Phytochemistry, 1979, Vol. 18, pp. 603–608. Pergamon Press. Printed in England.

RELATIONSHIPS BETWEEN THE PARTIAL AMINO ACID SEQUENCES OF PLASTOCYANIN FROM MEMBERS OF TEN FAMILIES OF FLOWERING PLANTS

D. Boulter*, D. Peacock†, A. Guise†, J. T. Gleaves† and G. Estabrook‡
Department of Botany, University of Durham, Durham, DH1 3LE, U.K.

(*Received* 22 *August* 1978)

Key Word Index—Angiosperm families; flowering plants; plastocyanin; cytochrome-*c*; amino acid sequence; plants; ancestral sequence; plant phylogeny.

Abstract—Partial amino acid sequences of plastocyanin from 40 members of 10 families of flowering plants were analysed using a compatibility method based upon the suggestions of numerical taxonomists. In addition, sequences from members of the Fabaceae, Solanaceae and Apiaceae were analysed separately by an ancestral amino acid sequence method. The relationships suggested by the data are discussed with reference to the plant cytochrome-*c* data set and the use of amino acid sequence comparisons in phylogenetic studies generally.

INTRODUCTION

Boulter [1] has used cytochrome-*c* amino acid sequence differences of flowering plants in an attempt to prove the usefulness of these characters in studies aimed at elucidating the evolutionary relationships of some families of flowering plants. Since the cytochrome-*c* data set contain parallel, back and presumably convergent substitutions, some distortion from the natural (phylogenetic) classification must occur when cytochrome-*c* data are used to construct topologies. In the absence of a natural system of classifications, which is known to be correct, it is difficult to assess the extent of the distortion in the cytochrome result. One possible approach to the problem is to compare the results obtained with cytochrome-*c* with those from a physiologically unrelated protein since it is unlikely that both will show similar convergent substitutions. To this end both complete and partial sequences (40 *N*-terminal residues) of some plastocyanins have also been determined. This paper describes the preliminary results obtained with 40 partial plastocyanin sequences and discusses these results in relationship to the present position of the usefulness of amino acid sequence data in phylogeny.

RESULTS

The species investigated and the families to which they belong are given in Table 1. The amino acid sequence data set are given in the accompanying appendix.

The ancestral sequence method [2] was applied only to a limited set of the plastocyanin data owing to the excessive computing time which would have otherwise been required; even so, many different equally parsimonious affinity trees were found. However a method based on the identification of the 'compatible' residues reduced the ambiguity in the results and gave the relationships summarised in Fig. 1. This method is based on an adaptation suggested by Estabrook and Landrum [3] and Sneath *et al.* [4] of the concept of compatibility put forward by Le Quesne [5] in an attempt to detect convergent evolution.

A set of characters defined in a group of species is said to be compatible if there is a pattern of divergence for the species that does not involve a parallel event in the evolution of the character states of any of the characters. A pair of characters, in this case residue positions in a protein amino acid sequence, can be tested for compatibility by setting up their co-occurrence table (see Table 2); an example is given below for two positions and six species (denoted a–f).

A series of 'rook's moves' through the table, beginning at species 'a' and moving horizontally or vertically from entry to entry, without reversing steps, will return after six moves to the starting point. The existence of such a closed path or circuit in the table can be shown to be a necessary and sufficient condition for the two positions to be incompatible. Any pattern of divergence for the six species requires that at least one parallel substitution to have occurred. For example, the substitution to alanine in species 'f' in the possible phylogeny is shown in Fig. 2.

Any residue position in a set of homologous sequences may be incompatible in this sense with a number of other positions. This method attempts to identify these parallel substitutions in the data by finding the group of species sharing a common amino acid at a particular position, for which the assumption of a parallel event explains the largest number of pair-wise incompatibilities. If the common amino acid in these species is replaced by a labelled version, i.e. by a nominally different amino acid to any present, then these incompatibilities disappear. By repeating this process and so removing the

---

* To whom reprint requests should be addressed.

† Present address: Department of Biochemistry, The University of Leicester, U.K.

‡ Present address: Department of Botany, University of Michigan, Michigan, U.S.A.

