IMPLICATIONS OF THE AMINO ACID COMPOSITION OF RECENT NEW ZEALAND BRACHIOPODS

by DEREK WALTON, MAGGIE CUSACK *and* GORDON B. CURRY

ABSTRACT. Intracrystalline proteins and amino acids were extracted from within the shells of several species of Recent brachiopods from New Zealand. These molecules were characterized both by amino acid analysis and by SDS–PAGE and the data derived from this summarized by multivariate statistical analysis. Results indicate that the amino acids from within the shell may be used as objective taxonomic discriminators, and that the groupings recognized by this method are generally in confirmation with those from the *Treatise on invertebrate paleontology*. The presence of amino acids containing taxonomic information within the shells of Recent brachiopods has implications for the use of amino acid analysis in the investigation of fossil species.

THE intracrystalline organic molecules which are present as a mixture within the shells of Recent brachiopods are a rich source of information regarding the taxonomy of the organisms (Collins *et al.* 1988, 1991*b*; Curry *et al.* 1991*b*, 1991*c*; Cusack *et al.* 1992). The molecules consist mostly of protein, although quantities of lipids (Curry *et al.* 1991*b*), alcohols (H. Clegg pers. comm.) and carbohydrate (Collins *et al.* 1991) are also present. Although the function of these molecules, and their role, if any, in the process of biomineralization is unclear, they are of significance in the field of molecular palaeontology.

Of the molecules present within the shells, lipids, alcohols and carbohydrates generally contain little information of significant taxonomic value. Proteins, however, contain information derived from the DNA of the organism, which may be utilized in studies of molecular taxonomy (e.g. Joysey 1988). Proteins are composed of relatively stable monomeric units, amino acids, which are present in a defined order along the chain of the molecule. It is this information, the primary structure, which is conventionally used for biochemical comparison of protein, as differences in the nucleotide bases in the DNA of organisms are usually (but not always) reflected in differences in the nature and order of amino acids (Crick *et al.* 1961). Analysis of the position of the changes in amino acids or nucleotide bases can only be completed by sequencing either the proteins or the DNA of the organism. The complete sequencing of proteins is a difficult and time-consuming process; even using automated sequencers, it is rare to be able to sequence more than 40 amino acids along a protein chain at any one time (counting from the *N*-terminus). Either enzymatic and/or chemical cleavage of proteins, or the use of DNA probes, is necessary to elucidate further sequence information, although there is no guarantee of success through either of these procedures.

Curry *et al.* (1991*a*) and Cusack *et al.* (1992) obtained *N*-terminal sequence information from brachiopod intracrystalline proteins. These proteins are homologous between different brachiopod species, although they do not correspond to the sequence of any other known proteins. The sequence of a small (6·5 kDa) protein analysed from three New Zealand genera (Cusack *et al.* 1992) shows few substitutions between the proteins. However, as yet, no study has completed the sequencing of more than a quarter of the total residues contained within any particular brachiopod shell protein.

Differences in the DNA between species will also be reflected by different relative proportions of amino acids within a sample, which may be easily quantified by amino acid analysis. This will reduce to some degree the level of information available from the proteins (as the primary structure is lost),

[Palaeontology, Vol. 36, Part 4, 1993, pp. 883–896.]

but this is compensated for by the much shorter time and much smaller sample size since *c.* 10 μg shell is sufficient for this type of analysis, while protein sequencing requires *c.* 100 g shell to provide sufficient information. Amino acid quantification is rapid and is able to identify variation in the relative proportions of amino acids, 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 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, although some changes in nucleotide base will not result in a changed amino acid residue in the protein (Crick *et al.* 1961).

It was with the inherent difficulties of protein sequencing in mind that the intracrystalline proteins from several Recent brachiopod species (listed in Table 1) were characterized by their amino acid compositions. This information was examined by multivariate statistical techniques to extract potential taxonomic data from the variation in concentration of 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.

TABLE 1. Species and localities used in SDS–PAGE and amino acid analysis.

Brachiopod	Locality
Neothyris lenticularis (Deshayes, 1839)	Paterson Inlet, Stewart Island
Neothyris parva (Cooper, 1982)	Continental shelf, 10 miles SW of Otago Peninsula
Waltonia inconspicua (Sowerby, 1846)	Otago Harbour
Notosaria nigricans (Sowerby, 1846)	Tikoraki Point, Moeraki, Otago
Terebratella sanguinea (Leach, 1814)	Paterson Inlet, Stewart Island
Terebratella haurakiensis (Allan, 1931)	Hauraki Gulf, North Island
Liothyrella neozealandica (Thomson, 1918)	Dusky Sound, Fiordland
Gyrothyris mawsoni (Thomson, 1918)	Continental shelf, 10 miles SW of Otago Peninsula

