

RESEARCH ARTICLE

Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colorless strawberry mutant

Karin Hjernø¹, Rikard Alm², Björn Canbäck³, Rune Matthiesen¹, Karin Trajkovski⁴, Lars Björk⁵, Peter Roepstorff¹ and Cecilia Emanuelsson²

¹Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

²Department of Biochemistry, Lund University, Sweden

³Department of Microbial Ecology, Lund University, Sweden

⁴Swedish Pomology Science Center, Sweden

⁵Department of Systematic Botany, Uppsala University, Sweden

Proteomic screening of strawberry (*Fragaria ananassa*) yielded a 58% success rate in protein identification in spite of the fact that no genomic sequence is available for this species. This was achieved by a combination of MALDI-MS/MS *de novo* sequencing of double-derivatized peptides and indel-tolerant searching against local protein databases built on both EST and full-length nucleotide sequences. The amino acid sequence of a strawberry allergen, homologous to the well-known major birch pollen allergen Bet v 1, was partially determined. This strawberry allergen, named Fra a 1 according to the nomenclature for allergen proteins, showed sequence identity of 54 and 77%, respectively, with corresponding allergens from birch and apple. Differential expression, as evaluated by 2-D DIGE, occurred in 10% of protein spots when red strawberries were compared to a colorless (white) strawberry mutant. White strawberries, known to be tolerated by individuals affected by allergy, were found to be virtually free from the strawberry allergen. Also several enzymes in the pathway for biosynthesis of flavonoids, to which the red color pelargonidin belongs, were down-regulated. This approach to assess differential protein expression without access to genomic sequence information can also be applied to other crop plants and phenotypic traits.

Received: June 24, 2005
Revised: August 10, 2005
Accepted: August 11, 2005

Keywords:

Allergens / *De novo* sequencing / Differential protein expression / Fluorescent two-dimensional differential gel electrophoresis / Protein identification

1 Introduction

Food allergy to vegetables, fruits, and berries is often caused by proteins homologous to birch pollen allergen proteins [1, 2]. A primary induction of antibodies, caused by inhalation of

birch pollen proteins, is followed by secondary allergic reactions due to antibody cross-reactivity to common epitopes in plant proteins [3]. Many plant-derived food allergens are homologs to the major allergen in birch pollen, the Bet v 1-protein [4]. Birch pollen related food allergy is a well-known epiphenomenon especially in northern Europe, and the structure of several allergens have been determined to atomic resolution, including a structure with a mAb bound to the Bet v 1 epitope [5]. Yet little is actually known about the biological function of Bet v 1 and its homologs.

For apple and cherry, variation in quantity of the Bet v 1-homologs Mal d 1 and Pru av 1 correlates with IgE binding [6, 7]. Different isoforms of allergens can be detected within a

Correspondence: Associate Professor Cecilia Sundby Emanuelsson, Department of Biochemistry, Center for Chemistry and Chemical Engineering, Lund University, PO Box 124, S-221 00 Lund, Sweden

E-mail: Cecilia.Emanuelsson@biokem.lu.se

Fax: +46-(0)-46-222-45-34

Abbreviation: SPITC, 4-sulfophenyl isothiocyanate

single fruit [7–9]. These phenomena may form the molecular basis for why birch pollen allergic patients react differently to different apple varieties. More than ten different isoforms of Bet v 1 present different IgE-binding capacity and T-cell reactivity [10].

In strawberry (*Fragaria ananassa*), we recently identified a protein homologous to the birch pollen allergen Bet v 1 by utilizing patient samples for cell-based assays and immunoblotting, in combination with protein analysis with LC-MS/MS and database searching using the MASCOT error tolerant search mode [11]. The immunoblot reaction could be inhibited by recombinant Bet v 1. Thus, the allergenic impact could be tied to a specific protein.

In this study we have separated extracted strawberry proteins by 2-D PAGE and analyzed protein spots by MS. As the sequence information from strawberry is very limited (only 290 proteins in NCBI from *F. ananassa*) we have applied several MS-based approaches to be able to identify proteins that are differentially expressed between colorless (white) and normal red strawberry varieties. The motivation for this comparison is that strawberry breeders have experienced that white strawberries are tolerated by individuals with adverse reactions to normal strawberry (Åke Truedsson, personal communication). Information on allergens and allergen isoforms is also crucial to identify varieties with low allergenic impact suitable for allergic individuals. Moreover, we hoped to gain evidence that this setup can be used to evaluate biological diversity, in terms of allergen content and of other properties, by analyzing proteome variation within and between varieties.

In the present study we have investigated proteins differentially expressed in red and white strawberries. We found that the allergen was down-regulated in white strawberries, as were also several enzymes in the biosynthetic pathway for synthesis of red color. Moreover, we have detected several allergen isoforms and deduced a representative amino acid sequence to be used for cloning and sequencing of the strawberry allergen.

2 Materials and methods

2.1 Strawberry proteins

Different varieties of strawberry, *F. ananassa*, cultivated in a breeding program directed by Karin Trajkovski at Balsgård, SLU, Sweden were selected for analysis. Eight varieties were used for the proteomic analysis, four ordinary red varieties (names Korona, Sengana, Honeoye, Cavendish) and four colorless (white) varieties (one variety named Chili Manzanar Alto, plus three novel varieties without registered names, referred to here with their breeding protocol numbers, 968603, 976902, 957401). Extracted proteins from one red and one white variety were used for a comparative proteome investigation (red variety: Korona and white variety: 968603; a crossing between two pink varieties, each derived from one

white and one red parent, one of which was Korona). Extracted protein from all above-mentioned eight varieties was pooled and used as reference in the DIGE analysis (see below). In addition, one other red variety, Elsanta, was submitted to detailed analysis of the strawberry allergen, to allow comparison to the nucleotide sequence data that were available from Elsanta. All varieties, except Elsanta, were grown outdoors on the same site and harvested on the same day and stored in -20°C freezer until protein extraction. Proteins were extracted by a phenol-extraction procedure in presence of protease inhibitors as described previously [11]. Shortly, strawberries were ground in liquid nitrogen into a frozen powder, proteins were extracted with phenol, and precipitated by 0.1 M ammonium acetate in methanol at -20°C with repeated washing. The yield of protein was approximately 200 μg per 200 mg frozen strawberry powder. Extracted protein was stored as dry pellets in -80°C freezer.

2.2 2-DE

2-D PAGE was performed at the Lund University Swegene Proteome Center (<http://www.swegene.se>). IEF was performed according to the protocols recommended for IPGphor (Amersham Biosciences, Sunnyvale, CA). Proteins, resolved in 8 M urea/2% CHAPS/IPG buffer 3–10 NL (5 μLmL)/DTT (2.8 mg/mL), were equilibrated with IPG strips (3–10 NL/24 cm) overnight at 20°C , and first dimension was run on IPGphor (prerunning conditions 1 h at 500 V, 2 h at 4000 V, and 1 h at 8000 V at 20°C , followed by 8000 V for 7 h). Second dimension was run according to [12] on 12.5% polyacrylamide gels on an Ettan DALT II system (Amersham Biosciences), 20°C , 10 W/gel. The amount of protein loaded to 2-D PAGE was either three times 50 μg labeled with the fluorescent cyanine dyes prior to 2-D PAGE (quantitative DIGE gels, see below), or 0.5 mg unlabeled protein (preparative gels, with either variety Korona or Elsanta), in which case spots were visualized by staining after 2-D PAGE with ruthenium II tris bathophenanthroline disulfonate [12] prior to spot picking for protein identification by MS.

2.3 Evaluation of differential expression on quantitative DIGE gels

Protein samples were labeled before 2-D DIGE [13], electrophoresis with spectrally resolvable fluorescent cyanine dyes Cy2TM, Cy3TM, and Cy5TM (Amersham Biosciences), allowing three samples at a time to be run under identical conditions in one gel.