Table 1. Plants examined for plastocyanin amino acid sequences

Name	Family	Number key*
Pastinaca sativa (parsnip)	Apiaceae	5
Anthriscus sylvestris (cow parsley)		6
Heracleum sphondylium (hogweed)		7a
Heracleum mantegazzianum (giant hogweed)		7b
Aegopodium podagraria (ground elder)		8
Ursinia anethoides	Asteraceae	35
Bellis perennis (daisy)		32
Rudbeckia sp.		36
Tussilago farfara (colt's foot)		33
Hieracium sp. (hawkweed)		31
Cirsium vulgare (thistle)		30
Senecio vulgaris (groundsel)		34
Capsella bursa-pastoris (shepherd's purse)	Brassicaceae	37
Brassica oleracea (spring cabbage)		38a
Brassica oleracea (cauliflower)		38b
Viburnum tinus	Caprifoliaceae	18
Sambucus nigra (elder)		19
Lonicera periclymenum		20
Cytisus battendieri	Fabaceae	9
Lupinus sp.		10
Robinia pseudoacacia		11
Daviesia latifolia		12
Pisum sativum (pea)		13
Trifolium medium		14
Vicia faba (broad bean)		15
Phaseolus vulgaris		17
Vigna radiata (mung bean)		16
Magnolia soulangeana	Magnoliaceae	1
Liriodendron tulipifera		2
Plantago major	Plantaginaceae	29
Prunus serrulata splendens	Rosaceae	3
Crataegus monogyna (hawthorn)		4
Antirrhinum majus (snapdragon)	Scrophulariaceae	26
Digitalis purpurea (foxglove)		27
Verbascum thapsus (mullein)		28
Lycopersicon esculentum (tomato)	Solanaceae	21
Solanum tuberosum (potato)		22
Nicotiana tabacum (tobacco)		23
Solanum crispum		24
Capsicum frutescens (pepper)		25

* Numbers refer to the taxa of Figs. 1, 3, 4 and 5.

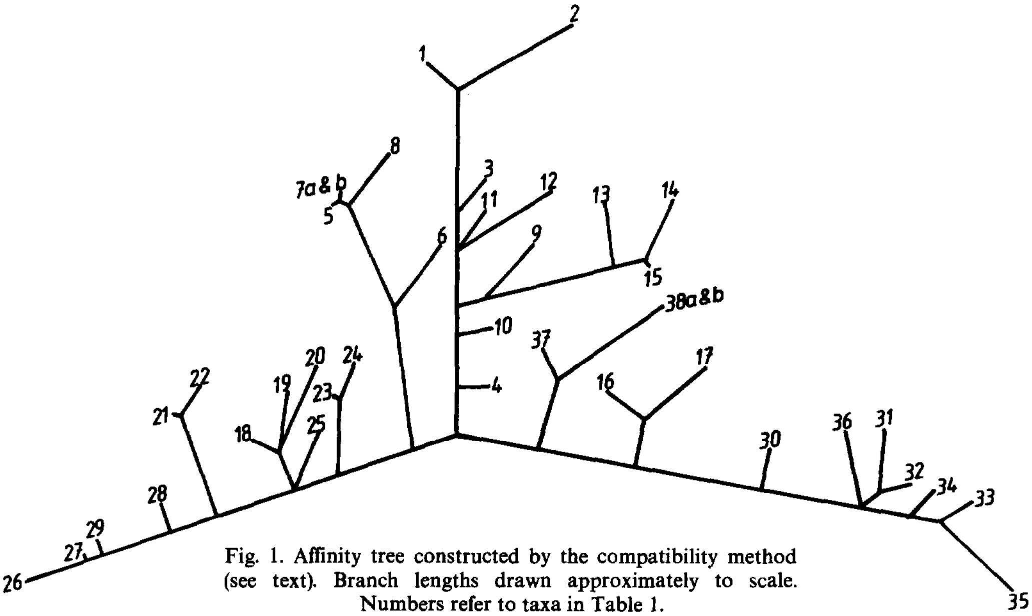

Fig. 1. Affinity tree constructed by the compatibility method (see text). Branch lengths drawn approximately to scale. Numbers refer to taxa in Table 1.

Table 2. Co-occurrence table

		Position 1			Amino acid residues at two positions (1 and 2)		
						1	2
		Ala	Ser	Gly	Species a	Ala	Ile
Position 2	Ile	a	b		b	Ser	Ile
	Val		c	d	c	Ser	Val
	Leu	f		e	d	Gly	Val
					e	Gly	Leu
					f	Ala	Leu

Co-occurrence table for two positions. The symbols a–f represent the species listed on the right.