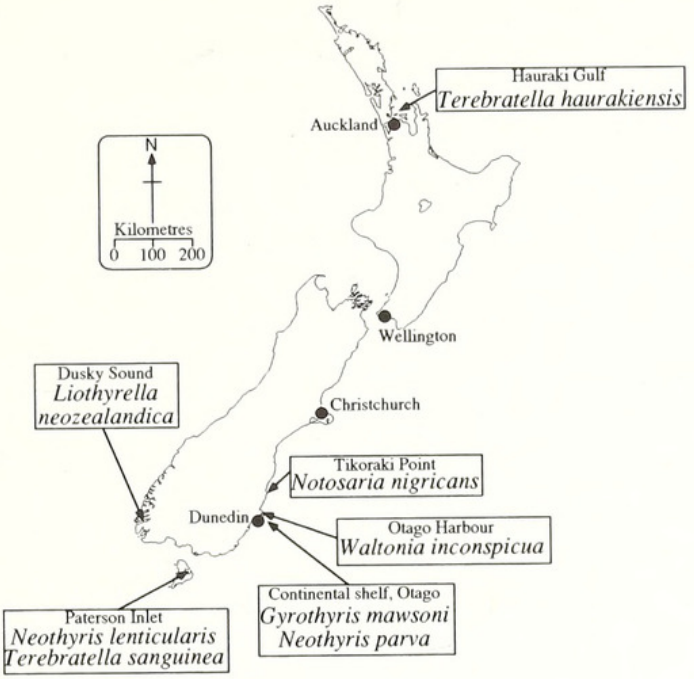
Other studies have shown that the number of proteins contained within the shell also varies taxonomically (Curry *et al.* 1991*b*). To provide a method of secondary confirmation of the taxonomic conclusions of the amino acid technique, the organic extract of several of the samples was fractionated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE) using the method of Schägger and Von Jagow (1987). As SDS–PAGE requires large amounts of shell material, *Terebratella haurakiensis*, *Gyrothyris mawsoni* and *Neothyris parva*, could not be analysed in this way. SDS–PAGE reveals which species have a similar number and size of proteins present within their shells, which provides a secondary method of confirmation of similarity (Curry *et al.* 1991*b*) in case of spurious relationships derived from the amino acid data. Throughout this study, the standard one- or three-letter codes are used to denote the amino acids.

METHODS AND MATERIALS

Preparation

Samples of Recent brachiopods were collected from the locations given in Text-figure 1 and Table 1.Wherever possible, samples which were excessively bored or fractured were omitted from further study.

Sediment was scrubbed from the sample and encrusting epifauna removed by scraping. Articulated shells were disarticulated and any remaining body tissue removed, before soaking 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 air dried. Samples were ground using a ceramic pestle and mortar, and the powdered samples incubated in an aqueous solution of bleach (10% v/v) under constant motion for 24 hours at room temperature to destroy intercrystalline molecules by oxidation. Samples were washed by repeated



TEXT-FIG. 1. Location of collection sites for samples of extant species used in the analyses.

agitation with MilliQ™ water and centrifugation (typically ten washes). Powders were lyophilized, allowing almost indefinite storage at room temperature.

To release the incarcerated molecules from within the calcium carbonate, it is necessary to dissolve the inorganic phase. Two methods were employed:

1. Hydrochloric acid. For amino acid analysis, an aqueous solution of 2 M HCl at a ratio of 11 μl/mg was used to dissolve the shell powder. Once demineralization was complete, samples were centrifuged (20 ***g***.h.) to remove any remaining insoluble particles. This method of preparation required no further concentration.

2. Ethylene diamino tetra acetic acid, disodium salt (Na_2-EDTA). For samples to be analysed by SDS–PAGE, an aqueous solution of Na_2-EDTA (20% w/v) at a ratio of 23 ml/g shell was used to dissolve the shell powder (Collins *et al.* 1988). Once demineralization was complete, samples were centrifuged (20 ***g***.h.) to remove any remaining insoluble particles. The sample was initially filtered and concentrated using the Minitan™ tangential flow system (Millipore; Cusack *et al.* 1992), which removed the EDTA/calcium complexes. The intracrystalline extract was further concentrated using a Minicon™ static concentrater (Amicon). Both filtration systems used 10 kDa cutoff filters.

