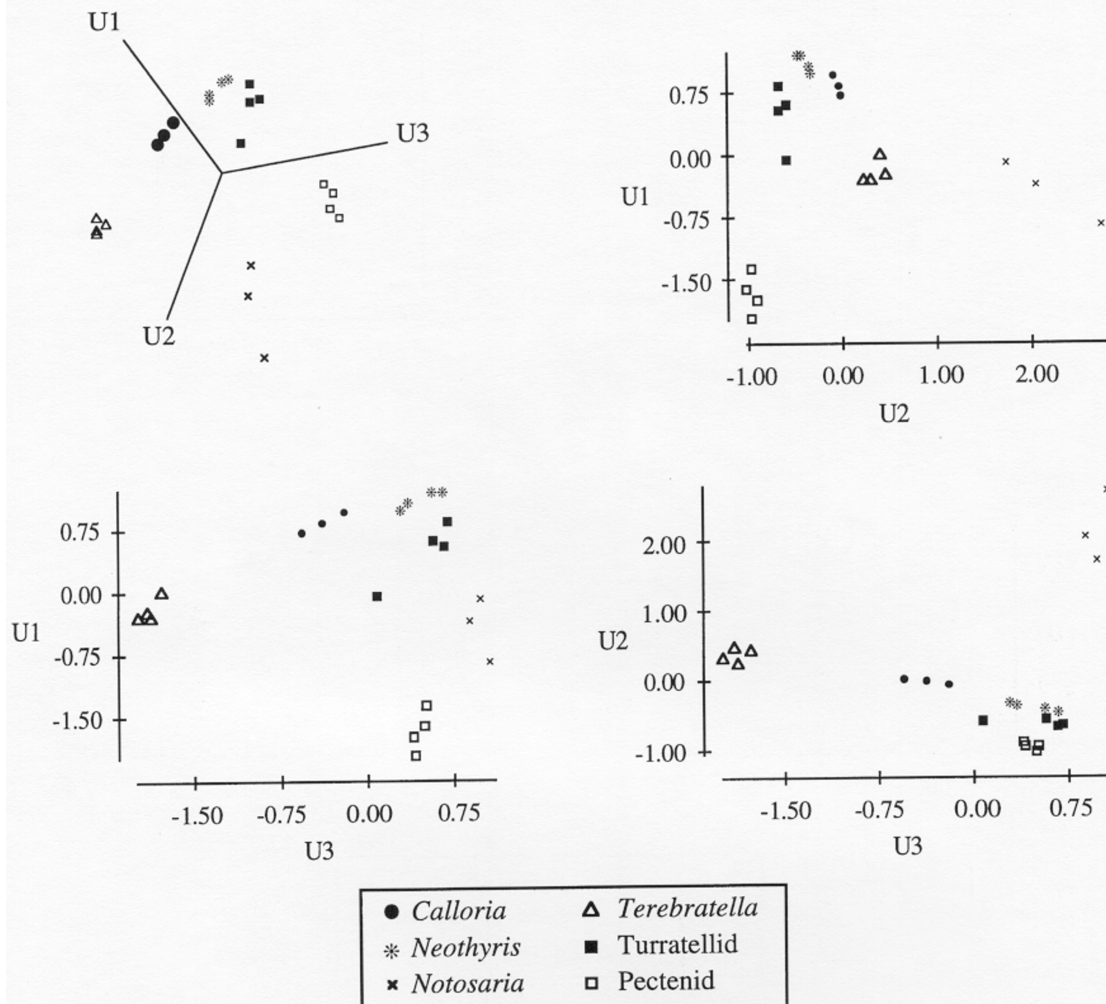


TEXT-FIG. 3. Plots of the first three principal components for the concentration of amino acids from samples collected from the Tainui Shellbed. All samples are well separated, with classification of the Terebratulida to the subordinal level (see text).

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, the effect of carbohydrates and of different mixtures of amino acids in the sample (Walton in press). However, as the amino acids are contained within a single time plane, and provided that there has been no homogenization of the amino acid composition of the samples in the horizon through time, similar methods of taxonomic discrimination can be used as for the Recent samples (Walton *et al.* 1993). Amino acids are referred to by their standard three letter codes (Appendix 1).

Within horizons

The Rapanui Marine Sand (*c.* 0·12 Ma) is the youngest of the horizons considered in the present study. The first three principal components (Table 2) contain 93·5 per cent. of the total variation of the dataset, mainly due to Glutamic acid (Glu) and Alanine (Ala) for the first, Tyrosine (Tyr) and Leucine (Leu) for the second, and Aspartic acid (Asp), Proline (Pro) and Valine (Val) for the third.

