The proteins of the primary cell walls of suspension cultured cells of five plant species, Arabidopsis, carrot, French bean, tomato, and tobacco, have been compared. The approach that has been adopted is differential extraction followed by SDS-polyacrylamide gel electrophoresis (PAGE), rather than two-dimensional gel analysis, to facilitate protein sequencing. Whole cells were washed sequentially with the following aqueous solutions, CaCl₂, CDTA (cyclohexane diaminotetraacetic acid), DTT (dithiothreitol), NaCl, and borate. SDS-PAGE analysis showed consistent differences between species. From the 233 proteins that were selected for sequencing, 63% gave N-terminal data. This analysis shows that (i) patterns of proteins revealed by SDS-PAGE are strikingly different for all five species, (ii) a large number of these proteins cannot be identified by data base searches indicating that a significant proportion of wall proteins have not been previously described, (iii) the major proteins that can be identified belong to very different classes of proteins, (iv) the majority of proteins found in the extracellular growth media are absent from their respective cell wall extracts, and (v) the results of the extraction process are indicative of higher order structure. It appears that aspects of speciation reside in the complement of extracellular wall proteins. The data represent a protein resource for cell wall studies complementary to EST (expressed sequence tag) and DNA sequencing strategies.

The plant cell wall is a dynamic system generally considered to be composed of more than 90% carbohydrate polymers. Proteins, phenolics and possibly lipids make up the remainder of the wall (1–3). To date, most research interest has been in the carbohydrate components because of considerations of their structural role and commercial interest. This has led to a number of models for the integration, interpolymeric association, and assembly of the wall (3, 4). By comparison, our knowledge of the complexity of protein in the plant cell wall is in a less advanced state. Much of the understanding of the range of structural wall proteins has come from cDNA and genomic cloning exercises and has led to the identification of glycin-, cysteine-, proline-, and hydroxyproline-rich subsets of wall proteins. In addition, many extracellular enzymes have been identified that are required for the restructuring and modification of this dynamic extracellular matrix which underpin its role in defense, detoxification, signaling, cell-cell recognition, cell expansion, cell adhesion, cell separation, translocation, differentiation, and morphogenesis (2, 5, 6). However, there is a lack of direct studies on the proteins themselves and the true range of extracellular proteins and their species differences remains to be elucidated. The present work describes the systematic extraction and sequencing of the major primary wall proteins from five species representing four families of plants.

Since whole plant tissue is complicated by the presence of different tissue types including lignified secondary walls, we have chosen to use suspension cultures as a source of experimental material due to their relative uniformity. There are a large number of studies which show that molecular probes for proteins and enzymes derived from tissue cultured cells locate in a predictable way in the intact plant, and thus tissue cultures can be a reliable guide to a substantial number of phenomena in the intact plant. To cover a diverse range of species, we have used suspension cultures of Arabidopsis, carrot, French bean, tomato, and tobacco. The reasons for choosing these species relates to their academic and commercial interest. Arabidopsis has not been as extensively used in tissue culture but is currently the target of an extensive EST sequencing exercise, and an international program aimed at the sequencing of its entire genome is under way (7, 8). Carrot, a member of the Umbelliferae, has been an important species in modeling embryogenesis and elongation growth (9). Tissue cultures of the leguminous species, French bean, have been extensively used as a model system for cell wall biosynthesis and modifications during responses to pathogens (2). The two solanaceous species tobacco and tomato allow a comparison of the extent of conservation of wall proteins within one family. Tomato is important commercially and was the subject of the first successful attempts to modify the expression of wall proteins by transformation (10). Tobacco is an important model plant for attempts to modify cellulose extractability by modifying lignification (11).

The aim is to generate a protein resource for the plant cell wall analogous to current efforts of EST and genomic sequencing, so allowing for the future identification of potential biochemical function arising from these exercises. Homology searches of the derived amino acid sequence data from the present study should provide a firm indication of the number of components of the cell wall to which function is still to be ascribed.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Supported by a BBSRC ROPA award.

‡ Supported in part by a grant from the BBSRC.

§ To whom correspondence should be addressed. Fax: 44-191-374-2417; E-mail: a.r.slabas@durham.ac.uk.

1 The abbreviations used are: EST, expressed sequence tag; CDTA, cyclohexane diaminotetraacetic acid; DTT, dithiothreitol; PAGE, polyacrylamide gel electrophoresis.
MATERIALS AND METHODS

Plant Material—The derivation and maintenance of cultures of Arabidopsis (12), tobacco (13), tomato (14), and French bean (15) have been described previously. The carrot cultures (previously unpublished) were grown in Murashige and Skoog (16) basal salts supplemented with 3% sucrose and 2 mg/ml 2,4-D. Suspension cultures were grown in 100-ml batches in 250-ml Erlenmeyer flasks under a 16-h photoperiod (Arabidopsis, tomato, and carrot) or in the dark (French bean and tobacco) at 24 °C while rotary shaken at 130 rpm and subcultured every 7–10 days.

Extraction and Preparation of Cell Wall Proteins—Cells growing 4–5 days after subculture were harvested by filtration on Miracloth. The cells were washed three times with dH2O (3 ml/g fr. weight). All subsequent manipulations were conducted at 4 °C. The cells were stirred for 30 min with three volumes of 0.2 M CaCl2, followed by two volumes of 50 mM DTT, 1 M NaCl, and finally 0.2 M borate, pH 7.5. The borate extraction was conducted at room temperature. Between extractions, cells were washed on the filter three times, each with three volumes of dH2O.