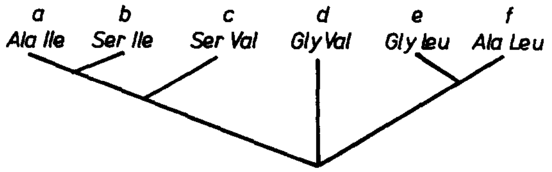

Fig. 2.

largest number of incompatibilities possible at each stage, the method produces a modified set of sequences with one or more labels marking the sites of possible (or suspected) parallel substitutions before tree building is started. Thus amino acid residues with different labels attached are nominally considered as different amino acids.

The labelled sequence data set is then used to construct a topology using the parsimony method and it is hoped that convergent similarity, as opposed to ancestral similarity, has been largely excluded; the topology is constructed by analytical methods made possible because of the reduction in ambiguity. The rationale for the method and full details of its operation will be the subject of a separate publication.

Figure 1 is a diagrammatic summary of the common features of different affinity trees constructed with the method by using the complete data set (40 partial sequences) and several different sub-sets of it.

Apart from a few exceptions, e.g. *Phaseolus vulgaris* and *Vigna radiata* (formerly *Phaseolus aureus*) which did not group with the rest of the Fabaceae, plastocyanin amino acid sequences belonging to members of a taxonomic family grouped accordingly. In the several topologies generated during the course of this investigation, the amino acid sequences from members of the Solanaceae were always separated from those of the Asteraceae. Furthermore, it was always possible to separate some of the families into three groups: (1) Asteraceae, (2) Solanaceae, Scrophulariaceae, Plantaginaceae, Caprifoliaceae and (3) Fabaceae, Rosaceae; the sequences of the Fabaceae, Rosaceae group were intermediate in similarity between the Asteraceae and Solanaceae. Although the sequences of the Apiaceae, Brassicaceae, and Magnoliaceae grouped according to taxonomic family, it was not possible to place these families in certain relationship to the three groups given above. In the case of the Apiaceae, there were similarities between their sequences and those of all three groups. In the case of Brassicaceae, and Magnoliaceae, the greatest similarity in the sequences was to the Fabaceae, Rosaceae group. Recently Wagenitz [6] has made a strong case on morphological and chemotaxonomic evidence that the families of the Asteridae are a monophyletic group. The protein data, however, do not support this suggestion. Figures 3, 4 and 5 are sequence topologies of some members of the Apiaceae, Solanaceae and Fabaceae obtained with the ancestral sequence method [2] as modified by Gleaves (unpublished). The global minimum was found by the search procedures used since all possible trees were evaluated. The sequence topology of the Asteraceae has been published already [7].

So far as the results presented in Figs. 3, 4 and 5 are concerned, since so few taxa are represented they can only be used as an indication of the potential of the method rather than to give taxonomic insight. Generally speaking, the Solanaceae and Apiaceae sequence results agree with the existing classifications although it is of interest to note the closer proximity of potato to tomato,

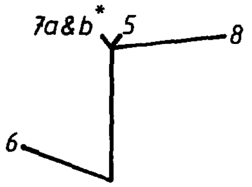

Fig. 3. Affinity tree constructed by a total search procedure (Gleaves, unpublished) using the ancestral sequence method [2]. Branch lengths drawn approximately to scale. * Two species of *Heracleum* had identical sequences which only differed from that of *Pastinaca* by the type of heterogeneity present.

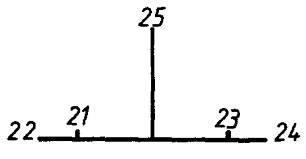

Fig. 4. Constructed as in Fig. 2.

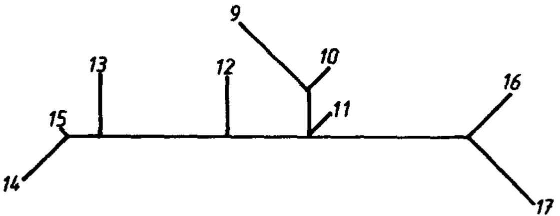

Fig. 5. Constructed as in Fig. 2.

than to the woody *Solanum*. With regard to the Fabaceae, it is well established that genera have often been ascribed to tribes artificially and that there is no satisfactory arrangement of the tribes. The most comprehensive treatment is that of [8], but the results presented here do not support this arrangement, particularly with regard to the placing of *Daviesia*, *Phaseolus* and *Vigna*.