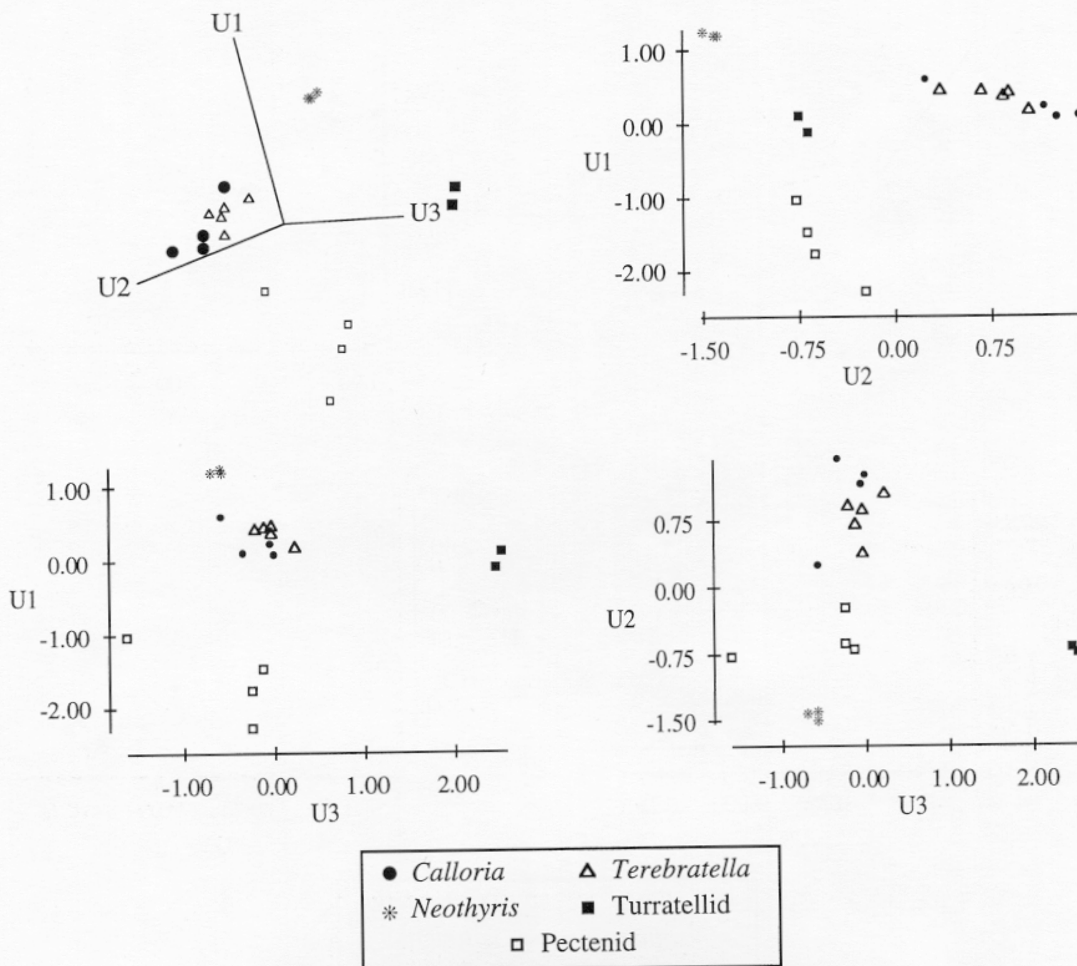


TEXT-FIG. 4. Plots of the first three principal components for the concentration of amino acids from samples collected from the Pinnacle Sand. All samples are well separated to the subordinal level (see text).

Graphical representation of the first three principal components (Text-fig. 2) shows that separation of samples by this method is good to at least the subordinal level. Specimens of *Neothyris lenticularis* present in the sample collected are derived (Walton 1992) and are not included in this analysis.

For the Tainui Shellbed (*c.* 0·40 Ma), PCA recalculates 90·6 per cent. of the variance within the first three eigenvalues (Table 2). The variability of the first principal component is caused mainly by Arginine (Arg) and Ala (Table 2), the second by Tyr and Leu, and the third by Pro and Val. A plot of the samples on the first three eigenvectors shows that there is good separation of the genera in space (Text-fig. 3). There has been no homogenization of the amino acid composition in samples through the horizon. The brachiopod samples are well separated at the ordinal level, with *Notosaria nigricans* (Rhynchonellida) plotting well away from the three species assigned to the Terebratulida. The three species in the Terebratulida may also be separated.

The first three principal components for the samples from the Pinnacle Sand (*c.* 0·42 Ma) contain 87·6 per cent. of the variation of the samples (Table 2). The first principal component has variation mainly due to the concentration of Arg and Lysine (Lys), the second due to Threonine (Thr) and



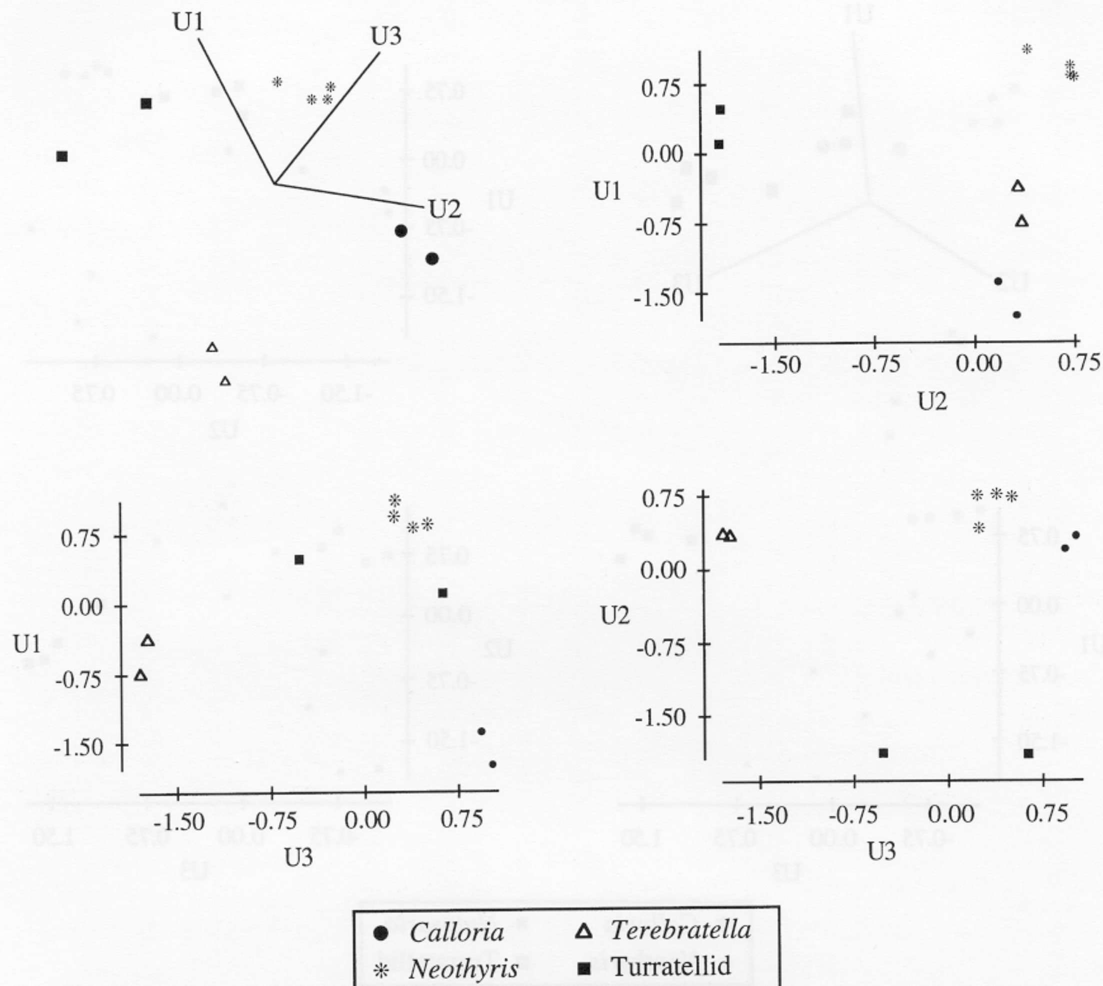
TEXT-FIG. 5. Plots of the first three principal components for the concentration of amino acids from samples collected from the Lower Castlecliff Shellbed. Note the merging of data points for the Terebratulida caused by the reduction of information available due to the degradation of amino acids in the sample (see text).