Extracts and the culture media were refiltered through GF/A paper and washed on the filter three times, each with three volumes of dH2O. The cell wall was removed from the filter and resuspended for further extraction. The cell wall was lyophilized before being dialyzed against a 10-fold excess of dH2O with three changes. The samples were lyophilized and reconstituted in SDS-PAGE loading buffer (17) prior to analysis. The lyophilized culture filtrate was extensively dialyzed against dH2O before being dialyzed against a 10-fold excess of dH2O with three changes. The samples were lyophilized and reconstituted in SDS-PAGE loading buffer (17) prior to analysis. The lyophilized culture filtrates were reconstituted at 30 μg/μl, whereas all the other extracts were reconstituted at 15 μg/μl in gel loading buffer.

Protein Analysis—SDS-PAGE was carried out on 10% gels, using Bio-Rad Mini-PROTEAN II apparatus and stained with Coomassie Brilliant Blue (17). Since protein recoveries varied between the different extractions within each species and also between species the amount of extract loaded per lane was optimized by SDS-PAGE, typically between 10 and 20 μl of reconstituted extract was used per lane. For sequencing purposes, this process was also necessary due to the complexity and uneven abundance of the components within each extract, and more than one loading was used to maximize the acquisition of sequence data for the minor protein bands. For N-terminal amino acid sequence determination, proteins were transferred onto Problot membrane (Applied Biosystems, Inc. (ABI), Foster City, CA) and visualized by Coomassie Blue staining as directed in the Problot manual. Amino acid sequence analysis was performed in an ABI model 477 Sequencer. Sequence similarities were determined using the electronic mail “BLAST” series of programs (18) against the non-redundant protein, non-redundant DNA, and non-redundant EST databases maintained by the National Center for Biotechnology Information.

RESULTS

Analytical Strategy

Although there have been a number of systematic protein sequencing projects in plants, usually based on two-dimensional isoelectric focusing/SDS-PAGE systems of whole plant extracts (19–22), these are limited in the amount of information that was obtained due to mass limitations of the resolved material. Whereas the resolving power of two-dimensional gels is much higher than that of SDS-PAGE gels, the loading of each band per gel is higher for the latter generating sufficient mass for routine N-terminal analysis; the main problem is resolution. Accordingly we have chosen to reduce the complexity of the initial mixture of proteins by differentially extracting the cell wall from intact suspension cultured cells successively with CaCl2, CDTA, DTT, NaCl, and borate. The rationale for the initial wash was based on the success of CaCl2 to extract wall proteins (29). CDTA would be expected to remove any proteins associated with the pectin fraction since calcium would promote such binding. DTT was chosen to remove any protein-protein interactions based upon cysteine disulfide bonds. NaCl was used to extract any strongly ionically bound proteins. Finally, borate was used to disrupt any interactions due to glycoprotein side chains and other saccharides in the wall. At all stages, intervening water washes were used. For all species, secreted proteins were also characterized in the culture filtrate. Microscopic examination of the cells after these extractions showed the cells to be plasmolyzed, demonstrating that the plasma membrane of the cells remained intact, indicating that none of the extracted components were cytosolic.

Extraction and Patterns of Cell Wall Proteins

SDS-PAGE analysis reveals that the subsets of wall proteins obtained by successively washing whole cells are distinct for each reagent employed and for each species. Sequential extracts are shown for Arabidopsis (Fig. 1A), carrot (Fig. 1B), French bean (Fig. 1C), tomato (Fig. 1D), and tobacco (Fig. 1E). It can be seen for each species that the proteins found in the culture filtrate exhibit a strikingly different profile from that of their respective wall fractions. Similarly, there are distinct subsets of proteins extracted by CaCl2, CDTA, DTT, NaCl, and borate (Fig. 1, A–E). Individual extractions, hereafter referred...
FIG. 2. SDS-PAGE analysis of the cell wall proteins extracted non-sequentially from intact suspension cultured cells of Arabidopsis (A) and tomato (B). The proteins were extracted by stirring whole cells in the requisite solution for 30 min. The extractants were 0.2 M CaCl₂, 50 mM CDTA, 2 mM DTT, 1 M NaCl, and 0.2 M sodium borate, pH 7.5. Contamination from the culture filtrate proteins was prevented by rigorously washing the cells with dH₂O beforehand. The extracts were extensively dialyzed against dH₂O and lyophilized before SDS-PAGE was carried out on 10% (w/v) discontinuous gels. Visualization was by Coomassie. Lane 1, molecular weight markers as indicated; lane 2, contains the proteins extracted with CaCl₂; lane 3, CDTA-extracted proteins; lane 4, DTT-extracted proteins; lane 5, NaCl-extracted proteins; lane 6, borate-extracted proteins; lane 7, molecular weight markers as indicated. Bands that are labeled are those that were subjected to protein sequencing.

A summary of the proteins from each species for which N-terminal sequencing was attempted is shown in Table I. The N-terminal sequences that were obtained, along with the amino acid yield in the first sequencing cycle and the corresponding protein molecular weights, are listed in Table II.