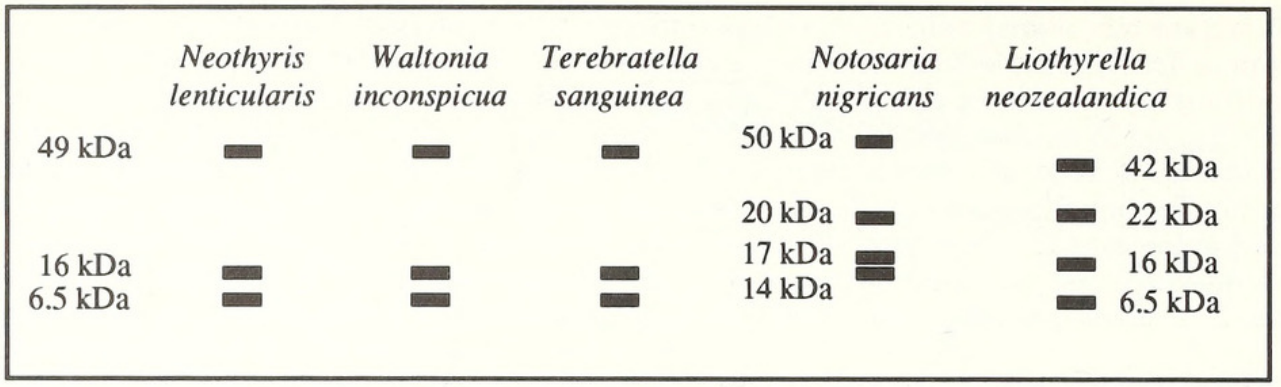
SDS–PAGE

The polyacrylamide gels used in this study were prepared to a method based on that of Schägger and Von Jagow (1987). 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 tracking dye (bromophenol blue) and incubated at 100 °C for 4 minutes to denature the protein and to allow SDS to bind to the protein. Proteins were separated by the application of a constant voltage of 50 V for 4 hours. Following electrophoresis, proteins were visualized by Coomassie Brilliant Blue R-250 stain.

Amino acid analysis

Hydrolysis of the peptide bonds which link individual amino acids in proteins and peptides is a prerequisite for amino acid analysis. Vapour phase hydrolysis, using 6 N HCl, was completed using automated hydrolysis on the Applied Biosystems 420H Amino Acid Analyser. Standard proteins and peptides were used during every analysis to ensure that hydrolysis proceeded to completion. Blank analyses were included to check for background levels of contamination.

Following hydrolysis, individual amino acids may be derivatized using Phenylisothiocyanate (PITC; Heinrikson and Meredith 1984) on the ABI 420H (Dupont *et al.* 1989), which absorbs UV light strongly at 254 nm. The derivatized amino acids were transferred to the dedicated ABI 130A narrowbore hplc system for separation and quantification (Dupont *et al.* 1989).



TEXT-FIG. 2. Representation of the size and distribution of the major protein bands from the > 10 kDa fraction of the organic extract of Recent brachiopods, separated by SDS–PAGE. Note the differing numbers and sizes of the proteins in the samples. (Original photographs of the SDS–PAGE separated samples are available from the senior author on request.)

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 by HCl without prior bleaching. Concentrations of amino acids were converted to weight percentages (wt%) using the EXCEL™ spreadsheet and analysed by the statistical program DATADESK™ on the Macintosh microcomputer.

Direct confirmation of the presence of intracrystalline amino acids

Within their shells, brachiopods contain organic molecules with antigenic properties (Collins *et al.* 1988). 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 *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

Gel electrophoresis

A composite gel of the samples separated by SDS–PAGE is shown in Text-figure 2. In all cases, with the exception of *Liothyrella neozealandica*, up to 20 replicates of each sample were completed. There was only sufficient material of *L. neozealandica* to allow a single replicate. Samples are grouped according to their taxonomic positions given by Williams *et al.* (1965).

Neothyris lenticularis, *Waltonia inconspicua* and *Terebratella sanguinea* (Order Terebratulida, Suborder Terebratellida) all show several major bands of approximate molecular weights 49, 16 and 6·5 kDa. *Liothyrella neozealandica* (Order Terebratulida, Suborder Terebratulidina) has 4 main bands of approximate molecular 42, 22, 16 and 6·5 kDa. *Notosaria nigricans* (Order Rhynchonellida) has 4 main bands of approximate molecular weights 50, 20, 17 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.

Direct confirmation of the intracrystalline nature of the amino acids

Results indicated that for every milligram of powder analysed in this way, there are 0·7 nanograms of amino acid. The shell powder was frozen and lyophilized again, and the inorganic phase demineralized by 2 N HCl as above. The concentration of amino acid yielded by this process was 169·4 ng/mg, slightly lower than the average for Recent *Terebratella 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 demineralization must therefore have been enclosed within the shell of the brachiopod (i.e. intracrystalline).

Absolute abundance of amino acids

Amino acid analysis of the samples that were not bleached shows that the molecules extracted from the intracrystalline fraction account in all three cases for 30–40% of the total amino acid present within the shell. On average, less than 10% of the total amino acid present within the shell of Recent brachiopods is in the form of free amino acids (Table 2), indicating that the vast majority are combined into proteins and peptides. A notable exception to this is tyrosine in *Notosaria nigricans*, which is *c.* 75% free. As there are generally very few free amino acids in these Recent samples, the indications are 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. Sample variability was between 5 and 8%. Concentrations which had a higher variability than this were repeatedly re-analysed until a series of consistent results was produced. *Liothyrella neozealandica* and *Neothyris lenticularis* both contain low concentrations of amino acid, which contrasts with the high concentrations found in *Notosaria nigricans*, a feature comparable to the values found for the total organic matter contained within the shell (Curry *et al.* 1989).

Decalcification of the *Notosaria nigricans* 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 partly of an acid-insoluble fraction, which is more resistant to the oxidative effect of the exposure to sodium hypochlorite (Collins *et al.* 1991*b*). The insoluble compounds were removed by centrifugation before amino acid analysis of the soluble fraction.