Tyr, and the third to Glycine (Gly), Pro and Val. Once again, there is good separation for all samples at the ordinal level (Text-fig. 4).

Samples from the Lower Castlecliff Shellbed (*c.* 0·44 Ma) are beginning to show the influence of time. The first three principal components contain 91·7 per cent. of the dataset variation (Table 2), which is due to Glu and Lys in the first principal component, the second by Gly, Tyr and Val, and the third has variation mainly due to Pro and Phenylalanine (Phe). Although the outgroups are well separated from the brachiopods (Text-fig. 5), and *Neothyris lenticularis* is separated, the brachiopod samples assigned to the subfamily Terebratellinae are plotting closer together and the data for the samples are beginning to merge, lowering the level of taxonomic information available.

Samples from the Kupe Formation (*c.* 0·5 Ma) did not include either *Notosaria nigricans* or a pectenid. The first three principal components contain 96·8 per cent. of the variation of the dataset (Table 2), due mainly to the variation of Glu and Ala for the first principal component, Thr and Leu for the second, and Thr for the third. All samples are well separated (Text-fig. 6).

The data for the Hautawa Shellbed (*c.* 2·20 Ma) show that 87·4 per cent. of the variation of the

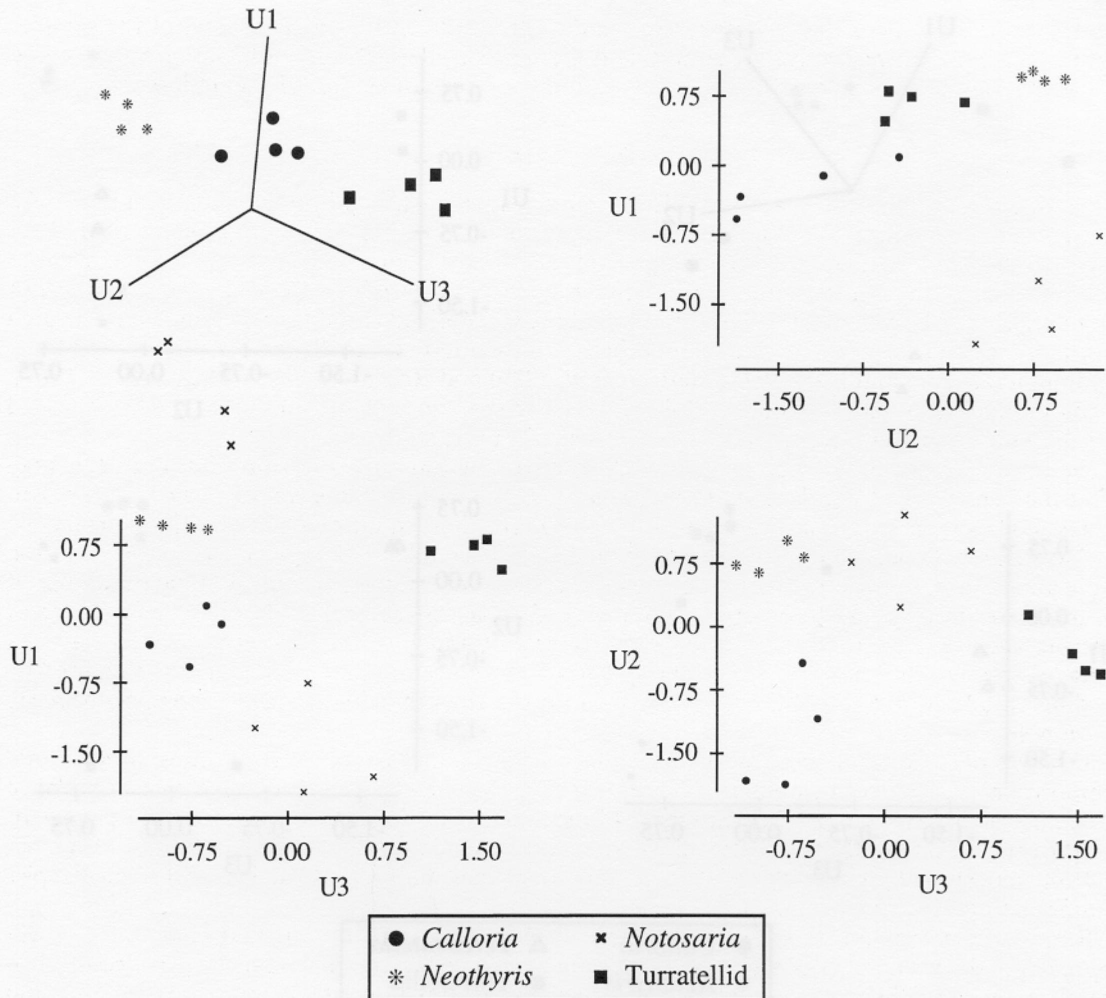


TEXT-FIG. 6. Plots of the first three principal components for the concentration of amino acids from samples collected from the Kupe Formation. Although separation is possible to below the subfamily level, there are fewer data points available and these tend to be more widely separated within a grouping (see text).

dataset is contained within the first three principal components (Table 2). This is due mainly to Thr and Ala for the first principal component, Glu and Pro for the second and Val and Leu for the third. No Arg remained in any sample and thus was omitted from the PCA. The samples are well separated by the amino acid data (Text-fig. 7), with both outgroups and *Notosaria nigricans* plotting away from the Terebratulida. Within this latter group, *Calloria inconspicua* and *Neothyris lenticularis* are also well separated, although the data points are more widely spaced for each taxon.

Between horizons

All samples analysed in this study were incorporated into the same dataset and a new PCA completed, in order to ascertain whether a taxonomic signal was preserved through geological time at a high enough level to allow similar samples to plot close together. The abundances of Serine (Ser), Arg and Thr were omitted from this calculation, as in some of the older samples they are completely decomposed.



TEXT-FIG. 7. Plots of the first three principal components for the concentration of amino acids from samples collected from the Hautawa Shellbed. Note the spreading of the data within the groupings caused by the loss of specificity due to amino acid degradation (see text).