Arabidopsis Cell Wall Proteins—From the 86 proteins selected for sequencing from the Arabidopsis extracts (Fig. 1A), 47 sequences were obtained, two of which yielded double sequences: band F in the CaCl₂ extract and band F in the non-sequential NaCl extract. In the case of band F from the CaCl₂ extract, it was possible to decipher the double sequence into separate sequences. This was performed by a process of subtraction since both the sequences within the CaCl₂ band F also appear as single sequences at other molecular weights within the same CaCl₂ extract, i.e. CaCl₂ band E and CaCl₂ band O. Discounting sequences that were present at either more than one molecular weight or were within other fractions left 31 unique sequences, which were found within the Arabidopsis wall extracts. The presumptive glycoprotein sequence starting “KVPV” is present in several extracts at more than one molecular weight (culture filtrate, band E; CaCl₂, bands E and F; non-sequential CDTA bands A and B), and may represent differential glycosylation. Evidence for microheterogeneity in primary sequence can be seen in bands F and G of the Arabidopsis culture filtrate proteins, which contained identical sequences except for the tyrosine or proline at position two. Bands K and L of the culture filtrate were found to contain identical sequences beginning with “ASSS.” The sequence beginning “NPNY” was found in the CaCl₂ extract, bands F, D, and O; the sequential CDTA extract, band D; and the non-sequential NaCl extract, band E. The sequence beginning with “IPCR” was found in the CaCl₂ extract, band R and the non-sequential CDTA extract, band K. The sequence beginning “RIPG” was found within the sequential DTT and NaCl extracts, bands B and H, respectively, and also within bands F and G of the non-sequential DTT extract. Bands I and J from the non-sequential CDTA extract were also found to contain the same sequence beginning with “KDLX.”

Carrot Cell Wall Proteins—All of the 10 carrot sequences were unique to carrot and were different from each other.

French Bean Cell Wall Proteins—From the 26 French bean proteins that were processed for sequencing (Fig. 1C), 21 sequences were obtained, of which six shared the same N-terminal four amino acids: “NYDK.” These sequences (culture filtrate, bands A and B; CaCl₂, bands D, O, and L) were identical apart from the CaCl₂ extract’s band O, where the sequence diverged substantially after the fourth amino acid. All of the other bean sequences were only seen in this species and were not represented at multiple molecular weights. Not counting sequences that appeared more than once left a total of 17 unique sequences in the bean extracts.

Tomato Cell Wall Proteins—Seventy-eight proteins were selected for sequencing from the tomato extracts (Fig. 1D), of which 46 yielded sequence information. Four of these proteins contained double sequences: culture filtrate band G, CaCl₂ bands D and O, and the non-sequential NaCl band L, all of which were deciphered into two individual sequences. This was possible by subtraction, in a similar way to deciphering the Arabidopsis double sequences. From the 46 sequences obtained, 30 were unique. The shortfall is due to a total of nine sequences, which were found at more than one position, or extract. Bands containing the sequence beginning “AKDF” were seen in the culture filtrate as bands G, H, and Ia (band I could be resolved into four more components by increasing the polyacrylamide concentration from 10 to 14% in the electrophoresis separating gel). The sequence beginning “KSTD” was also seen in band Ib of the culture filtrate and as the second component of band G of the same extract except that the N-terminal residue Lys was absent in the latter case. The sequence beginning “EQXG” was seen a total of eight times: culture filtrate, band E; CaCl₂, bands J, K, and L; the sequential CDTA extract, bands E and F; the sequential NaCl extract, band E; and the non-sequential DTT extract, band E. However, the CaCl₂ band J sequence was only the same up to the sixth residue, after which it diverged compared with culture filtrate band E. The sequence beginning “ANAK” appeared twice: once in the non-sequential NaCl extract, band G, and as one of the two short sequences that belong to CaCl₂ band D. Bands containing the sequence beginning with “VAGK” were CaCl₂ bands G and H; non-sequential NaCl extract, bands I and J; and band B of the non-sequential borate extract. These sequences were identical except residue 11, which was Ala in band G of the CaCl₂ extract and Leu in band J of the non-sequential NaCl extract. Sequences beginning with either “SNPN” or as “TNPN” appeared four times: twice beginning with TNPN, band Q in the CaCl₂ extract and as band G in the sequential CDTA extract, and twice beginning with SNPN as band F in the sequential NaCl extract and as band H in the non-sequential borate extract. Apart from this heterogeneity at the N terminus, all four sequences were identical. The sequence beginning “ALVE” appeared as band O in the non-sequential NaCl extract and as one of the two short sequences that were identified from band O of the CaCl₂ extract. Inciden-
the second of the short sequences from band O of the tobacco CaCl₂ extract (beginning “ADRE”) was also noted within the Arabidopsis CaCl₂ extract, band T. Bands containing the sequence beginning with “EVLY” were seen in the CaCl₂ extract, band F and in the non-sequential NaCl extract, band H. Finally, the sequence beginning with “ELQL” was seen in the non-sequential NaCl extract within bands L, M, and the non-sequential borate band D.

**Tobacco Cell Wall Proteins—**Twenty-seven of the protein bands from the tobacco extracts were selected for sequencing (Fig. 1E). Only two of the sequences were found to be either in more than one of the tobacco extracts or at more than one molecular weight within the same extract. The first begins with “GEQP” and was in the CaCl₂ extract as band C and band I. The sequence from band C was 18 amino acids long and that of band I, 17 amino acids long; both were identical, except in the case of band C’s sequence, which had an extra Asn present at amino acid position 13. The second tobacco sequence to appear in two different extracts began with “EQCQ” and was found within the CaCl₂ extract, band M and the CDTA extract, band E. All the other tobacco sequences are listed as they appear in Table II. Discounting any sequences that appeared more than once in these extracts left a total of 20 unique tobacco sequences.

**Overall Series of Sequencing Cell Wall Proteins—**On average 63% of the proteins from all of the plant species proved to be sequenceable, but these were only from the bands selected for analysis that were generated in this study. Generally the culture filtrate proteins yielded a high success rate in terms of the numbers of proteins that were sequenced: 92%, 11 out of 12, for Arabidopsis; 67%, 8 out of 12, for tomato; 100% for bean and tobacco: two and three proteins, respectively. In contrast, the sequential borate proteins were selected for sequencing, considerably less proteincs were amenable to this type of analysis. It was because of this continuing decline in successfully obtaining sequences that paralleled each step in the sequential extraction that none of the sequential borate proteins were selected for sequencing.