DISCUSSION

In view of the uncertainties in interpretation of topologies based on amino acid sequence data from plants (see later) and also for that matter from animals (see [9]), an assessment of the current situation is called for, especially in the light of some interpretations, e.g. [10].

Statements by Crick [11] in 1958 and Zuckerkandl and Pauling [12] in 1965 led to the ready and optimistic acceptance by many molecular biologists of the usefulness, as evolutionary markers, of comparisons of amino acid sequences of proteins from extant organisms. The first results, with a single protein cytochrome-*c* from vertebrates, were highly compatible with fossil evidence and with the classical ideas of the phylogeny of this

Table 3. Amino acid partial sequence data for plastocyanin

Plant	Amino acid partial sequence (from *N*-terminus)
Pastinaca	A E V K L G G D D G G L V F S P N S F T V A A G E K I T F K N N A G F P H N I V
Anthriscus	A E V K L G G D D G S L A F V P S S I T V A S G E T I T F K N N A G F P H N I V
	N
Heracleum (hogweed)	A E V K L G G D D G G L V F S P N S F T V A A G E K I T F K N N A G F P H N I V
Heracleum (giant hogweed)	A E V K L G G D D G G L V F S P N S F T V A A G E K I T F K N N A G F P H N I V
Aegopodium	A E L K L G A D D G G L V F S P S S F T V A A G E K I T F K N N. . . . .
	V
Capsella	I E V L L G G G D G S L A F V P N D F S I A K G E K I V F K N N A G F P H N V V
Brassica (spring cabbage)	I D V L L G S G D G A L A F V P N E F T I A K G E K I V F K N N A G F P H N V V
Brassica (cauliflower)	I D V L L G S G D G A L A F V P N E F T I A K G E K I V F K N N A G F P H N V V
Viburnum	I E I L L G G D D G S L A F V P G N F S V P S G E K I V F K N N A G F P H N. .
	V A
Sambucus	V E I L L G G E D G S L A F V P G N F S V P S G E K I T F K N N A G F P H N V V
	I S
Lonicera	I E V L L G G D D G S L A F V P G S F T V P S G E K I I F K N N A G F P H N V V
Cirsium	V E V L L G A S D G G L V F E P S T F T V A S G E K I V F K N N A G F P H N V V
	I N
Bellis	I E V L L G D N N G A L V F E P K T L S V A S G E K I V F K N N I G F P H N V V
	A G V S F
Hieracium	V E V L L G D N D G G L V F E P S T L S V A S G E K I V F K N N S G F P H N V V
Tussilago	I E V L L G D N D G G L N F E P S T F T V A P G E K I V F K N N S A F P H N V V
	V E A V L G
Senecio	I E V L L G D N D G G L A F E P S N F S V A A G E K I V F K N N S G F P H N V V
	A V T
Rudbeckia	L E I L L G G N L G G L V F E P A N F S V A A G E K I V F K N N S G F P H N V V
	D V D D A T P
Ursinia	A E V L L G D N D G G L N F E P A T L T V P A G E K I V F K N N S A F P H N V V
	M E A V K F I G
Cytisus	V E V L L G S D D G G L A F V P D N F S V S A G E K I V F K N N A G F P H N V V
Lupinus	V E V L L G S D D G G L A F V P N D F S V N P G E K I V F K N N A G F P H N V V
Robinia	V E V L L G G D D G S L A F V P N D F S V S P G E K I V F K N N A G F P H N V V
Daviesia	I E V L L G A S D G S L A F V P N S F S V S P G E K I T F K N N A G F P H N V V
Pisum	V E V L L G A S D G G L A F V P S S L E V S A G E T I V F K N N A G F P. . . .
Trifolium	V E V L L G A S D G G L A F V P N N F T V S A G D T I V F K N N A G F P H N V V
Vicia	V E V L L G A S D G G L A F V P N S F E V S A G D T I V F K N N A G F P H N V V
Vigna	L E V L L G A G D G S L V F V P S D F S V A S G E E I V F K N N A G F. . . . .
Phaseolus	L E V L L G S G D G S L V F V P S E F S V P S G E K I V F K N N A G F P H N V V
Magnolia	I E V L L G G S D G T L A F V P K E F S V S P G E K I V F K N N A G F P H N V V
Liriodendron	I E V L L G D S D G N L V F V P K E F S V A P G E K I V F K N N A G F P H N I V
Plantago	M D V L L G G D D G S L A F I P G S F E V A A G E K I T F K N N. . . . .
Prunus	I E V L L G G D D G S L A F V P N S F S I S P G E K I V F K N N A G F P H N I V
Crataegus	I E V L L G S D D G G L A F V P S S F S V A P G E K I V F K N N A G F P H N I I
	N
Lycopersicon	L E V L L G G D D G S L A F I P G N F S V S A G E K I T F K N N A G P F. . . .
Solanum tuberosum	L D V L L G G D D G S L A F I P G N F S V S A G E K I T F K N N A G F P H N V V
Nicotiana	I E V L L G S D D G G L A F V P G N F S V S A G E K I T F K N N A G F P H N V V
Solanum crispum	I E V L L G S D D G G L A F V P G N F S I S A G E K I T F K N N A G F P H N V V
Capsicum	I E V L L G G D D G S L A F V P G T F S V. S G E T I T F K N N A G F P H N V V
Antirrhinum	L D V L L G G D D G S L A F I P G T F E V A A G E K I V F K N N A G F P. . . .
	V
Digitalis	L D V L L G G D D G S L A F I P G S F E V A A G E K I T F K N N A G F P H N V V
Verbascum	I E V T L G G D D G S L A F I P Q N F E V A A G E K I V F K N N A G F P H N V V