For comparison between horizons the data was examined in two ways. Text-figure 8A shows the plot of the first three principal components derived from the absolute concentration of amino acids in the samples. The first three principal components contain 89·4 per cent. of the total variation present in the dataset, although the data points do not appear to contain any significant order and there is a great deal of overlap between the taxa. Text-figure 8B was constructed using the relative abundance of the amino acids, with 82·4 per cent. of the variation in the dataset being contained within the first three principal components. In this case the taxa may be split into four main groupings: Terebratulida, Rhynchonellida, pectenids and turratellids. There is clearly a major difference between the two datasets, although the groupings show that some degree of taxonomic separation is possible from a dataset that includes both Recent and fossil material, back to 2·2 Ma.

The two outgroups, pectenids and turratellids, form distinct groupings, as would be expected from members of different phyla. The brachiopods form two groups, with Rhynchonellida grouping away from Terebratulida. Within Terebratulida, no differentiation can be made, as the variation in

the data causes a spread that encompasses the data from the entire order. Several of the samples plot away from their respective groupings, and there is considerable spread within groups, caused by the differing ages and therefore differing amounts of decomposition of the amino acids.

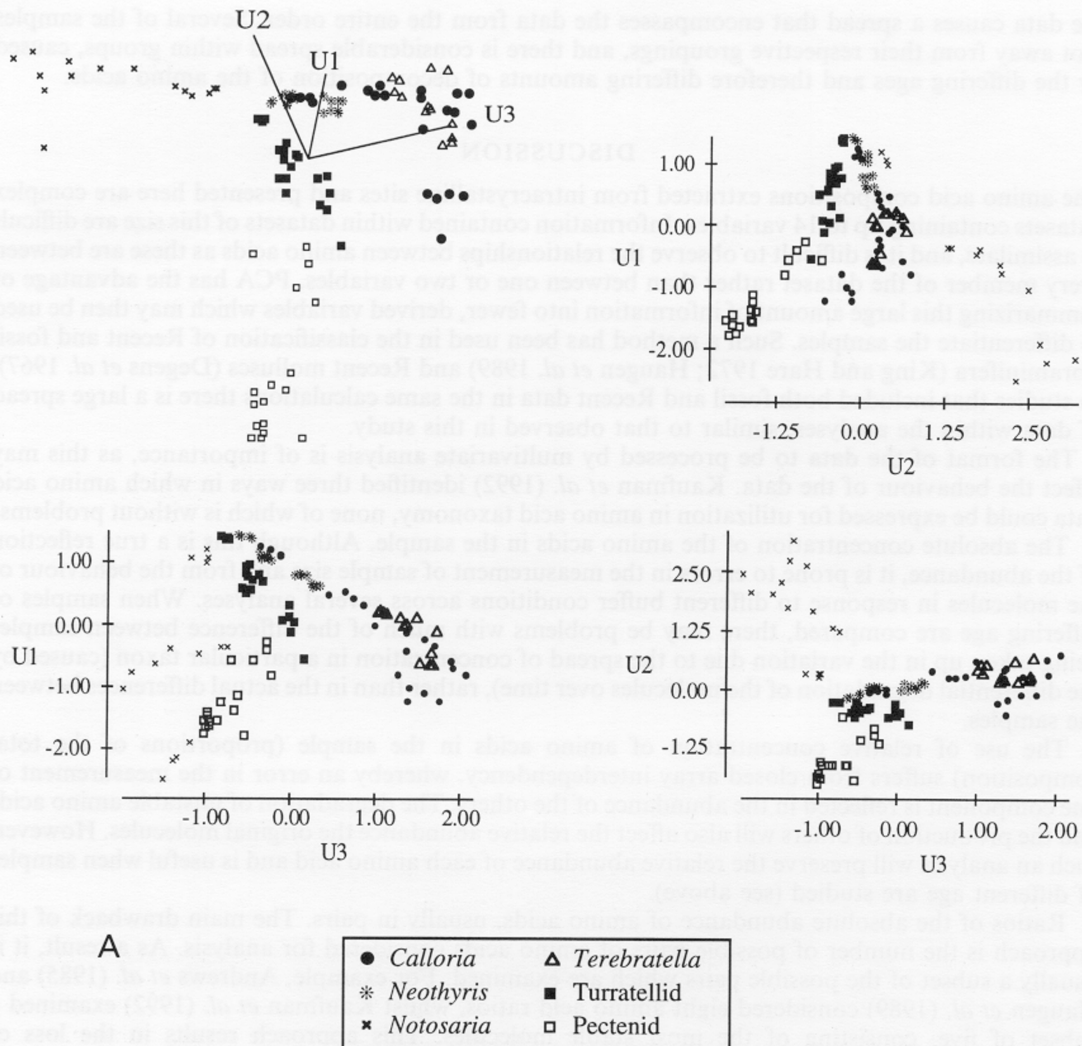
DISCUSSION

The amino acid compositions extracted from intracrystalline sites and presented here are complex datasets containing up to 14 variables. Information contained within datasets of this size are difficult to assimilate, and it is difficult to observe the relationships between amino acids as these are between every member of the dataset rather than between one or two variables. PCA has the advantage of summarizing this large amount of information into fewer, derived variables which may then be used to differentiate the samples. Such a method has been used in the classification of Recent and fossil Foraminifera (King and Hare 1972; Haugen *et al.* 1989) and Recent molluscs (Degens *et al.* 1967). In studies that included both fossil and Recent data in the same calculations there is a large spread of data within the analyses, similar to that observed in this study.

The format of the data to be processed by multivariate analysis is of importance, as this may affect the behaviour of the data. Kaufman *et al.* (1992) identified three ways in which amino acid data could be expressed for utilization in amino acid taxonomy, none of which is without problems:
1. The absolute concentration of the amino acids in the sample. Although this is a true reflection of the abundance, it is prone to errors in the measurement of sample size and from the behaviour of the molecules in response to different buffer conditions across several analyses. When samples of differing age are compared, there may be problems with much of the difference between samples being taken up in the variation due to the spread of concentration in a particular taxon (caused by the differential degradation of the molecules over time), rather than in the actual differences between the samples.
2. The use of relative concentration of amino acids in the sample (proportions of the total composition) suffers from closed array interdependency, whereby an error in the measurement of one component is reflected in the abundance of the others. The degradation of unstable amino acids and the production of others will also affect the relative abundance the original molecules. However, such an analysis will preserve the relative abundance of each amino acid and is useful when samples of different age are studied (see above).
3. Ratios of the absolute abundance of amino acids, usually in pairs. The main drawback of this approach is the number of possible pairs of amino acids considered for analysis. As a result, it is usually a subset of the possible pairs which are examined. For example, Andrews *et al.* (1985) and Haugen *et al.* (1989) considered eight amino acid ratios, whilst Kaufman *et al.* (1992) examined a subset of five, consisting of the most stable molecules. This approach results in the loss of information from the other amino acids not included in the samples.