**Homology Searches**

The identities given for each sequence similarity listed in Table II can be defined as the percentage of amino acid matches for that query sequence against a sequence found in a particular data base. It should be noted that this figure does not take into account any conservative substitution. Throughout this study there have been numerous examples of a single sequence that was either present at more than one molecular weight within the same extract, or was present in more than one extract. It was therefore general practice to terminate sequence runs after the first four or five sequencing reaction cycles if this was observed. However, it is clear that microheterogeneity is occasionally seen between almost identical sequences, especially when longer sequences were obtained.

In the case of Arabidopsis, the classic model plant, which is a Brassica (an order that also includes plants such as oil seed rape and cauliflower), the proteins that could be identified were a subtilisin-like protease, β-1,3-glucanase, two xyloglucan endotransglycosidases, an extracellular ribonuclease, cellulase, a carrot-like glycoprotein, an Arabidopsis hypothetical protein, and an expansin. Those that showed lower sequence similarities, with less than 65% identities, were a cytochrome P-450, triose-phosphate isomerase, and a mouse T-cell receptor. Nineteen proteins did not show any similarity to any member of the data bases searched.

The member of the Umbelliferae, carrot, was included because it has been used to mimic elongation growth and embryogenesis. Only one of the 10 sequences obtained from the carrot wall fractions bore any similarity to data base sequences, unlike as it may seem, which corresponded to an unknown genomic DNA sequence from human.

The legume, French bean, has been used to model both differentiation and elicitor-induced pathogen stress (2). Identified wall proteins include extensin and a hybrid proline-rich, cysteine-rich chitin-binding protein (2). The bean culture was the only one in this study that led to the identification of hydroxyproline-rich glycoproteins. There were also two proteins that had a degree of similarity to two bacterial proteins, one of which was an iron hydrogenase. Thirteen of the French bean proteins were unlike any members present within the data bases searched.

For the two model solanaceous species, tobacco and tomato, there were several proteins, common to both species, that could be identified in similarity searches and based on these results may therefore have similar functions. For example, both species appeared to have forms of osmotin, chitinase, and pathogen-related proteins: tobacco P7 curled leaf protein in the case of tomato extracts and tobacco protein P10 in the case of the tobacco extracts. The tomato extracts were also found to contain proteins that bore similarity to other known extracellular proteins, which were an extracellular ribonuclease, β-1,3-glucanase, subtilisin-like protease, β-exoglucanase, and a peroxidase. One of the other tomato proteins that was found to have an analogue within the data bases was a human IgG heavy chain fragment. Altogether, 18 of the tomato proteins could not be identified by data base searches. Whereas the tomato fractions only contained a single form of chitinase, similar to one from French bean, the tobacco extracts were found to have two forms, one resembling a French bean chitinase and the other a
tobacco chitinase. The tobacco extracts also contained a protein that bore similarity to an integral stone protein associated with the human urinary tract. Altogether, 15 of the tobacco proteins could not be identified by data base searches.

**DISCUSSION**

The proteins of the plant extracellular matrix comprise a subset that contains both structural proteins and enzymes. Many of these have been identified at the protein level, but a large number have also been characterized from gene sequences on the basis of a repeat motif or a secretory leader sequence. In comparison to present knowledge of the proteins of some metabolic pathways such as the Calvin cycle, shikimic acid, and lignin biosynthetic pathways, the components of which have been completely cloned, the proteins and their cognate genes of the extracellular matrix are poorly characterized. This is probably due in part to the relative inaccessibility of wall proteins in the plant. Additionally, differentiation of the

| Band | Accession no. | Mr (pmol) | Initial level | Sequence | Sequence similarity |
|------|---------------|-----------|---------------|----------|---------------------|
| Culture | B | P80854 | 70 28 | TTRTFLFLDEHTADLXFE | Arabidopsis subtilisin like protease (100%) (100%)a |
| Filtrate | C | P80826 | 65 30 | EDRTY | |
| D | P80827 | 60 5 | KGVNDCGT | |
| E | P80828 | 54 13 | KVPVDDQFRVNNNGATDTR | Carrot glycoprotein (75%) (75%)a |
| F | P80829 | 52 12 | EPFGVNYGGQADNL | Wheat β-1,3-glucanase (69%) (69%)a |
| G | P80830 | 42 10 | EVFGYN | Wheat β-1,3-glucanase (57%) |
| H | P80831 | 34 12 | EQDGN | |
| I | P80832 | 31 20 | IALTV | |
| J | P80833 | 30 15 | NFQRFTEITWGDMRR | Arabidopsis xyloglucan endotransglycosidase (87%) (87%)a |
| K | P42813 | 25 43 | ASSSEDFDFYFVQGXP | Arabidopsis extracellular ribonuclease (84%) (84%)a |
| L | P42813 | 23 16 | ASSSEDFDFY | Arabidopsis extracellular ribonuclease (100%) |
| CaC2 | A | P80833 | 96 12 | AVREYHWVFVE | Arabidopsis cellulase (100%) |
| D | P80855 | 60 30 | NPNYKEALSK | Arabidopsis cellulase (100%) |
| E | P80828 | 59 30 | KVPVDDQFR | Carrot glycoprotein (100%) |
| F | P80855 | 54 40 | KVPYD | Carrot glycoprotein (100%) |
| G | P80834 | 52 20 | ATLVFRRDN | Arabidopsis extracellular ribonuclease (100%) |
| H | P80835 | 44 5 | HLKYKDFEQG | |
| K | P80836 | 36 30 | ADRELHRSKA | |
| O | P80855 | 30 39 | NPNYKEALSK | |
| Q | P80837 | 25 100 | RPIGYSGQAGVNAHTFYG | Arabidopsis expansin (75%) (90%)a,b |
| R | P80838 | 23 57 | IPCRAIDVPFGXRYVXW | Arabidopsis xyloglucan endotransglycosidase (65%) (85%)a,b |
| S | P80839 | 18 13 | ADLRR | |
| T | P80840 | 17 30 | ADREPHNHYA | |
| A | P80841 | 44 10 | EATVMDPhil | |
| D | P80855 | 36 8 | NPNYK | |
| DTT | B | P80845 | 45 30 | ARKFF | Arabidopsis triose-phosphate isomerase (100%) |
| NaCl | H | P80845 | 22 24 | ARKFFVG | Arabidopsis triose-phosphate isomerase (100%) |
| CDTA | A | P80828 | 72 9 | KVPV | Carrot glycoprotein (100%) |
| B | P80828 | 69 8 | KVPVDDQFR | Carrot glycoprotein (100%) |
| I | P80842 | 39 16 | KDLXHRDDEKT | |
| J | P80842 | 36 16 | KDLXHRDDEKT | |
| K | P80838 | 33 18 | IPCRAIDVVVF | Arabidopsis xyloglucan endotransglycosidase (82%) |
| DTT | A | P80843 | 68 14 | AVPPRHYGTR | Arabidopsis triose-phosphate isomerase (62%) (69%)a,b |
| B | P80844 | 64 8 | MREPHNFPP | Arabidopsis triose-phosphate isomerase (100%) |
| F | P80845 | 36 70 | ARKFFGRRNPEL | |
| G | P80845 | 30 20 | ARKFFV | |
| NaCl | H | P42761 | 25 10 | VLTIYA | Mouse T cell receptor (55%) |
| A | P80846 | 62 5 | YKTTGKGYR | |
| B | P80847 | 60 10 | DVGKFK | |
| D | P80848 | 48 9 | DNPSSTTPLR | |
| E | P80855 | 46 25 | NPNYKEALSK | |
| F | P80855 | 38 43 | AFQSEG | |