In total, eight replicate gels were run; each of these eight gels contained three samples of 50 μg protein labeled with Cy2, Cy3, and Cy5. The gels were scanned using a Typhoon9400TM scanner (Amersham Biosciences) set at appropriate wavelengths for each of the cyanine dyes [13]. Spot detection and matching were performed using the software DeCyder (Amersham Biosciences), with the Cy5-labeled

pooled sample being the reference standard. Statistical analysis of protein abundance change between the red and the white strawberry varieties was performed in the DeCyder-BVA (Biological Variation Analysis module, v 5.02) with a *t*-test and average ratio calculated for each pair of spots as a measure of the differential protein expression. Of the eight replicate gels that were run, one of the gels had to be discarded due to a technical failure, hence data from seven replicates each of red and white variety were used in the evaluation.

2.4 Automated spot picking and protein identification from preparative gels

Preparative gels were matched to the DIGE data from the quantitative gels using DeCyder. Spots that were differentially expressed in red and white strawberries according to DIGE were selected on preparative gels for automated spot-picking (2 mm diameter gel plugs) and trypsin digestion (Porcine Sequencing Grade Modified Trypsin, Promega, Southampton, UK) with Ettan Spot Handling Workstation 2.1 (Amersham Biosciences), according to the protocols supplied by the manufacturer. After processing, the dry samples containing peptides were dissolved in 0.1% (w/v) TFA and finally spotted with the matrix α -hydroxycinnamic acid onto different MALDI target plates. MALDI-MS mass spectra were acquired in data-dependent mode on a Waters® Micromass® MALDI micro MX™ Mass Spectrometer (Waters, Manchester, UK) followed by automated protein identification by protein mass fingerprinting searching the NCBI database, limited to green plant (Viridiplantae), with either the search engine MASCOT (www.matrixscience.com [14]) or the software PIUMS® (www.hh.se/staff/bioinf [15]). PIUMS® incorporates features such as automated calibration and removal of contaminants.

2.5 Analysis by MALDI-MS/MS

Spot sets from another three preparative gels were subjected to the same automated spot-picking and digestion as described in Section 2.4, but instead of automated spotting on the MALDI-MS target plate, the dried peptides were redissolved in 0.1% TFA followed by desalting and concentration on microcolumns with POROS R2 resin (Applied Biosystems, Foster City, CA) packed in a GELoader tip (Eppendorf Scientific, Westbury, NY), as described previously [16]. Peptides were eluted from the microcolumns with 0.6 μ L matrix solution (5 mg/mL α -hydroxycinnamic acid in 70% (v/v) CH₃CN and 0.1% TFA) directly onto the MALDI target.

Manual data acquisition was performed using an Applied Biosystems 4700 Proteomics Analyzer with TOF/TOF™ optics (Applied Biosystems, Darmstadt, Germany), in positive reflector mode. Laser intensity and number of laser shots were varied depending on the peptide analyzed. In the manual setup, an MS spectrum was first recorded and analyzed manually by a first round of peptide mass search. Peaks were selected for MS/MS analysis based on the obtained results. MS/MS spectra were included in the analysis and the total scores reported.

For samples not clearly identified by the above-mentioned approaches, N-terminal sulfonation was performed using 4-sulphophenyl isothiocyanate (SPITC) as the derivatization reagent [17, 18]. The reagent solution was prepared by dissolving SPITC in 20 mM NaHCO₃ (pH 8.6) to a concentration of 10 μ g/ μ L. The vacuum dried peptides were redissolved in 20 mM NaHCO₃ (pH 8.6) and mixed with 8.5 μ L reagent solution and incubated at 50°C for 30 min followed by purification on microcolumns as described above.

In some cases, a solid-phase double-derivatization procedure was performed as described by Chen *et al.* [19] with some modifications: the tryptic peptide mixture was dissolved in 0.1% TFA and loaded onto microcolumns as described above. After washing the microcolumn with 0.1% TFA, lysine residues were transformed to homoarginine by application of 10 μ L of a solution of *O*-methylisourea hydrogen sulfate (0.5 M in 0.25 M Na₂CO₃, pH 9.4) onto the microcolumn for 10 min at 65°C, followed by rinsing with 0.1% TFA. N-terminal sulfonation was subsequently performed by applying excess volume of a 10 μ g/ μ L reagent solution of SPITC dissolved in 20 mM NaHCO₃ (pH 8.6) for 30 min at 50°C followed by rinsing with 0.1% TFA and direct elution onto the target as described above.

2.6 Protein identification and database searching

So called project-specific contaminants (peaks present in all mass spectra; mostly but not only originating from contaminating human keratin or the trypsin added for digestion of proteins) were identified and eliminated from the MS spectra using the software PeakEraser (<http://www.welcome.to/GPMAW> [20]). Furthermore, the mass accuracy of the MS spectra was improved by a recalibration procedure using PeakEraser [21]. Protein identification was performed either by searching the MALDI-MS spectra alone and/or in combination with MS/MS spectra from the same spot. Searches were performed against the database NCBIInr, limited to green plant (Viridiplantae). The following parameters were used for a standard database search: missed cleavages, 1; allowed modifications: carboxamidomethyl on cysteine, acrylamid adduct (propionamide) on cysteine, and oxidation of methionine, all as partial modifications. Significant hits were carefully inspected manually to eliminate false positives, based on criteria such as mass error distribution, likely/not-likely missed cleavages, and that detection of arginine-containing peptides are favored compared to lysine-containing peptides. For SPITC modified peptides, a user-defined modification was introduced in MASCOT adding 215 Da to the N-terminal of the peptides, and for double-derivatized peptides, a modification adding 42 Da to the lysine residues.

Data acquired by MALDI-MS/MS were subjected to peptide mass search by using MASCOT to search against all green plants in NCBIInr (Viridiplantae) followed by a search in an in-house constructed strawberry database (see below) using the search engine VEMSv3.0 (version III, <http://www.yass.sdu.dk/VEMSv3.0>) [22]. All MASCOT scores and *e*-values

reported for the MALDI-MS/MS search results in Table 1 were obtained using a fragment ion mass tolerance of 0.25 Da.

De novo sequencing was performed manually and assisted by the software VEMS and Data Explorer (v. 4.4, Applied Biosystems). Database searches were performed either using MSBlast (<http://dove.embl-heidelberg.de/Blast2/msblast.html> [23]) to search public databases or the Blast function in VEMS to search against the strawberry EST-database.

The software GPMW (<http://welcome.to/gpmw> [20]) was used for all sequence handling and storage. Multiple alignments and similarity calculations were performed using ClustalW (<http://www.ch.embnet.org/software/ClustalW.html> [24]) and SIM (<http://www.expasy.org/tools/simprot.html> [25]), respectively.

2.7 Construction of strawberry EST-database

A data set of 2169 EST sequences from strawberry (*F. ananassa*, cultivar Elsanta), kindly received from Elma Salentijn and Asap Aharoni at The Plant Research International, Wageningen, Netherlands, were added to a set of 5322 *F. ananassa* EST sequences collected from dbEST [26]. The software FrameD [27] was run to identify and correct possible frameshift errors using an organism specific Markov model build of 144 coding sequences from *F. ananassa* (downloaded from the coding sequence section of EMBL [28]). FrameD

was run together with blastx [29] output using an *e*-value threshold of $1e^{-10}$. The ESTs were assembled into contigs with the aid of the CAP3 software [30].

3 Results

3.1 The strawberry allergen is down-regulated in colorless, "white" variety

Approximately, 1000 protein spots with *pI* between 3 and 10 and molecular mass between 10 and 200 kDa could be resolved by 2-D PAGE (Fig. 1). Of these spots, 104 were differently expressed in red compared to white variety (difference in spot intensity >1.5 , with $p < 0.01$). It is evident from naked eye that the strawberry Bet v 1-homologous allergen is less abundant or absent in the white strawberry.

Quantification, as exemplified in Fig. 2 by the predominant allergen spot, spot no. 898, was performed using the software DeCyder. The 3-D presentation (Fig. 2B) shows that the spot 898 peak area density is much larger for red compared to white strawberry, in all seven replicates (Fig. 2C). The average ratio for spot 898 is -7.62 (Fig. 2D), indicating that the allergen is more than seven-fold more abundant in the red compared to white strawberry. The average ratios for all spots where proteins were identified are listed in Table 1.