Ratios between the amino acids have been the most common of the data formats thus far utilized for amino acid taxonomy of fossils (e.g. Jope 1967; Haugen *et al.* 1989; Kaufman *et al.* 1992). However, from the data presented in this study the ratios between the pairs of amino acids range over a wide scale, and there is an overlap between the ratios. Walton and Curry (1994) suggested utilizing relative abundances in PCA, although the level of information retrieved by this is less than when the absolute abundances are used (Text-fig. 9; cf. Text-fig. 5). For these reasons, and recognizing the problems outlined above, it is considered that the highest levels of taxonomic information in this case are revealed through the use of absolute abundances of amino acids.

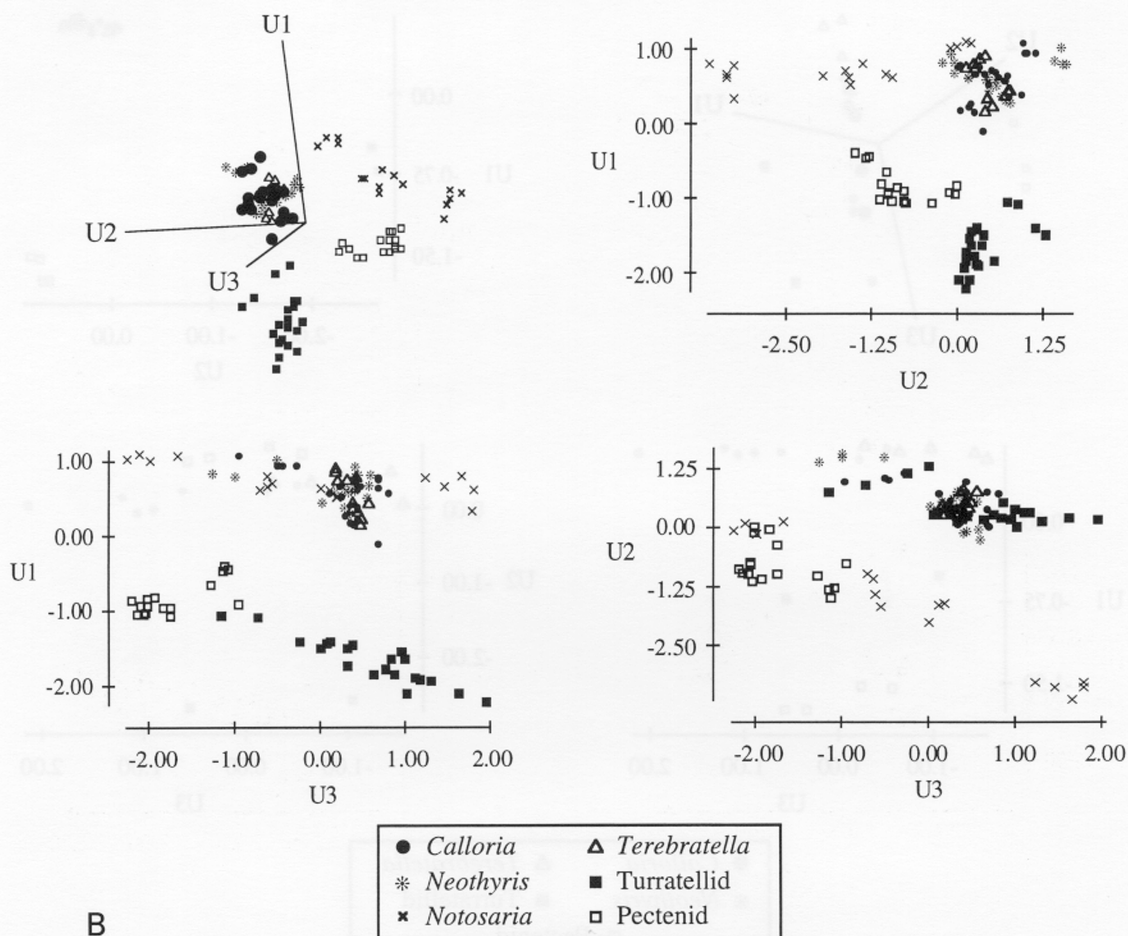
For each horizon in this study, every grouping of samples has a characteristic amino acid signature that is sufficiently different to allow separation of different taxa and convergence of similar taxa. Each major grouping is discrete, indicating that there has been no homogenization of the amino acids in the horizon. Samples that have a similar amino acid composition will plot closer together than those which have a different composition. Samples which are morphologically distinct (e.g. members of different phyla or classes) have amino acid compositions that are very different. Hence the brachiopods are well separated from the outgroups (molluscs) in all cases. Within a class,



TEXT-FIG. 8. For legend see opposite.

separations are also very distinct at the ordinal level (e.g. between Rhynchonellida and Terebratulida). These amino acid signatures must reflect original genetic differences between the samples.