Proteins were subjected to SDS/PAGE and blotted onto PVDF paper and stained with Coomassie Brilliant Blue. Protein bands, the annotation of which is represented in Figs. 1 and 2, were excised and subjected to sequencing. Only those sequences that yielded more than 4 residues are shown. Sequence similarities with the degree of identity, displayed in parentheses, are shown.
### Table II—continued

| Band | Accession no. | M<sub>r</sub> (pmol) | Initial level (pmol) | Sequence | Sequence similarity |
|------|--------------|---------------------|---------------------|---------|---------------------|
| G    | P80849       | 36                  | 10                  | EDLPETK | Maize cytochrome P450 (50%)<sup>a</sup> (75%)<sup>b</sup> |
| H    | P80850       | 33                  | 20                  | APQEFGNFGQLKLKYH |                      |
| I    | P80852       | 31                  | 12                  | SDREHLRSKAAAYFF | Arabidopsis hypothetical protein (67%) |
| J    | P80851       | 27                  | 10                  | VDTSLRFLTVNPPPTVV |                      |
| K    | P80837       | 25                  | 70                  | RIPGIYSGGWQNAHATFYG | Arabidopsis expansin (65%) |

Sequently extracted carrot proteins

| Calcium Carbonate (CaC<sub>2</sub>) | Culture | Accession no. | M<sub>r</sub> (pmol) | Initial level (pmol) | Sequence | Notes |
|-----------------------------------|---------|--------------|---------------------|---------------------|---------|-------|
| A                                 | P80750  | 68           | 6                   | EPFPYRLVDN         |         |       |
| B                                 | P80751  | 66           | 6                   | GPLNAQHQS          |         |       |
| C                                 | P80752  | 58           | 10                  | QLAELKTVI          |         |       |
| D                                 | P80753  | 46           | 7                   | DSLNLLSRVPNERSN    |         |       |
| E                                 | P80754  | 43           | 9                   | GVREDTYPDVXTA      |         |       |
| F                                 | P80755  | 32           | 8                   | AEPNDVNLTVYYWP    |         |       |
| G                                 | P80756  | 30           | 19                  | SEVGLVQFPRTF       |         |       |
| H                                 | P80757  | 24           | 8                   | AHSVATTPPARSKV     |         |       |
| CDTA                              | C       | P80758       | 56                  | 5                   | SQEDTPL  |       |
| E                                 | P80759  | 30           | 5                   | ATNPSSQ           |         |       |