Table 1. Proteins differentially expressed in red and white strawberries

Spot no.	Up/down-regulation average ratio ^{a)}	Accession no. ^{b)}	Protein name	Species matched ^{c)}	MASCOT score (MALDI-TOF/TOF data) ^{d)}	MASCOT <i>e</i> -values (MALDI-TOF/TOF data) ^{d)}	Additionally identified by
96	↓ -1.78	O24246	Phosphoglycerate mutase	<i>Prunus dulcis</i>	188	3.00E-14	^{e)}
118	↓ -1.84	P08927	Chaperonin precursor (only 1 peptide)	<i>Pisum sativum</i>	55	8.90E-04	
157	↓ -2.11	O64459	UDP-glucose pyrophosphorylase	<i>Pyrus pyrifolia</i>	358	3.00E-31	
161	↓ -1.38	Q42971	Enolase	<i>Oryza sativa</i> ^{f)}	177	3.80E-13	
212	↑ 1.37	P83372	Citrate synthase	<i>Fragaria ananassa</i>	84	7.90E-04	
213	↑ 1.28	P83372	Citrate synthase	<i>Fragaria ananassa</i>	–	–	^{g)}
250	↑ 1.76	Q9STA6	RAD23 protein	<i>Lycopersicon esculentum</i>	73	9.70E-03	^{g), h)}
251	↑ 1.75	Q9STA6	RAD23 protein	<i>Lycopersicon esculentum</i>	73	9.70E-03	^{g), h)}
261	↓ -1.76	Q84L31	RAD23-like protein	<i>Arabidopsis thaliana</i>	53	2.50E-02	^{g), h)}
274	↓ -1.69	Q40345	Isocitrate dehydrogenase (NADP ⁺)	<i>Apium graveolens</i>	330	1.90E-28	^{h)}
277	↓ -1.66	P17648	Alcohol dehydrogenase	<i>Fragaria ananassa</i>	87	4.10E-04	
281	↓ -1.48	P30168	Actin	<i>Solanum tuberosum</i>	52	4.70E-03	^{h)}
294	↓ -2.49	Q84L31	RAD23-like protein	<i>Arabidopsis thaliana</i>	111	1.50E-06	^{h)}
297	↓ -1.64	P83372	Citrate synthase	<i>Fragaria ananassa</i>	113	9.40E-07	

Table 1. Continued

Spot no.	Up/down-regulation average ratio ^{a)}	Accession no. ^{b)}	Protein name	Species matched ^{c)}	MASCOT score (MALDI-TOF/TOF data) ^{d)}	MASCOT e-values (MALDI-TOF/TOF data) ^{d)}	Additionally identified by
301	↓ -2.09	O22586	Chalcone synthase	<i>Onobrychis viciifolia</i> ^{f)}	181	1.50E-13	e), g)
324	↓ -1.77	O22586	Chalcone synthase	<i>Onobrychis viciifolia</i> ^{f)}	87	4.1E-04	h)
		P17648	Alcohol dehydrogenase	<i>Fragaria ananassa</i>	80	1.70E-03	h)
325	↓ -2.37	O22586	Chalcone synthase	<i>Onobrychis viciifolia</i> ^{f)}	160	1.90E-11	e), g)
334	↓ -1.94	Q43040	Chalcone synthase	<i>Petunia hybrida</i> ^{f)}	215	6.1E-17	e)
349	↓ -1.73	Q9FUG7	Cytosolic aldolase (only 1 peptide)	<i>Fragaria ananassa</i>	52	3.30E-03	g)
352	↓ -2.92	Q94B32	Similar to dihydroflavonol reductase	<i>Arabidopsis thaliana</i>	261	1.50E-21	e)
371	↓ -3.45	Q9ZRF1	Cinnamyl alcohol dehydrogenase	<i>Fragaria ananassa</i>	105	6.10E-06	h)
396	↓ -2.50	Q9M602	O-methyltransferase	<i>Fragaria ananassa</i>	–	–	g)
		Q66ME9	Flavanone 3-hydroxylase	<i>Fragaria ananassa</i>	95	5.60E-05	
397	↓ -3.13	Q9M602	O-methyltransferase	<i>Fragaria ananassa</i>	176	4.70E-13	
404	↓ -1.67	Q9M602	O-methyltransferase	<i>Fragaria ananassa</i>	205	6.00E-16	
409	↓ -2.37	Q9M602	O-methyltransferase	<i>Fragaria ananassa</i>	154	7.50E-11	
416	↑ 3.38	Q944T7	GA 3-P-DH	<i>Fragaria ananassa</i>	88	3.30E-04	
		Q84V25	Quinone oxidoreductase	<i>Fragaria ananassa</i>	86	5.30E-04	
418	↑ 2.58	Q84V25	Quinone oxidoreductase	<i>Fragaria ananassa</i>	277	3.80E-23	
433	↑ 2.24	Q941I0	Putative quinone oxidoreductase	<i>Fragaria ananassa</i>	367	3.80E-32	e)
455	↓ -1.38	Q944T7	GA 3-P-DH (only 1 peptide)	<i>Fragaria ananassa</i>	115	6.00E-07	
		Q9FUG7	Cytosolic aldolase	<i>Fragaria ananassa</i>	191	1.50E-14	e)
465	↓ -1.54	P93819	Putative malate dehydrogenase	<i>Arabidopsis thaliana</i>	99	2.50E-05	e)
478	↑ 2.04	Q08062	Cytoplasmic malate dehydrogenase	<i>Zea mays</i>	188	3.00E-14	
		Q84V25	Quinone oxidoreductase	<i>Fragaria ananassa</i>	129	2.40E-08	e)
487	↑ 1.80	P93819	Putative malate dehydrogenase	<i>Arabidopsis thaliana</i>	99	2.50E-05	
499	↓ -1.73	em:BM412365	EST586692Putative plastidic aldolase	<i>Lycopersicon esculentum</i>	165	1.10E-09	h)
		O65363	Malate dehydrogenase	<i>Glycine max</i>	110	1.90E-06	
509	↑ 2.54	Q43607	Prunin	<i>Prunus dulcis</i>	–	–	h)
663	↑ 1.36	O49997	14–3-3 isoform e	<i>Nicotiana tabacum</i> ^{f)}	220	1.90E-17	e)
664	↑ 1.48	O49997	14–3-3 isoform e	<i>Nicotiana tabacum</i> ^{f)}	249	2.40E-20	e)
708	↓ -1.95	Q9M2W6	Porin-like protein	<i>Arabidopsis thaliana</i>	161	1.50E-11	
715	↑ 1.30	rs: XM_483663	Proteasome alpha subunit (only 1 peptide)	<i>Oryza sativa</i>	97	4.00E-05	
745	↑ 1.37	O48551	Proteasome IOTA subunit	<i>Glycine max</i>	111	1.50E-06	
777	↑ 1.37	Q75K80	Putative nascent polypeptide ass complex chain	<i>Oryza sativa</i> ^{f)}	246	4.70E-20	
780	↑ 4.09	Q9S7F6	Cytosolic ascorbate peroxidase	<i>Fragaria ananassa</i>	–	–	e), g)
784	↑ 1.55	Q9S7F6	Cytosolic ascorbate peroxidase	<i>Fragaria ananassa</i>	272	1.20E-22	e)
825	↑ 1.67	Q39625	Expansin S1 precursor (only 1 peptide)	<i>Cucumis sativus</i>	67	3.50E-02	
834	↑ 2.25	Q84V25	Quinone oxidoreductase	<i>Fragaria ananassa</i>	133	1.70E-06	