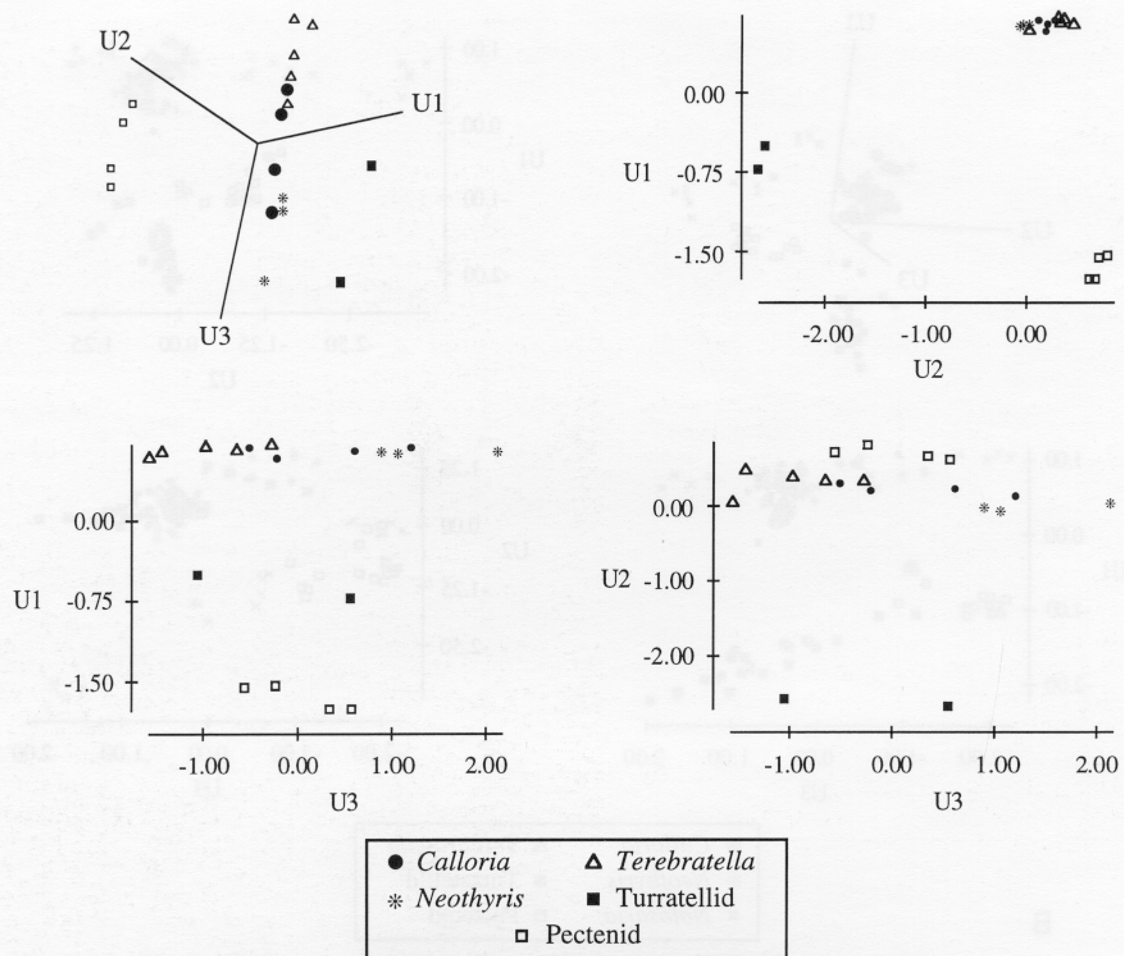
In fossil samples, as might be expected, the best separation of the taxa is gained when utilizing the youngest samples. As samples from successively older horizons are considered, the level of taxonomic information present within the shell generally decreases. This is due to the older samples containing macromolecules which have been degraded to a higher degree than have those of younger samples. This degradation is recognized by the merging of the formerly discrete groupings, representing the loss of differences between the amino acid compositions of the taxa. As degradation proceeds, differences between the relative amino acid composition will be reduced (by the loss of the less stable molecules and the gain, both relative and absolute, of others). The merging of datapoints represents the decay of unstable amino acid molecules and the diagenetic production of others which are important in differentiating between species. This process has an endpoint of the amino



TEXT-FIG. 8. Plots of the first three principal components for the concentration (A) and relative abundance (B) of amino acids in all samples combined together to examine the preservation of taxonomic signal in samples of differing ages. In A, it is not possible to recognize definite groupings. This is caused by much of the variation being taken up by the difference in abundance of the individual amino acids in the sample, rather than the difference in composition between the samples. However, in B, four groupings may easily be identified. In this case, the variation due to concentration in the sample size is removed by using the relative proportions of the amino acids which are preserved regardless of the concentration (see text).

acid content being similar (although not identical) in all samples. Merging of samples demonstrates the importance of retaining as much original information as possible; selecting groups of amino acids as the starting point for taxonomic analysis may reduce the level of taxonomic significance observed.

When samples of different ages are analysed together, a 'typical' amino acid composition is recognized which enables groupings of similar organisms to be made. The degradation of amino acids does not distort the amino acid signature of the sample to a level where it is similar to others from a different order. The degradation of unstable amino acids over time follows a pattern that is similar for all brachiopod species analysed (Walton 1996, in press). It is likely that the same will hold true for other samples. Once free from their proteins, the amino acids will behave as individual molecules and their degradation will no longer be influenced by the primary or higher order structure of the protein. No contaminating extraneous molecules will be included in the analysis,



TEXT-FIG. 9. Plots of the first three principal components for the relative proportions of amino acids from samples collected from the Lower Castlecliff Shellbed. Note the loss of detail in the analysis, resulting from lower amounts of information preserved by the relative proportions of amino acids (see text).

provided that the molecules remain within the shell and are not released by shell recrystallization, etc. Degradation of the amino acids occurs, 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, which represents this decay and diagenetic production of amino acids.

Using standard amino acid analysers, the level of information described here may possibly be the highest to be gained routinely from fossil samples. This is not as high as was initially hoped for amino acids recovered from intracrystalline sites, as these were thought to be better protected (Curry 1988). In Recent samples, this method can distinguish between genera in all cases, and possibly also species (investigated with *Neothyris*; Walton *et al.* 1993). The degradation of the molecules has led to a decrease in the amount of information retained which may be recorded by the instrumentation used. It is likely that further analyses using other techniques, such as GC-MS, may refine this information level by quantifying the degradative remains of amino acids. In addition to the amino acids there is a range of other molecules present within the shell that may provide

further phylogenetic information, or may mask a true relationship. In particular, taxonomically important molecules will be formed from the original amino acids through a range of degradative reactions (Walton in press) and the products may not be amino acids and hence will not be recorded. Indeed, there will be a range of intermediates, but degradation will ultimately lead to the formation of short-chain hydrocarbons (Thompson and Creath 1966).

If the degradative pathways are known, then the reaction products can be assayed and the original amino acid composition restored to extract the taxonomic information. This is similar to the suggestion of Kaufman *et al.* (1992) who attempted to reconstruct the amino acid composition by calculating the rate of degradation based on the rate of amino acid racemization. These compositions were related to Recent counterparts for identification. However, the method of Kaufman *et al.* (1992) relies upon there being a recognized Recent representative of taxa used in comparison studies and the absence of significant evolution of the protein over geological time. Clearly, if amino acid taxonomy is to be of general use in palaeontology, both of these problems must be overcome. Reconstruction of the original amino acid composition of the fossil through analysis of the degradation products will enable taxa with no living representatives to undergo this type of analysis.