Sequently extracted French bean proteins

| Culture | Accession no. | M<sub>r</sub> (pmol) | Initial level (pmol) | Sequence | Notes |
|---------|--------------|---------------------|---------------------|---------|-------|
| Culture A | P80760       | 41                  | 20                  | NYDKPOVEKPOYKPOEVPY | Proline-rich protein (contains 3 repeating blocks of 4 amino acids) |
| Filtrate B | P80761       | 36                  | 3                   | NYDKPOVEKP       | Proline-rich protein |
| CaCl<sub>2</sub> | A | P80762       | 230                 | 8                   | NMYLPOVOOOOOOVPTF | Extensin |
| Filter B | P80763       | 140                 | 48                  | NHYSYSSSSSSSSVEST | Proline-rich protein (see bean culture filtrate, A) |
| C | P80760       | 138                 | 10                  | NYDKPOVEKPOYK     | Proline-rich protein (see bean culture filtrate, A) |
| E | P80764       | 65                  | 7                   | EDPVRFNLG         |         |       |
| F | P80765       | 60                  | 8                   | EDAYKFTTW          |         |       |
| G | P80766       | 53                  | 10                  | VAGRSVVKIAEGYL     |         |       |
| H | P80767       | 46                  | 5                   | KPDEPavlV          |         |       |
| I | P80768       | 45                  | 7                   | NPKPPEALILVKXSQ    |         |       |
| J | P80769       | 44                  | 10                  | SHTKPDHLKLFEKLDDLLSVHNA |         |       |
| K | P80770       | 43                  | 24                  | YDKVKSIIILFGVNG    |         |       |
| L | P80761       | 42                  | 30                  | NYDKPOVEKPOYKPOEVPYK | Proline-rich protein (see bean culture filtrate, A) |
| M | P80771       | 35                  | 12                  | ELPVNFYALNLTDNINIGY |         |       |
| O | P80772       | 32                  | 10                  | NYDKNPYEYDTP     |         |       |
| P | P80773       | 26                  | 5                   | EYPVVFVKLFFGKG    |         |       |
| Q | P80774       | 22                  | 25                  | QNPQPDFANOpFNPQNA |         |       |
| CDTA | C | P80775       | 36                  | 100                 | DVNGGGHTLTPQPLYQTTVVL |         |       |
| D | P80776       | 33                  | 20                  | AVGDPAPPAVYVTNG    | Desulfovibrio desulfuricans periplasmic Fe hydrogenase (75%) |
| E | P80777       | 30                  | 10                  | MGQGAVEQGFLFYNVQ  | Pseudomonas fluorescens protein F (80%) |

Sequently extracted tomato proteins

| Culture | Accession no. | M<sub>r</sub> (pmol) | Initial level (pmol) | Sequence | Notes |
|---------|--------------|---------------------|---------------------|---------|-------|
| Culture A | P80798       | 65                  | 23                  | EKAIAGLAKPRMDST | French bean chitinase (73%) |
| D | P80799       | 35                  | 20                  | EQPDEFDFDIT      | Tomato extracellular ribonuclease (100%) |
| E | P80800       | 31                  | 200                 | EQGXSAGAGLARGL   | Tomato extracellular ribonuclease (69%)<sup>b</sup> (65%)<sup>a</sup> |
| G | P80802       | 25                  | 8                   | STDFFDFN        | Tomato extracellular ribonuclease (100%) |
| F | P80801       | 14                  | 1                  | AKDDDFYFFQGQP    |         |       |
| H | P80801       | 24                  | 173                 | AKDFDDFYFVQQWPGYDTPKQPKQ |         |       |
| In<sup>a</sup> | P80801     | 23                  | 112                 | AKDFD         |         |       |
| Ib | P80802       | 22                  | 30                  | KSTDFDYNKKANYD   |         |       |
| Ic | P80803       | 21                  | 15                  | SNAVALNXXEXM    |         |       |
| D | P80817       | 64                  | 10                  | ANAVKPSP        |         |       |
| E | P80804       | 62                  | 6                   | EVPLDDTGL       |         |       |

<sup>a</sup> Maize cytochrome P450 (50%)<sup>b</sup> (75%)<sup>b</sup>

<sup>a</sup> Arabidopsis hypothetical protein (67%)

<sup>a</sup> Arabidopsis expansin (65%)
cells, which in part resides in profound changes in wall structure, and the extent to which individual proteins are immobilized within these structures increases complexity. One approach to increase knowledge of the range of wall proteins is to use tissue-cultured cells, which can be grown in bulk to allow characterization of proteins of walls that resemble primary

| Band | Accession no. | M (pmol) | Initial level | Sequence | Sequence similarity |
|------|--------------|----------|---------------|----------|---------------------|
| F    | P80805       | 50       | 8             | EVLYIPVTTDA | Human Ig heavy chain DJ region (64%) |
| G    | P80806       | 44       | 18            | VAGKSFPVIALGRQ | Tobacco P7 curled leaf protein (64%) |
| H    | P80825       | 40       | 13            | VAK     | Tobacco P7 curled leaf protein (100%) |
| I    | P80807       | 39       | 5             | SPV     | |
| J    | P80808       | 35       | 8             | EQXGR    | French bean chitinase (70%) |
| K    | P80809       | 34       | 9             | EQXGSA   | French bean chitinase (86%) |
| L    | P80810       | 33       | 7             | EQXGSA   | French bean chitinase (98%) |
| O    | P80840       | 28       | 15            | ADREP   | |
| P    | P80809       | 27       | 18            | TGVINPSLGNNLP | Tobacco β-1,3-glucanase (77%) (85%)\(^a\) |
| Q    | Q41350       | 23       | 100           | TNPFFI    | Tomato osmotin (100%) |
| E    | P80808       | 34       | 10            | EQXGS    | French bean chitinase (100%) |
| F    | P80831       | 31       | 8             | EQXGS    | French bean chitinase (100%) |
| G    | Q41350       | 23       | 8             | TNPFF    | Tobacco osmotin (100%) |
| NaC  | E P80808     | 34       | 38            | EQXGSAQAGGA | French bean chitinase (90%) |
| F    | F P80821     | 30       | 10            | SNFNFLTI  | Tomato osmotin (90%) |
| G    | G P80810     | 23       | 8             | ANPEVRNNLP | |
|     |              |          |               |          |                     |