Table 1. Continued

Spot no.	Up/down-regulation average ratio ^{a)}	Accession no. ^{b)}	Protein name	Species matched ^{c)}	MASCOT score (MALDI-TOF/TOF data) ^{d)}	MASCOT e-values (MALDI-TOF/TOF data) ^{d)}	Additionally identified by
880	↑ 2.13	P54152	Methionine sulfoxide reductase (only 1 peptide)	<i>Fragaria ananassa</i>	72	1.10E-02	
898	↓ -7.62	P43211	Mal d 1 protein (only 1 peptide) (The Bet v 1-homologous strawberry allergen, Fra a 1)	<i>Malus domestica</i> ^{f)} See Figs. 4, 6	74	7.30E-03	i)
899	↓ -3.66	Q8S3L0	Peroxiredoxin	<i>P. tremula</i> × <i>P. tremuloides</i>	74	6.10E-03	
903	↑ 4.56	Q8S3L0	Peroxiredoxin	<i>P. tremula</i> × <i>P. tremuloides</i>	54	7.70E-01	g)
905	↓ -7.42	Q8S3L0	Peroxiredoxin	<i>P. tremula</i> × <i>P. tremuloides</i>	136	4.70E-09	
919	↓ -7.14	No acc nr	The Bet v 1-homologous strawberry allergen, Fra a 1	See Figs. 4, 6	–	–	i)
920	↓ -2.86	No acc nr	The Bet v 1-homologous strawberry allergen, Fra a 1	See Figs. 4, 6	–	–	i)
963	↑ 2.15	Q24246	Phosphoglycerate mutase	<i>Prunus dulcis</i>	105	6.00E-06	
		Q93X24	Thioredoxin h (only 1 peptide)	<i>Pisum sativum</i> ^{f)}	90	1.70E-04	
1026	↓ -1.54	Q41119	Cyclophilin	<i>Phaseolus vulgaris</i>	175	6.1E-13	
1051	↓ -2.56	No acc nr	The Bet v 1-homologous strawberry allergen, Fra a 1	See Figs. 4, 6	–	–	i)
1112	↓ -1.23	Q946X8	Citrate synthase	<i>Prunus persica</i> ^{f)}	210	1.90E-16	

a) Arrows pointing downwards (↓) and with accession number and name marked gray-shaded, represent 32 spots that were down-regulated in white compared to red strawberries. Arrows pointing up-wards (↑) and with no shading of accession number and name represent 21 spots that were up-regulated in white compared to red strawberries. Average ratios for the different expression determined in seven replicates are given as negative and positive values, respectively, for down- and up-regulation.

b) Accession numbers are given according to Swiss-Prot and TrEMBL except for two entries from EMBL (em) or RefSeq (rs).

c) Protein reported is the protein given the highest score except when a protein sequence from *Fragaria ananassa* appeared further down in the list with a significant score, then the strawberry sequence is instead reported. Multiple protein entries representing homologs from different species were often found with equivalent scores, these are not reported.

d) MASCOT scores and e-values are given for a combined search of manually obtained MALDI-TOF/TOF MS and MS/MS data in MASCOT against all plants in NCBI. According to MASCOT, a protein score above 65 was in this case significant ($p < 0.05$).

e) Identified by MALDI-TOF MS data in MASCOT search.

f) Proteins also found in the strawberry database.

g) Confirmed by similarity between MS using clustering program under development (Alm, R. *et al.*, to be published).

h) Confirmed by *de novo* sequencing of at least three SPITC-derivatized peptides subjected to MSBlast or a significant MASCOT score (applied to 20 spots in total).

i) All spots containing isoforms of the Bet v 1-homologous strawberry allergen, Fra a 1 (marked in black) were down-regulated in white compared to red strawberries. Further sequence information on the Bet v 1-homologous strawberry allergen is given in Figs. 4, 6.

3.2 Identification of proteins differentially expressed in red and white strawberries

Of the 96 protein spots that were differentially expressed, protein identification was successful for 55. Several proteins were detected in more than one spot and seven spots were found to contain two proteins each, resulting in identification of a total of 32 different proteins (Table 1).

Nine proteins were identified by matching to some of the few hundred strawberry sequences that are present in NCBI. The majority (23) of the proteins were identified based on sequence homology to proteins in other plant species. The mass spectrometric data were additionally searched against strawberry EST sequences with the software VEMS. The corresponding *F. ananassa* homologs were identified for 6 out of 23 proteins that

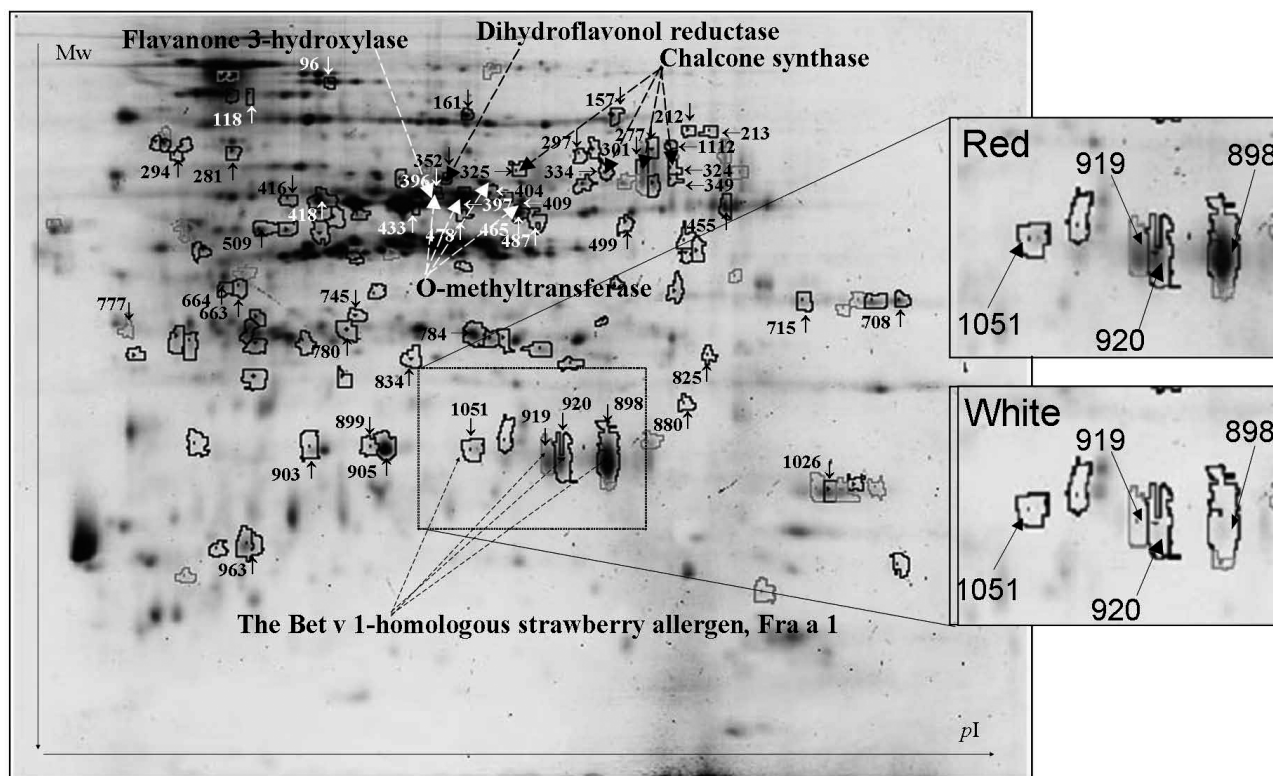


Figure 1. Gel image of strawberry proteins separated by 2-D PAGE. Spots encircled were differentially expressed in red and white strawberries, evaluated as described in Fig. 2. Spots with protein names stated were all down-regulated in the white strawberry variety. Spots marked with arrow and number are listed in Table 1 with protein identifications. Closeups show images of the spots containing the Bet v 1-homologous strawberry allergen [11] from one gel each with red and white strawberry, respectively.

had been identified by sequence homology (marked with F in Table 1), but no new identifications were obtained.

MS/MS MASCOT searches in Viridiplantae at NCBI initially gave a significant hit to the Bet v 1 homolog for only one spot (898) and this was based on only a single peptide. By a combination of Error Tolerant Search in MASCOT, and a parallel feature in the software VEMS, which also allows for indels (insertions/deletions), the mass spectrometric data were researched against the homologous proteins and supplemented by manual *de novo* sequencing of peptides. By this approach, all but three low abundant peaks in the MS spectra could be assigned to the allergen (all peaks confirmed by MS/MS) or to minor project specific contaminants and tryptic autodigest ion products (Fig. 3). Utilizing this method, the allergen was additionally found to be the dominating protein in three other spots (919, 920, 1051, data not shown).