* Full species names and families are given in main paper. Isolated residues refer to the sequence where the position is heterogeneous.

group. As additional results from other proteins such as haemoglobin were obtained, so-called 'anomalous positionings' became apparent. Data obtained with different proteins which supported one another and the accepted phylogeny were still common, but in some cases protein data, although internally consistent, disagreed with the classical picture. In addition, there were examples where the results from different proteins disagreed among themselves (see [9]), and the pendulum has now swung to the opposite extreme, e.g. Williams [13] concludes that "the prospect is not exciting for the protein chemist". Williams [13] has pointed out that so far macromolecular data have not given many phylogenetic insights when compared with morphological characters. However, to date, there have been very little molecular data assembled compared to the vast amount of morphological data available, since taxonomists have traditionally used morphological characters. In any case, since the existing classifications can be thought to approximate to the natural one in some groups, new methods cannot be expected to give many new insights. *A priori* proteins should be excellent taxonomic characters since they are the direct expression via *m*RNA of the genetic information of an organism, and further they are chemically complex and usually monogenic.

A posteriori, however, sequence data are seen to suffer from the same limitations of interpretation as do other characters of present day organisms when used to reconstruct phylogenetic relationships. The major distorting factors are unequal rates of evolution and the fact that convergent or parallel substitutions have occurred. How far the molecular clock varies is still a matter of considerable debate [14] but in any case, as pointed out by Boulter *et al.* [15], methods such as the ancestral amino acid sequence method and other parsimony methods may not give seriously distorted results even though rates of evolution have not been constant. The problem of convergent and parallel evolution however can be grossly distorting [16]. It is for these reasons that different protein data sets vary in their usefulness with the same group of organisms, i.e. they may reflect different rates of change, degrees of parallel evolution etc., so that different proteins from the same group of organisms may give conflicting results. A further distorting factor likely to be of importance, is failure to recognise paralogous, as opposed to orthologous, gene products although many proteins do not exhibit this type of behaviour.

In addition, different methods of data-handling vary in that they contain different assumptions either with regard to the mechanism of evolution itself, e.g. whether or not evolutionary rates are assumed to be constant, whether parsimony is accepted, or in their alogorithms, e.g. the measure of differences used and the method of selection of the desirable tree. This must be borne in mind when two proteins from the same group of organisms have been subjected to different data-handling methods. These conceptional, methological and technical difficulties lead us to the conclusion that it is essential to use several genetically and physiologically unrelated proteins from each organism in order to establish an acceptable phylogeny. The more proteins investigated the closer we may come to establishing the phylogeny, but it is essential to use amino acid sequence data in conjuction with other available biological data. Cronquist [10] suggested that the importance of amino acid sequence data is that they can act as an independent check to 'classical' schema. Such statements are symptomatic of the incorrect tendency to place the use of amino acid data into an 'all or nothing' position. The difficulties in considering the evidence from 'classical taxonomy' however are considerable. As Heywood [17] points out "there is room for considerable debate about content and circumscription" at all levels, family, order and even sub-class. Again in speaking of phylogenetic trees, he rightly states that "trees are basically a false representation of the pathways and patterns of evolutionary processes. There is seldom any discernible basis for the branching pattern, not even for the major branches which in the Angiosperms are in dispute". Heywood [17] concludes that "practically nothing is known about the course of evolution in the Angiosperms". Basically, there was little disagreement with this viewpoint at the recent Hamberg meeting—see [18].