Non-sequentially extracted tobacco proteins

| Band | Accession no. | M (pmol) | Initial level | Sequence |
|------|--------------|----------|---------------|----------|
| DTT  | A P80811     | 62       | 3             | MEKGYDLES |
|     | C P80812     | 40       | 18            | ANDDPFYTQANRP |
|     | E P80808     | 35       | 20            | EQXGSA |
|     | F P80813     | 22       | 8             | MNIPPGD |
| NaC  | A P80814     | 80       | 30            | STHTSDFLKL |
|     | C P80815     | 76       | 20            | STRTFEFLGLDNQCGVWA |
|     | E P80816     | 66       | 21            | GYMKYKDPKQPLLGRXRDX |
|     | G P80817     | 64       | 65            | ANAKVPSTHISNPF |
|     | H P80805     | 49       | 22            | E VilY |
|     | I P80825     | 47       | 15            | VAK     |
|     | J P80825     | 44       | 60            | VAGKSFPVIALGRQSKQTPF |
|     | K P80818     | 40       | 60            | GPVEIYQLQADARK |
|     | L P80824     | 38       | 25            | VKIGTYELLKGDFSV |
|     | M P80823     | 38       | 38            | ELQNYNTTKSXXRAE |
|     | M P80823     | 35       | 12            | ELQNYNTTKSXXRAE |
|     | O P80819     | 23       | 12            | ALVEDPQMQKYK |
|     | B P80825     | 44       | 10            | VAGKSFVIALGRQSKQTPF |
|     | C P80820     | 20       | 4             | HPVEI |
|     | D P80823     | 40       | 8             | ELQNY |
|     | H P80821     | 23       | 14            | SNFNFLTNYMNVPYTTWPA |
|     | I P80822     | 20       | 5             | EYIFPHFENW |

Sequentially extracted tobacco proteins

| Band | Accession no. | M (pmol) | Initial level | Sequence |
|------|--------------|----------|---------------|----------|
| Culture | A P80778 | 70       | 6             | GLVPPADKY |
| Filtrate | B P80779 | 41       | 31            | AVNGGAPATLEQPI |
| CaCl₂ | C P80780 | 34       | 10            | AHVIEEEPSL |
|     | A P80781 | 67       | 11            | STIEVRRNSSPYSYD |
|     | B P80782 | 66       | 9             | VPWAVWNSXNXYS |
|     | C P80783 | 57       | 10            | GEQPPGDQARQRRNPXNN |
|     | D P80784 | 56       | 7             | QDPYVDFLK |
|     | E P80785 | 52       | 8             | AQPQPAQDFL |
|     | F P80786 | 50       | 5             | FYYALIHTLTVNTFYNNPSAP |
|     | G P80787 | 48       | 9             | QYVPDPDQIVVARIFLQLQLVQR |
|     | H P80788 | 47       | 40            | ATIECVNLFLYYYYYYSSKAWSWN |
|     | I P80783 | 46       | 8             | GEQPPGDQARQRRNPXNN |
|     | J P80789 | 45       | 8             | QDARFLLTHTYVG |
|     | K P80790 | 40       | 5             | QPEEYFFA |
|     | L P80791 | 34       | 38            | WXPXQIFSAVRGAXV |
|     | M P80792 | 33       | 7             | EPCQDMAAGG |
|     | N P80793 | 28       | 11            | IWVIGYKHSFLYFQ |
|     | O P80794 | 27       | 28            | GPYPRKXPDVFPTTN |
|     | P P80795 | 15       | 10            | TIEKLLNPYYVVA |
|     | Q P80796 | 14       | 12            | AVFVIHTNVT |
| CDTA | B P80797 | 14       | 12            | AVFVIHTNVT |
|     | E P80798 | 33       | 5             | EOQDMAGGAR |

\(^a\) Sequences which have homologies to *Arabidopsis* ESTs.

\(^b\) Sequences which have homologies to ESTs from species other than *Arabidopsis*.

\(^c\) This band could be separated into 4 more by increasing the strength of the resolving gel from 10 to 14% (w/v) acrylamide.

Table II—continued

---

*Plant Cell Wall Proteins* 15847
walls. These cultures can also be stimulated to mimic developmental processes such as elongation growth and embryogenesis in carrot cells or xylogenesis and secondary wall formation as in *Zinnia* or French bean (24). A wide range of cells have also been used to mimic aspects of pathogen stress using fungal elicitors. A distinct advantage of cell suspensions is that they can be subjected to washing regimes that elute non-covalently bound proteins without disrupting the cells. Our studies demonstrate, by comparing the plant species, that (i) patterns of proteins revealed by SDS-PAGE are strikingly different, (ii) a large number of these proteins that were sequenced cannot be identified by database searches, and (iii) the major proteins that can be identified belong to very different classes of proteins. It appears that aspects of speciation resides in the complement of extracellular and cell wall proteins. Although it is difficult to discriminate between an extracellular protein and a cell wall protein *in planta*, this is not the case with suspension-cultured cells and is exemplified since most of the protein sequences found within the culture filtrates were absent from the subsequent salt-eluted extracts. Moreover, the culture filtrate profiles for each individual species was different to any of their respective wall extracts (Fig. 1, A–E).

The use of tissue cultures to model structures and processes in the whole plant has been validated by a large number of studies. These show that purification of proteins and cloning of cDNAs that originate through this route give rise to antibodies and cDNAs that can be used on the whole plant and locate to the predicted tissues and cells. Examples include structural proteins such as glycine-rich proteins (25), hydroxyproline-rich glycoproteins (5), proline-rich proteins (26), and enzymes such as peroxidases (23), chitinases, and laccases (27). Analyses of the sequences generated in this study validate the methodology, since many are known wall proteins and the cells remain intact. Only two potential cytoplasmic proteins have been revealed, a triose-phosphate isomerase and a P-450, of which their degree of identity was relatively low. Indeed, the types of extracellular proteins we have come across and identified through homology searches encompass the carbohydrate-modifying enzymes such as xylanoglucan endotransglycosidases, cellulases, and glucanases; other examples of wall proteins are the expansin, peroxidase, protease, ribonuclease, chitinases, extenсин, and proline-rich protein.