Taken together, the measures taken to improve identification finally yielded a success rate at 58%. Thus out of 96 spots, proteins were successfully identified in 55. For the remaining 41 spots the identification failed, even when, as in most cases, the mass spectra were of high quality. This is presumably due to the limited number of strawberry protein- and EST-sequences available and limited homology to other plant species.

To exclude the possibility that down-regulation could be balanced by up-regulation of proteins in other spots, the mass spectra from all up-regulated spots were reanalyzed using the software PeakErazor. The peptide masses for the allergen and flavonoid biosynthesis enzymes were placed in the erazor list. In this way PeakErazor would reveal if any of the spectra contained peaks, which might be derived from the mentioned proteins. In addition, all mass spectra were checked using a new software tool being developed for clustering of MS spectra in order to identify isoforms from 2-D gels (Alm, R. *et al.*, to be published). Neither analysis with PeakErazor nor clustering analysis gave support for the presence of allergen or flavonoid biosynthesis enzymes in up-regulated spots.

3.3 Sequence information for the strawberry allergen Fra a 1

The strawberry Bet v 1-homologous allergen, only recently discovered by us [11], should be named Fra a 1 according to present allergen nomenclature. At least two EST-sequences (contig_Fra from the variety Strawberry Festival and jB125 from Elsanta) were found to match MS/MS data representing the allergen. Contig_Fra consists of four overlapping

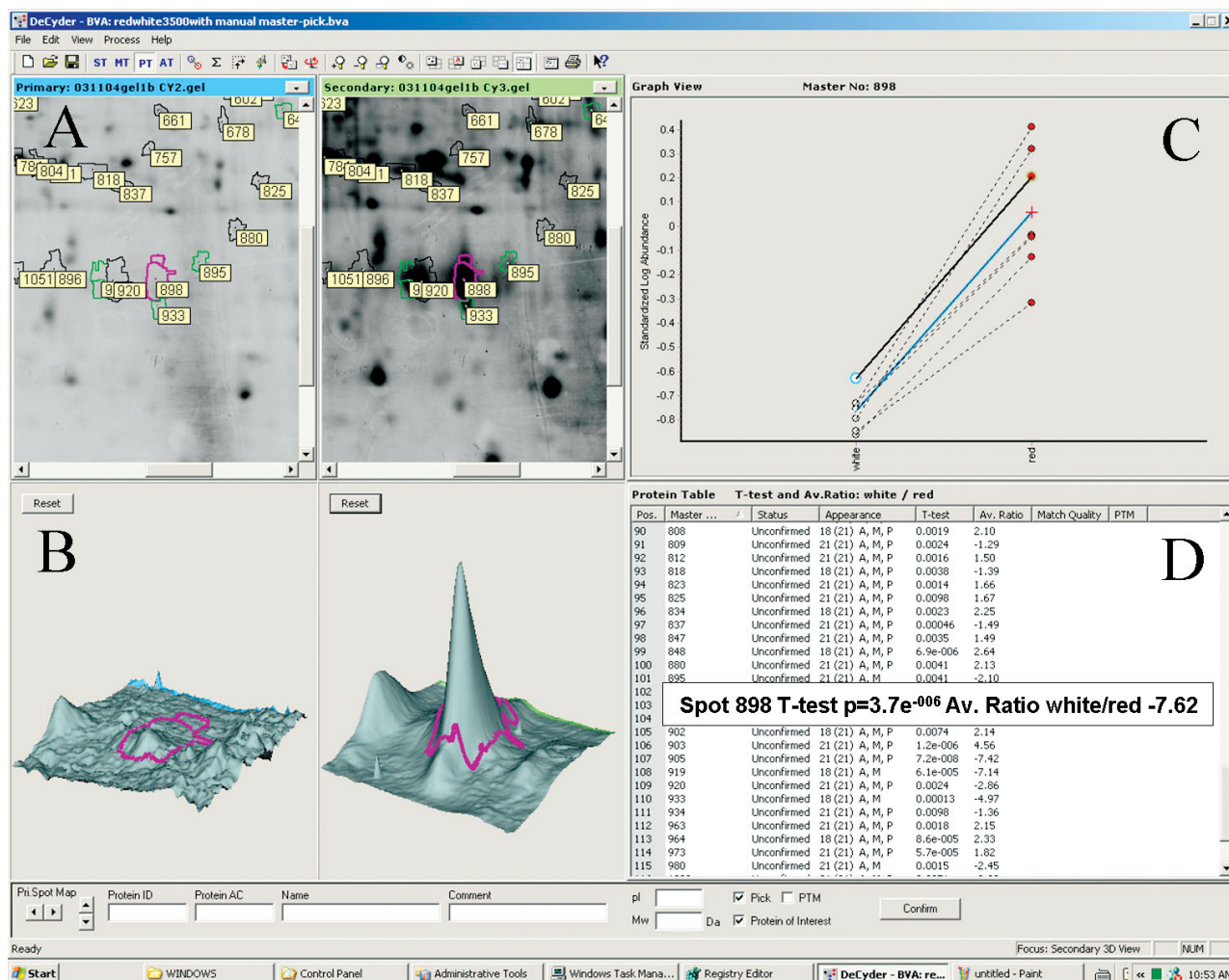


Figure 2. Evaluation of differential expression in red and white strawberry varieties by 2-D DIGE. Data obtained from 2-D DIGE using the software DeCyder Biological Variation Analysis module for spot 898, the major allergen spot. Panels: (A) region with allergen-containing spots on from one gel each of red and white strawberry, respectively, (B) 3-D representation of the difference in spot intensity between the two gels in A, (C) graph view of intensity difference for seven replicate gels each of red and white strawberries, (D) table listing *t*-test, *p*-value, and average ratio (=spot intensity in white variety divided by spot intensity in red variety; a negative value is used to signify down-regulation in white compared to red).

ESTs and covers the entire protein sequence. As can be seen in Fig. 4, these ESTs align well with Bet v 1 and Bet v 1 homologous sequences. Another nucleotide sequence from strawberry variety Elsanta (AY679601) was discovered by a homology search. The nucleotide sequence translation into protein made in TrEMBL was, however, not correct. A new translation, based on homology to the Mal d 1 protein utilizing two frameshifts (marked by black triangles) is also included in the alignment in Fig. 4. The EST-sequence designed Contig_Fra showed 77% amino acid identity with Mal d 1 (160 residues overlap, score 634.0, gap frequency 0.6%) and 54% amino acid identity with Bet v 1 (160 residues overlap, score 469, no gap introduced).

Additional amino acid sequence information was generated from an allergen-containing spot in Elsanta after derivatization. The peptide mixture obtained from in-gel digestion was treated with *O*-methylisourea [31] to convert lysines into homoarginines (no arginines are present in the strawberry allergen, see Fig. 4), and with SPITC to assist fragmentation [17, 19]. The spectra yielded near-complete γ -ion series for 11 such double-derivatized peptides (Table 2, column 5), 6 of these are shown in Fig. 5. Several of these are peptides not identified by MS/MS of underivatized peptides from the Korona variety. Complete amino acid sequence was obtained by the double-derivatization method for two relatively large peptides (MH^+ 1930.7 and 2382.0 Da), which were only partially sequenced from Korona.