Against this background the criticism of Cronquist [10] that the cytochrome-*c* amino acid data conflict with the widely used schemes loses much of its impact. Cronquist's second main criticism of the amino acid sequence data is that the molecular tree is not in agreement with the fossil record. Here one must distinguish between the statements about possible time of orgin of the Flowering Plants and those about the tree itself. It was clearly recognised by Ramshaw *et al.* [19] that the dating of the origins depended on several assumptions not least the evolutionary clock hypothesis; Cronquist [10] chose to quote our statements out of context so that this point was omitted. So far as the molecular tree not fitting the fossil record goes, the force of the criticism depends on one's interpretation as to how good a fossil record exists. Walker [20] concluded that while tricolpate-derived pollen can be identified in the fossil record as being of angiosperm origin rather than gymnosperm, it can inform us little about the early evolution of the angiosperms. At the critical biological period, pollen of primitive angiosperms and their ancestors was monosulcate and consequently difficult, if not impossible to distinguish from gymnosperm pollen. Stebbins [21] makes the point that in the absence of adequate fossil record (which still does not exist in the angiosperms) it would prove impossible to reconstruct it by extrapolation from modern plant families.

Some phylogenetic considerations with regard to the plant data

When we started on the present amino acid sequence programme several years ago, there was only one complete sequence of a plant protein available, that of wheat germ cytochrome-*c* [22]. At that time the priority was to increase the amount of data. Now helped by the advent of automatic sequence methods [23], extensive plant cytochrome-*c* and plastocyanin data sets exist as well as limited data on ferredoxins. Whilst automatic sequencing has quickened the pace of data acquisition and to some extent changed the strategy of its acquisition, e.g. determination of the *N*-terminal 40 residues or so, such data, with its lack of redundant information, must be used cautiously. Thus only a very low level of error in the determination of amino acid sequences is permissible as sometimes one or two incorrect residue positions can give a change in the predicted phylogeny.

The results obtained with cytochrome-*c* led us to several speculations [1, 24] which it was hoped would be supported or refuted by the plastocyanin data presented here. The rate of evolutionary change of plastocyanin is fast compared to cytochrome-*c* however, and this makes plastocyanin useful at the familial and tribal levels rather than at greater evolutionary distances, and confirmation or otherwise must now await the sequencing of a protein whose rate of change is more similar to that of cytochrome-*c*; it would take too much work to use plastocyanin since only data from relatively closely related families can be compared so that very many families would have to be investigated to cover the evolution of the Angiosperms. In presenting the cytochrome-*c* tree in 1972 [15] and pointing out that it did not support the widely accepted schemes of Cronquist, Takhtajan and Thorne, we did so in the spirit of Constance's suggestion [25], i.e. to serve as an intellectual stimulus and to focus attention on the fact that the basis of such schemes, as also pointed out by Heywood [17], are rarely published.

The facts that amino acid sequences from closely related taxa are more similar (e.g. plastocyanin from the two varieties of *Brassica* are identical, sequences from genera clearly belonging to a family are more similar than to any others suggest that much of the results make taxonomic sense). Even so, at this stage in the investigation it is not possible to say to what extent amino acid sequence data give new insights into the phylogeny of the Flowering Plants since 'random' evolutionary amino acid substitutions lead to distortions when 'trees' are constructed from amino acid sequence data using the present data handling methods. The distinction between useful new information and the distortions will only become fully apparent when more protein data sets are assembled and it is better to await phylogenetic speculation till that time. In the meantime, scheme formulators should consider the strength of their evidence. New data handling methods perhaps along the lines of those outlined in this paper would also be a great benefit.