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 standardization is necessary. In this study, the concentrations were converted to weight percentages. These relative abundances are shown in Table 2, which shows the variation in the amino acid content of the samples. These datasets are based on between 2 and 10 replicates of each sample, with 2 to 4 plotted in each case, depending upon the number of replicates. The most striking variations lie in high Asp/Asn in *Notosaria nigricans* (36·94 wt%) and *Liothyrella 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. *Neothyris lenticularis* and *Neothyris parva* contain high Gly (51·51 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 from the matrix proteins considered in these other studies.

The low proportion of free amino acids within the samples (Table 2) is important as it indicates intact protein and peptide survival within the biocrystals of Recent shells, and also the absence of contamination by sample handling (Walton and Curry 1991).

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. 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.

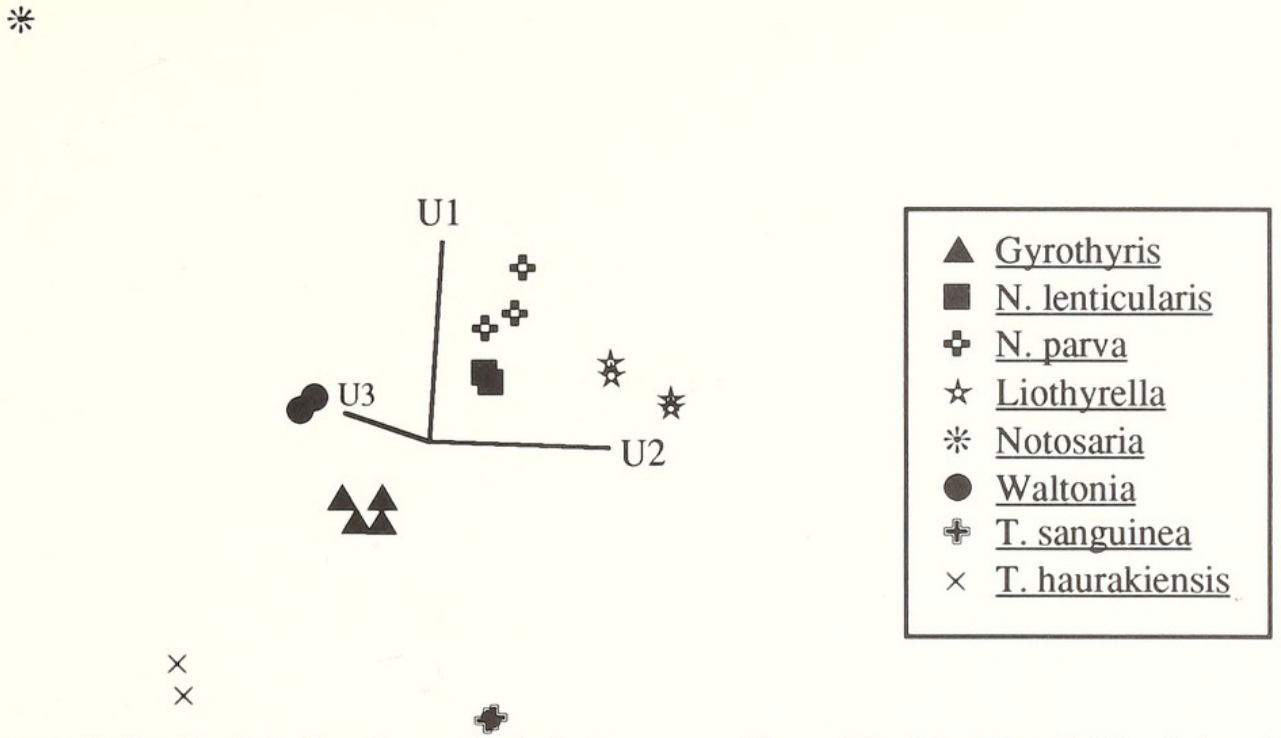
Direct comparisons of the relative abundance of amino acids between samples are hard to make, as the overall change of so many variables is difficult to observe. The illustration of the scale and direction of variation of the amino acid content of the samples is not possible, as this would require

TABLE 2. Amino acid composition of the total organic extract from Recent brachiopods. Absolute concentrations in ng/mg, relative proportion in weight percentage.

	Amino acid													
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	2·99
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 2. Amino acid composition of the total organic extract from Recent brachiopods. Absolute concentrations in ng/mg, relative proportion in weight percentage.