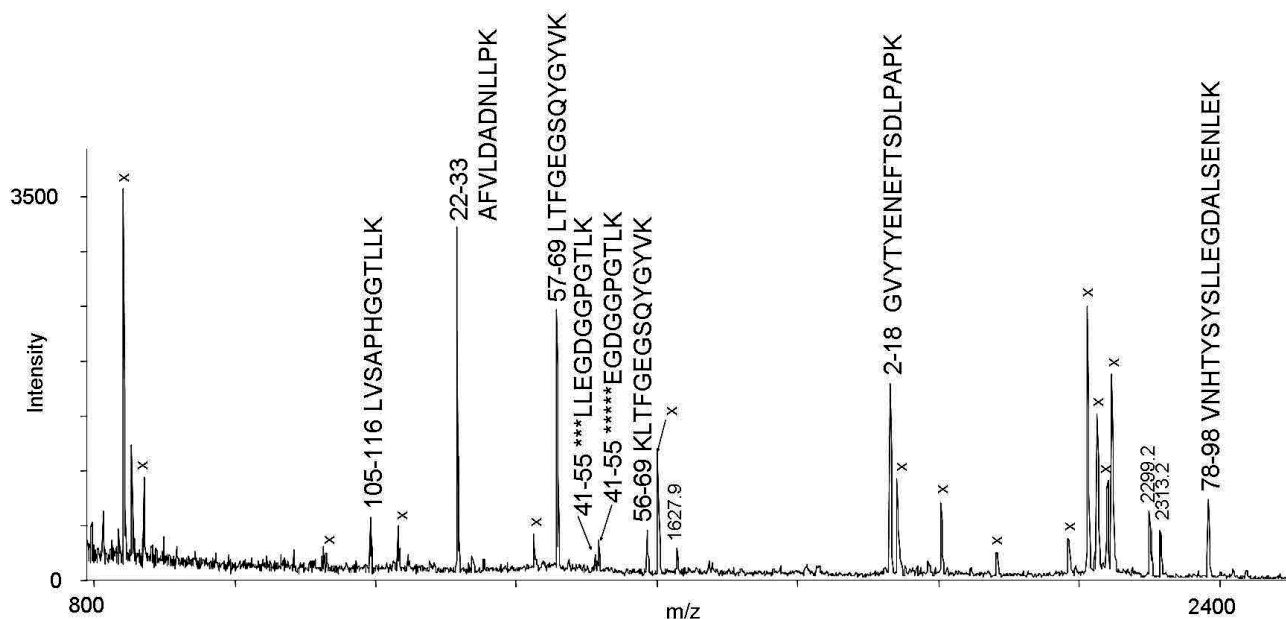


Figure 3. MALDI-TOF mass spectrum of tryptic digest obtained from spot 898 containing the Bet v 1-homologous strawberry allergen. Peaks assigned to peptides derived from the allergen are marked by their amino acid sequence and project-specific contaminants are marked by x. Peptide masses and sequences are summarized in Table 2, column 1 and their location in the sequence indicated in the alignment in Fig. 6.

An earlier study of the strawberry variety Camarosa had been performed with 1-D gel and nano-LC MS/MS [11]. These data were researched against our new sequence data and as a result two more peptides were confirmed by MS/MS as belonging to Fra a 1. (Table 2, column 6). All mass spectrometric data obtained for allergen peptides in Table 2 are aligned in Fig. 6. Sequence coverage of 75% (119 out of 159) was obtained. Complete sequence coverage is difficult to obtain due to loss of small peptides in the micropurification step [32] and the low ionization efficiency for, e.g., arginine-deficient peptides with a mass below 1000 Da.

3.4 Strawberry allergen isoforms and amino acid variability

The appearance of proteins in multiple spots in 2-D PAGE can be due to PTMs [33–35], splice variants, protein degradation, or allelic variation. Isoforms are reported for birch and apple homologs [7, 9]. Multiple spots reported in, e.g., barley [36] might be typical for polyploidal plants. In strawberry, which is octaploid, 16 alleles can in principle encode many different gene products so detection of several spots representing the same protein, or several gene products within one spot, due to allelic variation is likely.

Sequence analysis of peptides from the four allergen spots revealed some variability as shown by the triplicate mass spectra in Fig. 7 for the peptide LVSAPHGGTLLK (aa 105–116 in Contig_Fra). The peptide variants differ in mass by 14 and 57, respectively (underivatized peptide mass values

at m/z 1178.8, 1192.8, and 1135.7, Table 2). *De novo* sequencing of the double-derivatized peptides showed that mass deviations were due to substitution of leucine (LVSAPHGGTLLK, 1192.8) to valine (LVSAPHGGTVLK, 1178.8), and to a deletion of a glycine (LVSAPHG_TLLK, 1135.7) in either position 111 or 112. One other peptide was similarly detected as two isoforms (AAHLFK, 686.4 and ASHLFK, 702.4, aa 141–146 in Contig_Fra, mass spectra not shown but see Table 2).

The MS-derived amino acid sequence information and amino acid sequence information derived from translation of nucleotide sequences are assembled and compared in Fig. 6. The data are from different strawberry varieties (Korona, Elsanta, Strawberry Festival, Camarosa). Only in approximately 15% of the positions there was conflicting data, which might be caused either by allelic variation or by sequencing errors in the ESTs. The overall impression is that sequence from different varieties is very similar based on the present data.

4 Discussion

We have demonstrated that a 2-DE-based proteomics approach can be used to visualize and quantitate differences in the protein expression for a red and a white strawberry variety, and to identify proteins in up- and down-regulated spots. We have found that the allergen as well as enzymes in the biosynthetic pathway for synthesis of red color are down-