There are a number of reasons why certain known wall proteins may be absent from this study. The most conspicuous example is perhaps members of the hydroxyproline-rich glycoprotein family, only two of which were detected in the bean extracts. Their absence could be because they are present as minor components of the wall and consequently not detected here, or that they are only expressed in response to a particular environmental or developmental response that is not present within the culture regimes (28). The proteins could be N-terminally blocked, or it may also be that they are not readily extracted by these reagents, or that the more heavily glycosylated types are not stained with the Coomassie dye. It would seem that many of the enzyme activities associated with the wall are only known through function so may not yet be present within the data bases and are still waiting to be characterized by purification and subsequent sequencing.

For certain sequences within each species, there also existed a certain degree of heterogeneity in terms of the occasional amino acid substitution and also their appearance at different molecular weights. The former may be explained by the proteins originating from different genes and the latter by post-translational mechanisms such as glycosylation.

Cloning of any of the protein sequences reported here will undoubtedly be accelerated by the ever increasing numbers of ESTs that are being characterized. One surprising outcome of this research is that the sequenced *Arabidopsis* proteins did not have a larger proportion of “hits” in the EST data base. This could be because the majority of ESTs do not represent full-length cDNAs. Greater exploitation of the EST data base would thus require extensive internal amino acid sequence data to be generated if this were the case. It is envisaged that, within the foreseeable future, most if not all of the genes in several plants will be available in public access data bases. However, it has also been estimated that the function of at least 70% of the *Arabidopsis* genes alone, at present, cannot be identified by sequence homology to genes or proteins that have already been designated a function (7, 8). Therefore, having prior knowledge relating to a gene, such as its expression pattern or the eventual location of the expressed protein, may help in elucidating function. This information should therefore complement new data from systematic DNA sequencing exercises in that it gives a direct localization of the protein identified in this way.

REFERENCES

1. Fry, S. C. (1995) Annu. Rev. Plant Physiol. Mol. Biol. 46, 497–520
2. Bolwell, G. P. (1993) *Int. Rev. Cytol.* 146, 261–324
3. Caprara, N. C., and Gibeaut, D. M. (1993) *Plant J.* 3, 1–3
4. Roberts, K. (1990) *Curr. Opin. Cell Biol.* 2, 920–928
5. Showalter, A. M. (1993) *Plant Cell* 5, 9–23
6. Kieliszewski, M. J., and Lampert, D. T. A. (1994) *Plant J.* 5, 157–172
7. Newman, T., de Bruijn, F. J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohr, J., Raikhel, J. S., Slee, T., and Thomashow, M. F. (1994) *Plant Physiol.* 106, 1241–1255
8. Cook, R., Raynal, M., Launie, M., Grillet, F., Delseny, M., Morris, P. C., Guerrier, D., Giraudat, J., Quigley, F., Claubault, G., Li, Y. F., Mache, B., Krivitsky, M. S., Miyake, S., Lecharny, A., Petzmann, Y., Marbach, J. P., Fleck, J., Clement, B., and Phillips, G. (1994) *Plant Physiol.* 106, 1261–1271
9. Kreuger, M., and van Hoot, G. J. (1993) *Plant Physiol.* 102, 243–248
10. Smith, C. J. S., Watson, C. F., Ray, J., Bird, C. R., Morris, P. C., Schuch, W., and Grierdon, D. (1988) *Nature* 334, 724–726
11. Elkin, V., Edwards, R., Madzalende, M., Hedrick, S. A., Ribak, O., Dixon, R. A., and Lamb, C. J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 9057–9061
12. May, M. J., and Leaver, C. J. (1993) *Plant Physiol.* 102, 621–627
13. Memelink, J., Hoge, J. H. C., and Schippers, R. A. (1987) *EMBO J.* 6, 3575–3583
14. Fry, S. C. (1988) The Growing Plant Cell Wall, Longman, Harlow, United Kingdom
15. Diam, R. A., and Lamb, C. J. (1979) *Biochim. Biophys. Acta* 586, 456–463
16. Murashige, T., and Skoog, F. (1962) *Physiol. Plant.* 15, 473–479
17. Laemmli, U. K. (1970) *Nature* 227, 680–695
18. Altshul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410
19. Kamo, M., Kawakami, T., Miyatake, N., and Tsugita, A. (1995) *Electrophoresis* 16, 425–450
20. Tsugita, A., Kawakami, T., Uchiyama, Y., Kamo, M., Miyatake, N., and Nana, Y. (1994) *Electrophoresis* 15, 708–720
21. Baw, G., De Loose, M., Inze, D., Van Montagu, M., and Vanderkerveke, J. (1986) *Proc. Natl. Acad. Sci. U. S. A.* 84, 4806–4810
22. Baw, G., Rasmussen, H., Van den Bulcke, M., Van Damme, J., Puppe, M., Geesaer, B., Celes, J. E., and Vanderkerveke, J. (1990) *Electrophoresis* 11, 528–536
23. Smith, J. J., Muldoon, E. P., and Lampert, D. T. A. (1984) *Phytochemistry* 23, 1233–1239
24. Bolwell, G. P., and Robertson, D. (1997) in *Morphogenesis in Plant Tissue Culture* (Soh, W. Y., ed) Kluwer Academic Publishers, Dordrecht, The Netherlands, in press
25. Keller, B. (1993) *Plant Physiol.* 101, 1127–1130
26. Datta, K., Schmidt, A., and Marcus, A. (1989) *Plant Cell* 1, 945–952
27. Faye, L., Johnston, K. D., Storm, A., and Chrispeels, M. J. (1989) *Plant Physiol.* 75, 309–314
28. Miller, J. G., and Fry, S. C. (1992) *Plant Cell Tissue Organ Cult.* 31, 61–66