	Amino acid													
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	2·99
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

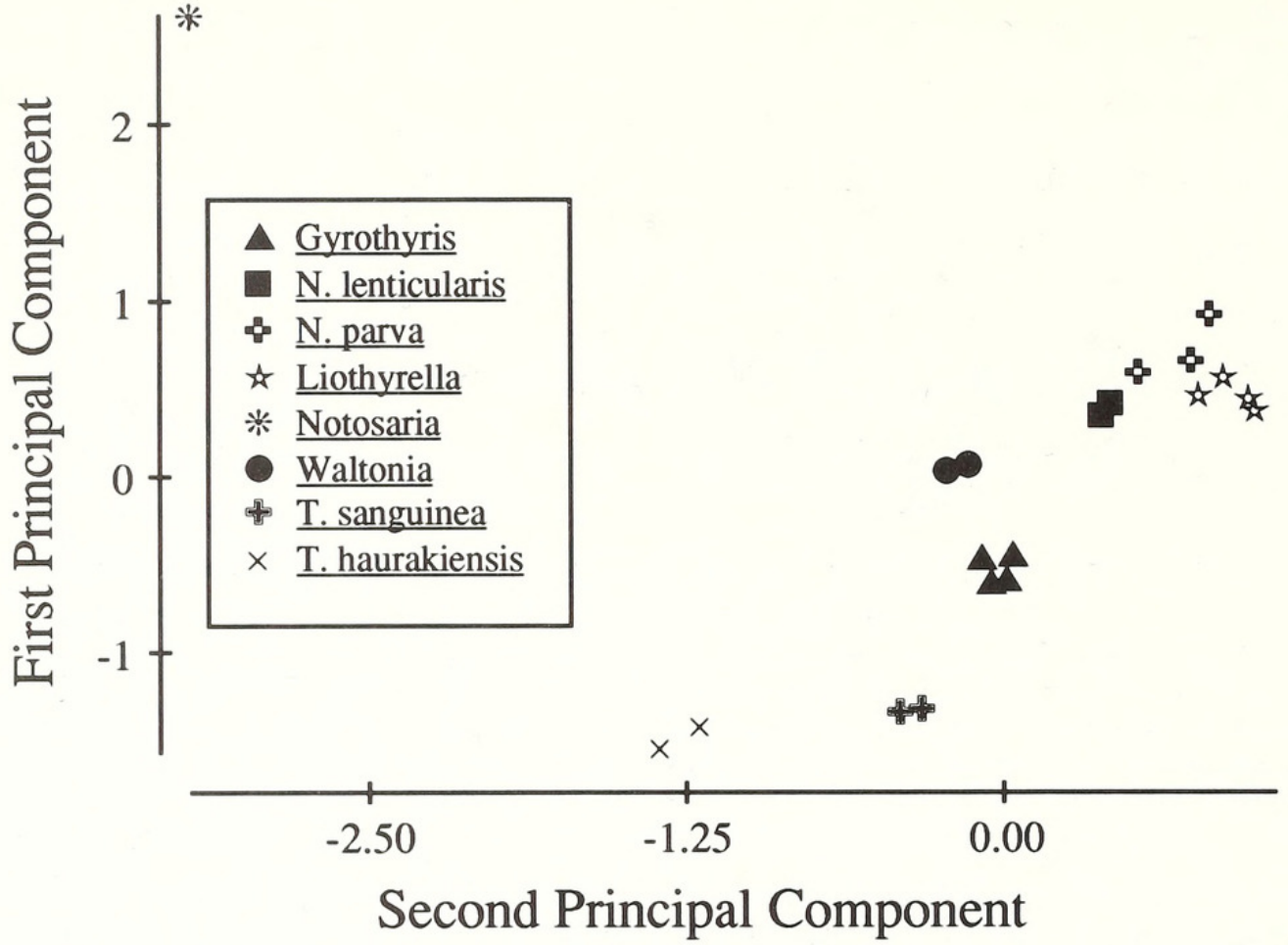


TEXT-FIG. 3. Graph of the first three principal components (denoted by U1, U2 and U3) of the relative proportion data for amino acids from Recent brachiopods (complete organic extract). Note close groupings of related species (e.g. *Neothyris lenticularis* and *N. parva*) and the distance of more distantly related species (e.g. *Notosaria*).

multi-dimensional space which cannot be used. For this reason, multivariate statistical techniques were applied to the dataset to summarize the variables and produce more meaningful information regarding the variation of the samples. Principal component analysis (PCA) is a method of analysing the dataset in a way which summarizes it into derived variables, which may then be compared in graphical form. Text-figure 3 shows the plot of the first three principal components extracted from the dataset, presenting a summary of the data in which most of the variation is contained within fewer, derived variables. The new dataset is plotted into a rotating plot, allowing three dimensional observation (Text-fig. 3), and onto a scatterplot (Text-fig. 4). In Text-figures 3 and 4, samples that are closely related plot 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 difference between the genomes of the organisms is reflected by different amino acid compositions. PCA 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 3) of the variability of the dataset, and this variation is mainly due to differences between samples in the proportion of Ser (−0·328), 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 differences in the proportion of Gly (0·332), Tyr (−0·500), Val (0·335) and Phe (−0·362). The third principal component contains 10·5% of the variation, due mainly to variation in the abundance of Asp/Asn (−0·304), Gly (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 these data are valid (Sneath and Sokal 1973).