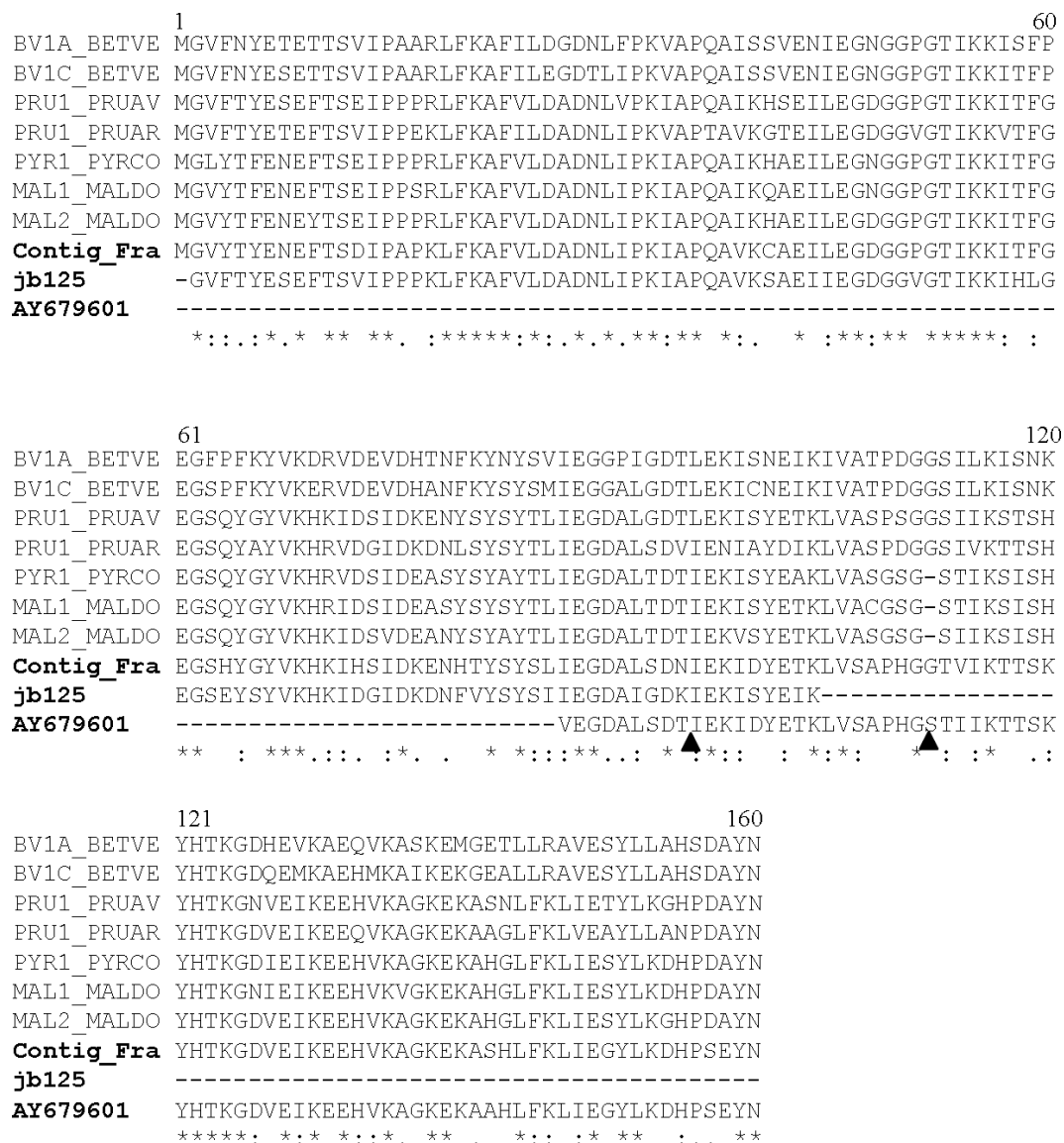


Figure 4. Multiple sequence alignment of birch pollen allergen Bet v 1 homologs including sequences from birch, cherry, apricot, pear, apple, and strawberry. Strawberry amino acid sequences (names in bold; Contig_Fra, Jb125 and AY679601) were translated from strawberry nucleotide sequences. Black triangles represent frameshifts introduced in the translation of AY679601. Positions with absolute conservation are noted with an asterisk and conserved and semiconserved substitutions are noted with two and one dots, respectively. Amino acid sequences derived from Swiss-Prot have the following accession numbers: BV1A_BETVE (P15494), BV1C_BETVE (P43176), PRU1_PRUAV (Q24248), PRU1_PRUAR (O50001), PYR1_PYRCO (O65200), MAL1_MALDO (P43211), and MAL2_MALDO (Q40280).

regulated in white strawberries. The down-regulation of this allergen offers an explanation as to why individuals affected by allergy to red strawberries can tolerate white strawberries.

In addition we have detected the presence of different isoforms of the strawberry Bet v 1 homolog and deduced a representative amino acid sequence for the allergen, which we named Fra a 1, according to the nomenclature for allergen proteins. This was done by a combination of MS and

nucleotide sequence information found in different databases. The deduced amino acid sequence can be used for cloning and further analysis of the allergen.

The entire genome of only two plants, *Arabidopsis thaliana* and rice, has been sequenced to date. However, far more plant species are of interest for proteome analysis such as for example strawberry, which has been investigated in the present study. Here protein identification based on sequence

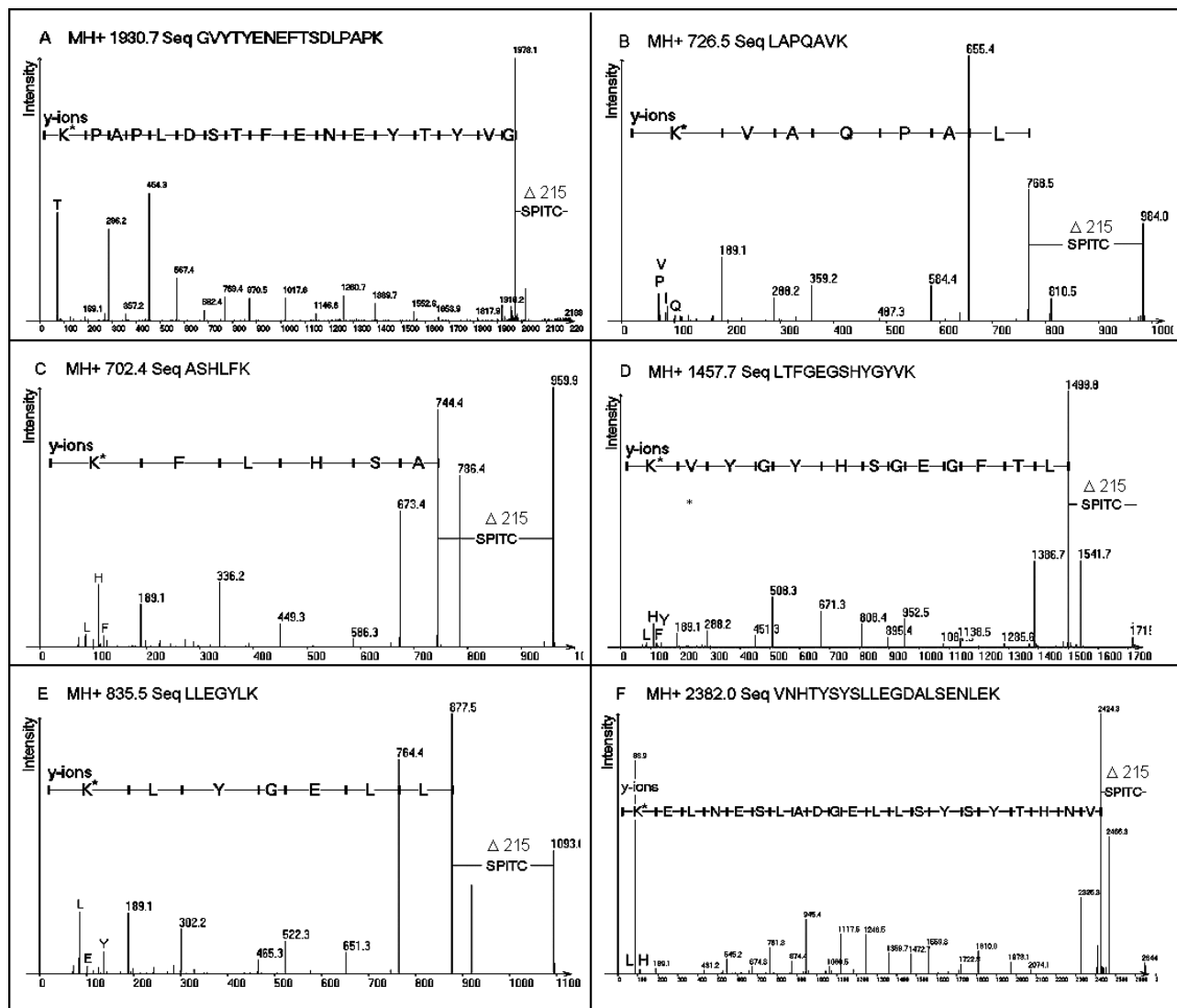


Figure 5. MALDI-MS/MS spectra obtained for the Bet v 1-homologous strawberry allergen after double derivatization of the peptides. *De novo* sequence information was generated after double derivatization with *O*-methylisourea and SPITC whereby the loss of the sulfonic acid derivatives from Y_N is observed in the MS/MS spectra as a neutral loss of 215 Da (indicated as $\Delta 215$ SPITC). Y_1 is observed as 189 Da due to the conversion of lysine into homoarginines ($\Delta 42$ Da, indicated as K^*). Immonium ions are indicated with the corresponding amino acid one-letter code. Peptide masses and sequences of these six peptides in panels A–F are summarized together with five other peptides in Table 2, column 5.

homology was improved by manual interpretation of MALDI-MS/MS spectra, *de novo* sequencing after peptide derivatization and indel-tolerant searching against local databases. This yielded a 58% success rate of protein identification. The here described proteomics set-up, therefore, also provides proof-of-principle for its capacity to discriminate between varieties in terms of differential protein expression, and a potential to be utilized with various other crop plants as a tool in breeding and searching for valuable phenotypic traits.