Acknowledgements—We thank Mr. J. Gilroy, Miss E. Jones and Mrs. R. McIntosh for their technical help.

REFERENCES

1. Boulter, D. (1974) in *Chemistry in Botanical Classification* (Bendz, G. and Santesson, J., eds.) Nobel Symposium No. 25, p. 211. Academic Press, New York.
2. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5. National Biomedical Research Foundation, Silver Spring, Maryland, U.S.A.
3. Estabrook, G. F. and Landrum, L. (1975) *Taxon* **24**, 53.
4. Sneath, P. H. A., Sackin, M. J. and Ambler, R. P. (1975) *Syst. Zool.* **24**, 311.
5. LeQuesne, W. J. (1969) *Syst. Zool.* **18**, 210.
6. Wagenitz, G. (1977) in *Flowering Plants, Evolution and Classification of Higher Categories* (Kubitzki, K., ed.) p. 375. Springer, Wien.
7. Boulter, D., Gleaves, J. T., Haslett, B. G. and Peacock, D. (1978) *Phytochemistry* **17**, 1585.
8. Hutchinson, J. (1964) *The Genera of Flowering Plants*, Vol. 1. Clarendon Press, Oxford.
9. de Jong, W. W., Gleaves, J. T. and Boulter, D. (1977) *J. Mol. Evol.* **10**, 123.
10. Cronquist, A. (1976) *Brittonia* **28**, 1.
11. Crick, F. H. C. (1958) *Biological Replication of Macromolecules*, Symp. Soc. Exp. Biol. No. 12, p. 138. Cambridge University Press.
12. Zuckerkandle, E. and Pauling, L. (1965) *J. Theor. Biol.* **8**, 357.
13. Williams, J. (1974) *MTP International Review of Science, Chemistry of Macromolecules* (Gutfreund, H., ed.) Biochemical Series Vol. 1, p. 1. Butterworth, University Park Press.
14. Wilson, A. C., Carlson, S. S. and White, T. J. (1977) *Annu. Rev. Biochem.* **46**, 573.
15. Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M. and Brown, R. H. (1972) *Proc. Roy. Soc. Lond.* **181**, 441.
16. Peacock, D. and Boulter, D. (1975) *J. Mol. Biol.* **95**, 513.
17. Heywood, V. H. (1977) in *Flowering Plants, Evolution and Classification of Higher Categories* (Kubitzki, K., ed.) p. 1. Springer, Wien.
18. Kubitzki, K. (1977) in *Flowering Plants, Evolution and Classification of Higher Categories* (Kubitzki, K., ed.) p. 5. Springer, Wien.
19. Ramshaw, J. A. M., Richardson, D. L., Meatyard, B. T., Brown, R. H., Richardson, M., Thompson, E. W. and Boulter, D. (1972) *New Phytol.* **71**, 773.
20. Walker, J. W. (1976) in *Origin and Early Evolution of Angiosperms* (Beck, C. B., ed.) p. 241. Columbia Univ. Press, New York.
21. Stebbins, G. L. (1975) *Ann. Mo. Bot. Gard.* **62**, 825.
22. Stevens, F. C., Glazer, A. N. and Smith, E. L. (1967) *J. Biol. Chem.* **242**, 2764.
23. Haslett, B. G., Evans, I. M. and Boulter, D. (1978) *Phytochemistry* **17**, 735.
24. Boulter, D. (1973) *Syst. Zool.* **22**, 549.
25. Constance, L. (1964) *Taxon* **13**, 257.
26. Haslett, B. G. and Boulter, D. (1976) *Biochem J.* **153**, 33.

APPENDIX

Partial Amino Acid Sequences of Plastocyanin from Members of Ten Families of the Flowering Plants: Data Set

B. G. HASLETT, D. BOULTER, J. A. M. RAMSHAW* and M. D. SCAWEN†

Department of Botany, University of Durham, Durham,
DH1 3LE, U.K.

The above paper presents the relationships suggested by the plastocyanin data set using a compatibility method. This appendix presents the sequence data themselves (Table 3).

The sequences are based on the results of a single automatic sequential degradation in each case using a Beckman Spinning Cup Sequencer. The methods and criteria for identification are given in [26].

---

* Department of Comparative Zoology, Harvard University, Cambridge, Mass., U.S.A.

† Microbiological Research Establishment, Porton, Salisbury, Wiltshire, U.K.