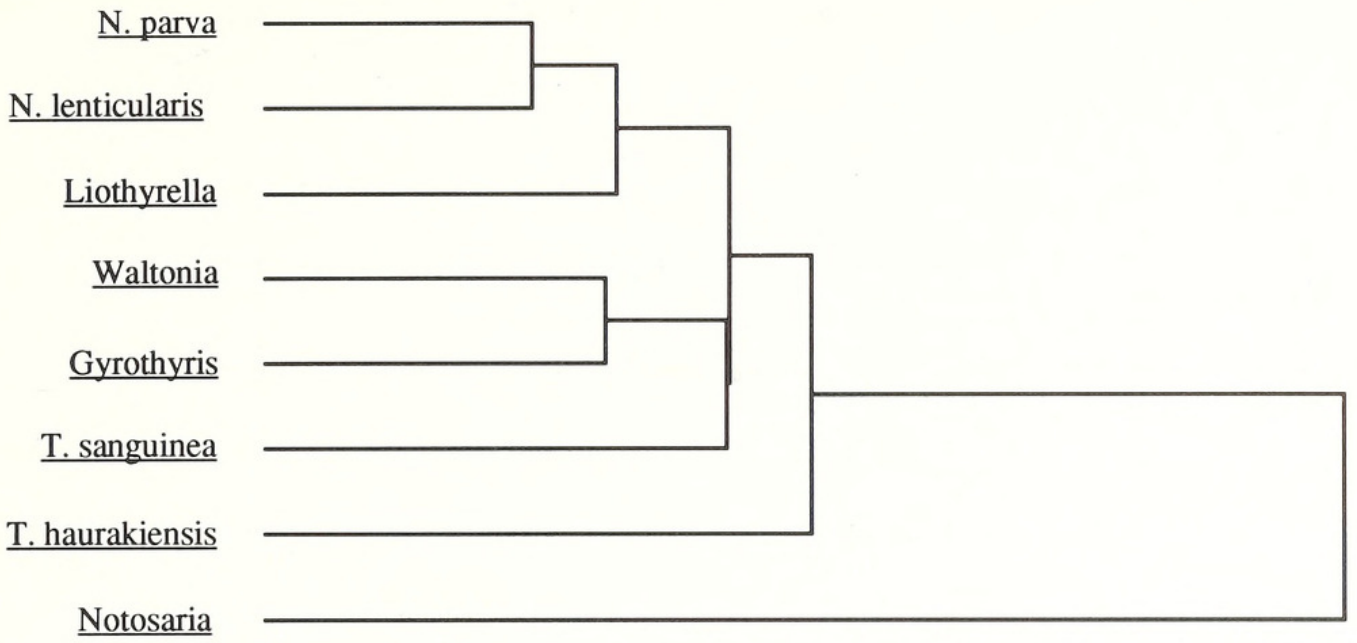
Taxonomic information is difficult to derive from the principal component data alone, as this method provides only discrete groupings of samples which are not linked. The method, at this scale at least, is unable to separate the two species of the genus *Neothyris*. Cluster analysis is a method of assaying the degree of 'relatedness' of the dataset, and the variables were analysed by single



TEXT-FIG. 4. Scatterplot of the first two principal components of the relative proportions of amino acids from Recent brachiopods.

TABLE 3. Principal component analysis of the relative proportion data, calculated for samples of Recent brachiopods. Only the first three eigenvectors and eigenvalues are given.

	Eigenvectors				Eigenvalues	
	V1	V2	V3		value	variance proportion
D	0·267	−0·294	−0·304	e1	6·694	47·8
E	−0·091	0·437	0·019	e2	3·839	27·4
S	−0·328	−0·032	0·149	e3	1·470	10·5
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			



TEXT-FIG. 5. Single linkage cluster analysis showing apparent relationships of extant brachiopods from New Zealand. Note the aberrant position of *Liothyrella*, caused by chance similar ratios of amino acids. *Liothyrella* contains a different number and size of proteins (Text-fig. 2), indicating that the extracts should only be compared with caution.

linkage analysis, which is equivalent to the nearest neighbour analysis of Sneath and Sokal (1973). The dendrogram in Text-Figure 5 is derived from cluster analysis of the relative proportion data shown in Table 2. 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. *Waltonia inconspicua* and *Gyrothyris mawsoni* are closely related. The two species of the genus *Terebratella*, however, do not appear to be as closely related as would be expected, while *Liothyrella neozealandica* appears to be related to the genus *Neothyris*.

DISCUSSION

Application of this method to numerical taxonomy

Numerical taxonomy was defined by Sneath and Sokal (1973, p. 4) as 'the grouping by numerical methods of taxonomic units into taxa on the basis of their character states', and implied repeatability and objectivity. Any method which relies on objective data collection may therefore be utilized in a study of numerical taxonomy. The work of Degens *et al.* (1967) provides a theoretical and practical basis for the characterization of species by the amino acid composition of proteins from their body tissue, and illustrates the use of multivariate statistical analysis of amino acid differences for taxonomic purposes using the principles of numerical taxonomy, i.e. utilizing non-subjective criteria.