Even though it is more than 40 years since the first amino acids were recovered from the shells of fossils (Abelson 1954), we still know very little regarding many of the rates and pathways of protein and amino acid degradation. Some reactions are known, however: for example, one of the degradation products of Arg is ornithine. The concentration of ornithine in shells varies inversely to the concentration of Arg (Walton in press). This is the only pathway by which ornithine can be formed in the shell and therefore represents an unambiguous link with the parent molecule. Recognition of such linkages should be possible for many of the original molecules and therefore the original composition may be reconstructed. However, not all molecules will have such an unambiguous pathway. Ser degrades (through a number of intermediates) to form Ala (Bada *et al.* 1978), resulting in the increased level of Ala seen in brachiopods (Walton 1996), in Foraminifera (Haugen *et al.* 1989) and molluscs (Kaufman *et al.* 1992). This Ala will be indistinguishable from the original Ala in the sample and will therefore distort the analysis. However, the degradative pathways of other amino acids (e.g. Val, Leu) are unknown or poorly understood and must be recognized prior to any attempted reconstruction of the amino acids for use in taxonomy.

CONCLUSIONS

The results of this study show that, despite high levels of amino acid degradation, taxonomic information is preserved in intracrystalline molecules. This information may be observed by using graphical presentation of multivariate statistical analysis of the relative proportions of amino acids. In all samples, separation is possible to at least subordinal level and in some cases to subfamilial level on the basis of amino acid composition alone. The diagrams may be considered as analogous to geochemical discrimination diagrams, as the majority of the groupings described above would be recognized, even if morphologically derived groupings were not known.

The degree of taxonomic discrimination is less than was hoped at the start of this study, but still represents the preservation of characteristic amino acid signatures. This may be refined by examination of the degradative remains of fossils. A full understanding of degradative pathways, to allow the reconstruction of the parent molecules from the degradation products, is a prerequisite to allow detailed taxonomic information to be retrieved from the organic component of shells. Amino acid data alone may not be sufficient in the fossil record to fulfil the aims of a molecular taxonomy.

Acknowledgements. This work was undertaken during the tenure of a UK NERC studentship (GT4/89/GS/42) at the University of Glasgow, and was written during a University of Derby sabbatical, both of which are gratefully acknowledged. This manuscript benefited from the critical reading of Maggie Cusack, Matthew

Collins and an anonymous referee. Helen Wilkins and Ann Agarogda (Derby) and Sandra McCormack (Glasgow) are thanked for technical assistance.

REFERENCES

ABELSON, P. H. 1954. Organic constituents of fossils. *Yearbook of the Carnegie Institute of Washington*, **53**, 97–101.

—— 1955. Organic constituents of fossils. *Yearbook of the Carnegie Institute of Washington*, **54**, 107–109.

AKIYAMA, M. 1971. The amino acid composition of fossil scallop shell proteins and non-proteins. *Biomineralisation*, **3**, 65–70.

ALBECK, S., AIZENBERG, L., ADDADI, L. and WEINER, S. 1993. Interactions of various skeletal intracrystalline components with calcite crystals. *Journal of the American Chemical Society*, **115**, 11691–11697.

ANDERTON, P. W. 1981. Structure and evolution of the South Wanganui Basin, New Zealand. *New Zealand Journal of Geology and Geophysics*, **24**, 39–83.

ANDREWS, J. T., MILLER, G. H., DAVIES, D. C. and DAVIES, K. H. 1985. Generic identification of fragmentary Quaternary molluscs by amino acid chromatography – a tool for Quaternary and palaeontological research. *Geological Journal*, **20**, 1–20.

ARMSTRONG, W. G., HALSTEAD, L. B., REED, F. B. and WOOD, L. 1983. Fossil proteins in vertebrate calcified tissues. *Philosophical Transactions of the Royal Society of London, Series B*, **301**, 301–343.

BADA, J. L., SHOU MING-YUNG, MAN, E. H. and SCHROEDER, R. A. 1978. Decomposition of hydroxy amino acids in foraminifera tests; kinetics, mechanism and geochronological implications. *Earth and Planetary Science Letters*, **41**, 67–76.

COHEN, B. L. 1994. Immuno-taxonomy and the reconstruction of brachiopod phylogeny. *Palaeontology*, **37**, 907–911.

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. *Historical Biology*, **1**, 207–224.

—— MUYZER, G., CURRY, G. B., SANDBERG, P. and WESTBROEK, P. 1991. Macromolecules in brachiopod shells: characterization and diagenesis. *Lethaia*, **24**, 387–397.

CORNISH-BOWDEN, A. 1979. How reliably do amino acid composition comparisons predict sequence similarities between proteins? *The Journal of Theoretical Biology*, **76**, 369–386.

—— 1983. Relating proteins by amino acid composition. *Methods in Enzymology*, **91**, 60–75.

CURRY, G. B. 1988. Amino acids and proteins from fossils. 20–33. *In* RUNNEGAR, B. and SCHOPF, J. W. (eds). *Molecular evolution and the fossil record.* Short Courses in Paleontology, 1. Paleontological Society, Knoxville, 167 pp.

CUSACK, M., CURRY, G. B., CLEGG, H. and ABBOTT, G. 1992. An intracrystalline chromoprotein from red brachiopod shells: implications for the process of biomineralisation. *Comparative Biochemistry and Physiology, Series B*, **102**, 93–95.

DAVIS, J. C. 1986. *Statistics and data analysis in geology*. John Wiley & Sons, New York, 646 pp.

DEGENS, E. T., SPENCER, D. W. and PARKER, R. H. 1967. Paleobiochemistry of molluscan shell proteins. *Comparative Biochemistry and Physiology*, **20**, 553–579.

DUPONT, D. R., KEIM, P. S., CHUI, A., BELLO, R., BOZZINI, M. and WILSON, K. J. 1989. A comprehensive approach to amino acid analysis. 284–294. *In* HUGLI T. E. (ed.). *Techniques in protein chemistry*. Academic Press, New York, 612 pp.

DUSSART, G. B. J. 1983. The amino acid composition of fresh-water mollusc shells in relation to phylogeny and environment. *Journal of Molluscan Studies*, **49**, 213–223.

FLEMING, C. A. 1953. *The geology of the Wanganui Subdivision.* Department of Scientific and Industrial Research, Wellington, 362 pp.