The theory of the relationship of proteins based on their differences in amino acid composition has important implications for the use of amino acid analysis of intracrystalline molecules as a taxonomic tool. The technique has been criticized in the literature for not providing enough data on the taxonomic difference between samples, and also for being difficult to interpret (e.g. Collins *et al.* 1991; Logan *et al.* 1991). 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 by Walton and Curry (1991). In that study, the amino acid compositions of finger tips and sediments showed a gross similarity with the relative proportions of amino acids present in fossils, a relationship which did not stand up to reappraisal by more rigorous multivariate statistical techniques which revealed a clear discrimination between the data. Comparisons of the

ratios of amino acids are misleading, and the use of more powerful multivariate statistical techniques, such as cluster analysis and PCA, can lead to more certain discrimination between data points.

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 entirely consistent with the phylogeny produced from conventional morphological 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, rendering interpretation of the data difficult. Q-mode factor analysis of shell proteins was used in a phylogenetic study of Recent and fossil planktonic foraminifers (King and Hare 1972), where the data were expressed on a triangular plot, representing the equivalents of the first three principal components (cf. Text-fig. 3). King and Hare (1972) were also the first to conclude that chemotaxonomy of amino acids was viable in extinct species. Haugen *et al.* (1989) completed a chemotaxonomic study of benthic foraminifers using amino acid data and combined the use of ratios and PCA to infer relationships between the species.

All the above studies utilized analysis of the bulk extracts of the shell (i.e. both intercrystalline and intracrystalline molecules) to examine taxonomy. Intercrystalline molecules are in a position whereby they may be easily contaminated. This possibility is not present when the intracrystalline molecules alone are analysed, as is the case in this study.

Taxonomic conclusions

The following taxonomic conclusions may be reached for the samples in this study. These conclusions utilize the results from both SDS–PAGE analysis of samples as well as the amino acid analyses. The amino acid concentrations are used entirely objectively. Samples of the genus *Neothyris* are considered to be members of separate species, in terms of their position in the cluster diagrams (Text-fig. 5). The plot of the first three principal components (Text-fig. 3) is unable to separate the samples, indicating that they are closely related. *Gyrothyris mawsoni* and *Waltonia 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 on invertebrate paleontology* (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*. As these species are classified as being members of the same genus, it was expected that they would plot together on the PCA plot (Text-fig. 3), and form a discrete cluster (Text-fig. 5), similar to the two species of *Neothyris*. However, the bulk amino acid compositions of the two species of *Terebratella* are markedly different (Table 2), with *T. haurakiensis* having a higher overall concentration of the amino acids, and also a different relative proportion of the amino acids. This leads to the separation of the data points for these samples in both the PCA plots and in cluster analysis. It therefore appears likely, from these bulk amino acid data, that the classification of this genus may need revising. Further work on the composition and size of the intracrystalline molecules from *Terebratella* species, especially from an area where the two species coexist, is necessary in order to confirm the need for revision of the classification. Unfortunately, only a limited amount of material was available for the study of *T. haurakiensis*, rendering it impossible to employ SDS–PAGE to assess whether the differences in the amino acid composition between the two samples are due to changes in the actual composition of similar sized proteins in the species, or whether the samples contain different numbers of proteins of different sizes and are therefore more distantly related than currently thought (cf. the amino acid composition of *Liothyrella neozealandica* below). 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 *Waltonia inconspicua*, which is 'not uncommon' (D. E. Lee pers. comm.), or it could be due to a more fundamental difference, as the amino acid data would tend to suggest.

The species *Liothyrella neozealandica* is problematical; although belonging to the same order as the Terebratellida, it is a member of a different superfamily. However, in PCA (Text-fig. 3) *L. neozealandica* plots close to the other data, and in the cluster analysis (Text-fig. 5) 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 artefact of the data, and that similar relative concentrations of amino acids are present in different proteins. The amino acid composition of these 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 *Notosaria nigricans* would be expected to be very much different from the remaining samples. This brachiopod differs from the remaining samples at the ordinal level, and also contains shell proteins of different molecular weights than those of 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 (Text-figs 5 and 3), which accurately record the differences between the amino acids present within the species. This shows similarity with the immunological investigation of Endo (1992), where antibodies produced in response to shell macromolecules from *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 with only objective methodology. The technique is rapid and requires very small quantities of material, making it available for the study of material in short supply, which is not available in quantities sufficient for protein sequencing. The relationships between the samples are easily seen by the graphical presentation of the PCA data.

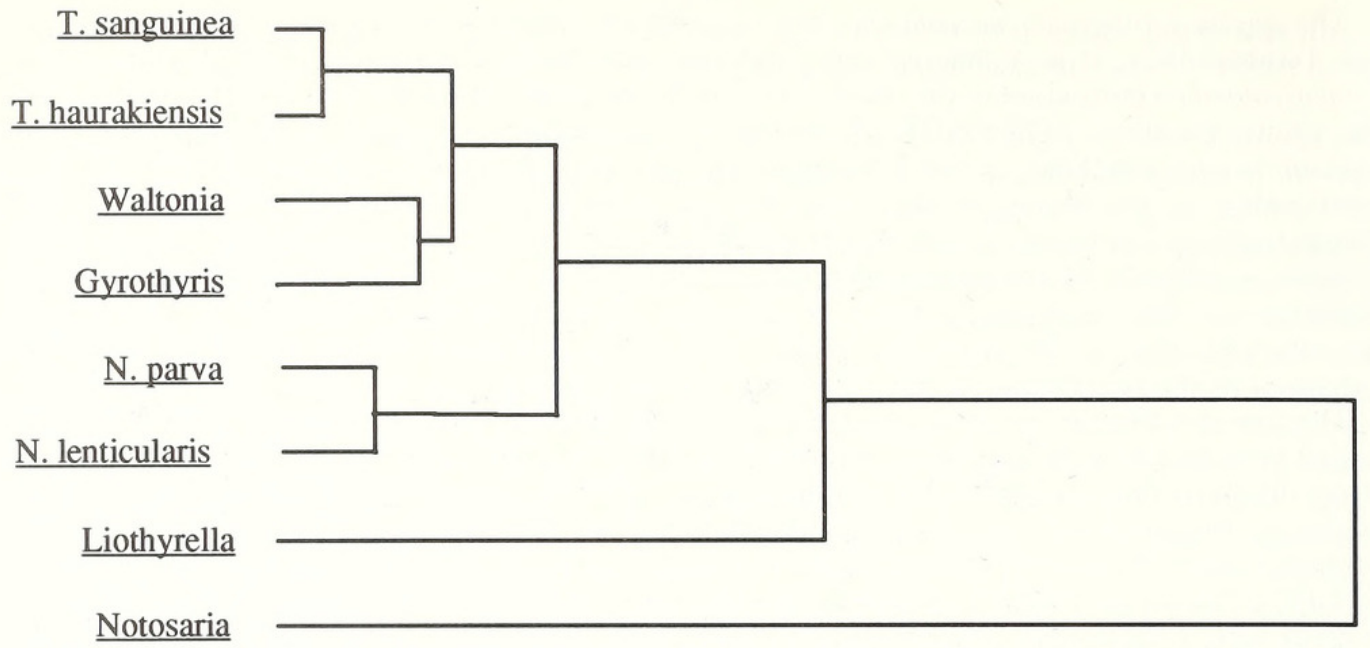
Anomalies between partial sequence and amino acid analyses

The high proportions of Gly found within the bulk amino acid composition of the samples contrast with the partial sequence data so far derived for brachiopod intracrystalline proteins (Curry *et al.* 1991*a*; Cusack *et al.* 1992) which include only moderate amounts of Gly. This could be due to different sample preparation methods: the proteins for sequencing undergo concentration by filtration which removes most compounds with a molecular weight of less than 10,000 (10 kDa), although this also depends on the conformation of the protein. Polypeptides with a molecular weight of less than this may contain larger abundances of Gly.

As discussed previously, this method has significant advantages over methods of protein sequencing in terms of rapidity and required sample size. An alternative approach would be to separate proteins from Recent brachiopods using SDS–PAGE and to obtain the amino acid composition of the purified protein. Taxonomic conclusions could then be reached by application of the theories of Cornish-Bowden (1979, 1983). This approach would increase the level of information gained from such molecules and has indeed been attempted by the authors. However, consistent results have not yet been achieved due to the interaction of the support membrane with the automated hydrolysis of the analysis system. This approach also has the same disadvantage as protein sequencing in requiring large amounts of shell powder for analysis, in contrast to the method described here.

Implications for the study of fossil molecules

The information contained within intracrystalline molecules was sealed by the mineralization 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



TEXT-FIG. 6. Dendrogram showing morphological relationships from the *Treatise* (after Williams *et al.* 1965).

composition of the intercrystalline shell matrix (e.g. Jope 1967*a*, 1987*b*; Konesnikov 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 is degraded in less than one year, and that after this time the fibres 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). These studies indicate that intracrystalline molecules from fossils should be a much more reliable source of historical information. In any study of fossil molecules, however, study of Recent samples is required as a starting point. 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 present.

The degradation products of the fossil equivalents of these Recent intracrystalline molecules will remain trapped within fossil shells until the inorganic phase is dissolved. The trapped molecules will record molecular taxonomic differences in the fossil record. This is in contrast to the changes between Recent and fossil molecules in the intercrystalline matrix of the shells, which may be diagenetic rather than genetic. The study described here is an essential starting point for the study of fossilized amino acids and proteins from brachiopod shells.

CONCLUSIONS

Taxonomic analysis of brachiopods using the weight percentage of intracrystalline shell amino acids reinforces the morphological taxonomy. Using the classification of the brachiopods from the *Treatise* (Williams *et al.* 1965), the data are divided into two orders, the Order Rhynchonellida, represented by *Notosaria*, and the Order Terebratulida, which encompasses all 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 these data, an approximate dendrogram may be produced (Text-fig 6) to show suggested relationships. It would be expected that *Liothyrella* and both species of *Neothyris* would plot away from the Terebratellinae.