HARE, P. E. 1974. Amino acid dating of bone – the influence of water. *Yearbook of the Carnegie Institute of Washington*, **73**, 576–581.

—— and HOERING, T. C. 1977. The organic constituents of fossil mollusc shells. *Yearbook of the Carnegie Institute of Washington*, **76**, 625–631.

—— and MITTERER, R. M. 1969. Laboratory simulation of amino acid diagenesis in fossils. *Yearbook of the Carnegie Institute of Washington*, **67**, 205–208.

HAUGEN, J.-E., SEJRUP, H.-P. and VOGT, N. B. 1989. Chemotaxonomy of Quaternary benthic foraminifera using amino acids. *Journal of Foraminiferal Research*, **19**, 38–51.

HEINRIKSON, R. L. and MEREDITH, S. C. 1984. Amino acid analysis by reverse-phase high-performance liquid chromatography: precolumn derivatization with phenylisothiocyanate. *Analytical Biochemistry*, **136**, 65–74.

JOPE, M. 1967. The protein of brachiopod shell – II. Shell protein from fossil articulates: amino acid composition. *Comparative Biochemistry and Physiology*, **20**, 601–605.

KAUFMAN, D. S., MILLER, G. H. and ANDREWS, J. T. 1992. Amino acid composition as a taxonomic tool for molluscan fossils: an example from Pliocene–Pleistocene Arctic marine deposits. *Geochimica et Cosmochimica Acta*, **56**, 2445–2453.

KING, K., Jr and HARE, P. E. 1972. Amino acid composition of the test as a taxonomic character for living and fossil planktonic foraminifera. *Micropaleontology*, **18**, 285–293.

MACFIE, H. J. H., LIGHT, N. D. and BAILEY, A. J. 1988. Natural taxonomy of collagen based on amino acid composition. *The Journal of Theoretical Biology*, **131**, 401–418.

NORRIS, R. M. and GRANT-TAYLOR, T. L. 1989. Late Quaternary shellbeds, Western Shelf, New Zealand. *New Zealand Journal of Geology and Geophysics*, **32**, 343–356.

QIAN YAORONG, ENGEL, M. H., GOODFRIEND, G. A. and MACKO, S. A. 1995. Abundance and stable carbon isotope composition of amino acids in molecular weight fractions of fossil and artificially aged mollusk shells. *Geochimica et Cosmochimica Acta*, **59**, 1113–1124.

ROBBINS, L. L. and BREW, K. 1990. Proteins from the organic matrix of core top and fossil planktonic foraminifers. *Geochimica et Cosmochimica Acta*, **54**, 2285–2292.

—— and DONACHY, J. E. 1991. Mineral regulating proteins from fossil planktonic foraminifera. 139–148. *In* SIKES, C. S. and WHEELER, A. P. (eds). *Surface reactive peptides and proteins*. American Chemical Society, Washington DC, 416 pp.

—— and HEALY-WILLIAMS, N. 1991. Towards a classification of planktonic foraminifera based on biochemical, geochemical and morphological criteria. *Journal of Foraminiferal Research*, **21**, 159–167.

SNEATH, P. H. A. and SOKAL, R. R. 1973. *Numerical taxonomy*. W. H. Freeman and Company, San Francisco, 573 pp.

SUCOV, H. M., BENSON, S., ROBINSON, J. R., BRITTEN, R. Y., 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. *Developmental Biology*, **120**, 507–519.

SYKES, G. A., COLLINS, M. J. and WALTON, D. I. 1995. The significance of a geochemically isolated intracrystalline organic fraction within biominerals. *Organic Geochemistry*, **23**, 1059–1065.

THOMPSON, R. R. and CREATH, W. B. 1966. Low molecular weight hydrocarbons in Recent and fossil shells. *Geochimica et Cosmochimica Acta*, **30**, 1137–1152.

TOWE, K. M. 1980. Preserved organic ultrastructure: an unreliable indicator for Paleozoic amino acid biogeochemistry. 65–74. *In* HARE, P. E., HOERING, T. C. and KING, K. J. (eds). *Biogeochemistry of amino acids*. John Wiley & Sons, New York, 558 pp.

WALTON, D. I. 1992. Biogeochemistry of brachiopod intracrystalline proteins and amino acids. Unpublished Ph.D. thesis, University of Glasgow.

—— 1996. Degraded intracrystalline proteins and amino acids from fossil brachiopods and considerations for amino acid taxonomy. 289–297. *In* COPPER, P. and JIN, JUISO (eds). *Brachiopods*. Balkema Press, Rotterdam, 373 pp.

—— in press. Degradation of intracrystalline proteins and amino acids in fossil brachiopods. *Organic Geochemistry*.

—— and CURRY, G. B. 1994. Extraction, analysis and interpretation of intracrystalline amino acids from fossils. *Lethaia*, **27**, 179–184.

—— CUSACK, M. and CURRY, G. B. 1993. Implications of the amino acid composition of Recent New Zealand brachiopods. *Palaeontology*, **36**, 883–896.

WEHMILLER, J. F., YORK, L. L. and BART, M. L. 1995. Amino acid racemization geochronology of reworked Quaternary mollusks on US Atlantic coast beaches: implications for chronostratigraphy, taphonomy, and coastal sediment transport. *Marine Geology*, **124**, 303–337.

WEINER, S., LOWENSTAM, H. A. and HOOD, L. 1976. Characterization of 80 million year old mollusk shell proteins. *Proceedings of the National Academy of Science, USA*, **73**, 2541–2545.

WYCKOFF, R. W. G. 1972. *The biochemistry of animal fossils*. Scientechnica, Bristol, 151 pp.

DEREK WALTON

Division of Earth Sciences
University of Derby
Kedleston Road
Derby DE22 1GB, UK

Typescript received 9 December 1996
Revised typescript received 27 August 1997

APPENDIX

The one letter and three letter codes for the amino acids used in this study.

Amino acid	Three letter code	One letter code	Amino acid	Three letter code	One letter code
Alanine	Ala	A	Lysine	Lys	K
Arginine	Arg	R	Phenylalanine	Phe	F
Aspartic acid	Asp	D	Proline	Pro	P
Glutamic acid	Glu	E	Serine	Ser	S
Glycine	Gly	G	Threonine	Thr	T
Isoleucine	Ile	I	Tyrosine	Tyr	Y
Leucine	Leu	L	Valine	Val	V