Of the differentially expressed spots, approximately half were up-regulated and half were down-regulated in the white compared to red strawberry variety. All four spots containing

the strawberry Bet v 1-homologous allergen were highly down-regulated. Among other spots that were down-regulated were chalcone synthase (four spots), dihydroflavonol reductase (one spot), flavanone 3-hydroxylase (one spot), and methyltransferase (four spots), see Table 1 and Fig. 1. These enzymes belong to the well-characterized flavonoid biosynthesis pathway, which produces several compounds including the pelargonidin, the red color of strawberry [37, 38]. Chalcone synthase, dihydroflavonol reductase, and flavanone 3-hydroxylase were also down-regulated in a colorless white mutant of bilberry, *Vaccinium myrtillus*, where genes in the flavonoid pathway were probed by digoxigenin-dUTP-labeling PCR [38]. Many colorless plant mutants appear to be

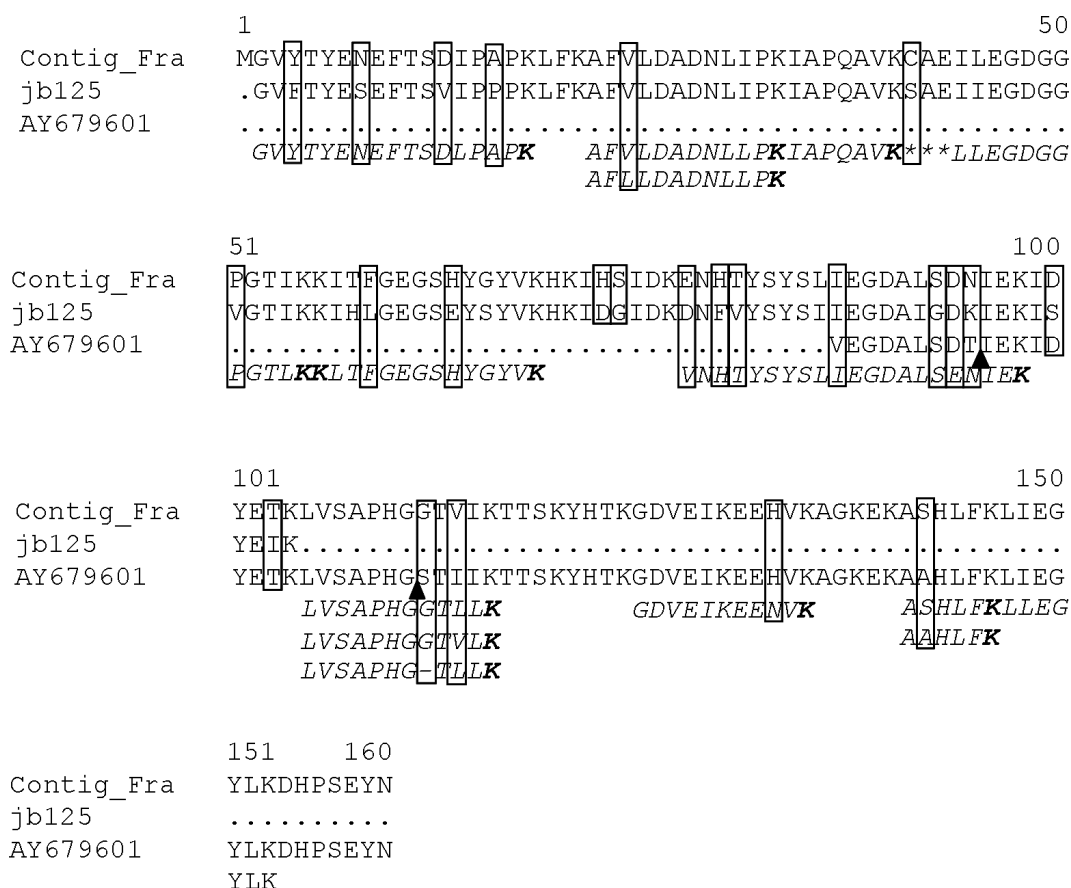


Figure 6. Comparison of amino acid sequence information obtained by MS and by translation of nucleotide sequence information for the strawberry Bet v 1-homologous allergen. All MS-derived sequence information (listed in Table 2) is aligned (in italics, observed tryptic cleavage sites in bold) below the translated nucleotide sequences information for the full-length protein (Contig_Fra) or only the C-terminal or the N-terminal half, respectively (AY679601 and Jb125). Frameshifts introduced in the translation of AY679601 are marked by black triangles. Residues found to differ between any two of the different sources of sequence information are boxed (24 positions out of 160). Stars indicate that the MS data were insufficient to obtain the full sequence. In the MS-derived sequence information the abbreviation L stands for both leucine and isoleucine, since the two could not be distinguished.

transcription factor mutants in which several of the enzymes in the flavonoid biosynthesis pathway are simultaneously down-regulated [39, 40], but present data cannot tell whether the down-regulation in white strawberries is due to failing transcription or lack of genes.

The biological function of the omnipresent Bet v 1-homologs in plants is not clear, although Bet v 1 is a well-studied allergen in terms of structure, epitopes, and epidemiology [1]. The here reported concerted down-regulation of the strawberry allergen and the flavonoid biosynthesis provide independent support to a role for Bet v 1 as a carrier of hydrophobic compounds, as recently suggested based on ANS (8-anilino-1-naphthalenesulphonic acid)-replacement binding studies *in vitro* with hydrophobic ligands such as flavone and naringenin to purified Bet v 1-protein [41]. Further support for the involvement of the flavonoid synthesis

pathway is found by 45% sequence identity between Bet v 1 and Hyp-1, the enzyme responsible for formation of hypericin, the red-colored naphthodianthrone in *Hypericum perforatum* [42]. If the main role of the Bet v 1 homologs in plants is to assist transport of flavonoid-like molecules, it makes sense that it is down-regulated along with the structural enzymes in the flavonoid biosynthesis pathway in the white strawberries, either due to cotranscriptional down-regulation of its synthesis or by increased degradation if it is “unemployed” in white strawberries.

The majority of the 1000 resolved proteins in the strawberry proteome were *not* differentially expressed in white compared to red strawberries, including many of the abundant proteins. Among the spots that were up-regulated in white strawberries were ascorbate peroxidase (two spots), methionine sulfoxide reductase (one spot), and proteasome

Table 2. Summary of mass spectrometric data obtained for peptides from the Bet v 1-homologous strawberry allergen. Peptide sequence information obtained from underivatized peptides from spots 898, 919, 920, and 1051 (see Fig. 1) is presented in column nr 1–4. Column nr 5 presents peptide sequence data obtained from one of the allergen spots (920) from the strawberry variety Elsanta, for which peptides were double derivatized to improve the MS/MS spectra. Column 6 presents peptide sequence data from a third strawberry variety, Camarosa, obtained by searching against the translated nucleotide sequences in Fig. 4 with previously recorded mass spectra obtained by LC-nESI-QTOF of a 20 kDa protein band excised from a 1-D gel [11]

MH ⁺ ^{a)}	Peptide sequence ^{b)}	Position in se- quence	1 898 MALDI- TOF/TOF, Korona	2 919 MALDI- TOF/TOF, Korona	3 920 MALDI- TOF/TOF, Korona	4 1051 MALDI- TOF/TOF, Korona	5 920 MALDI- TOF/TOF, Elsanta, double derivatiza- tion	6 20 kDa 1-D gel, LC nESI QTOF Camarosa
686.4	AAHLFK	141–146	ND	ND	ND	ND	c), d)	
702.4	ASHLFK	141–146	ND	ND	ND	ND	c), d)	
726.5	IAPQAVK	34–40	c)				c), d)	
835.5	LLEGYLK	147–153	c)		c)		c), d)	c), d)
1135.7	LVSAPHGTLLK	e) 105–116			c)		c), d)	
1178.8	LVSAPHGGTVLK	105–116			c), d)		c), d)	c), d)
1192.8	LVSAPHGGTLLK	105–116	c), d)	c), d)	c), d)	c), d)	c), d)	c), d)
1282.7	GDVELKEEHVK	124–134			c)	c)		c), d)
1315.6	AFVLDADNLLPK	22–33	c), d)		c)	c)	c), d)	c), d)
1328.7	AFLDADNLLPK	22–33						c), d)
1457.7	LTFGEGSHYGYVK	57–69	c), d)	c), d)	c)	c), d)	c), d)	c), d)
1511.7	(355)LLEGDGGPGTLK	41–55	c), d)	c)	c), d)	c)		
1516.7	(586)EGDGGPGTLK	41–55	c), d)	c)	c)			
1585.8	KLTFGEGSHYGYVK	56–69	c), d)		c)			c), d)
1930.7	GVYTYENEFTSDLPAK	2–18	c), d)		c)		c), d)	
2382.0	VNHTYSYSLLEGDALSENLEK	78–98	c), d)		c)		c), d)	c), d)

a) MH⁺ values refers to the underivatized peptides.

b) Amino acid residues leucine and isoleucine are both symbolized by L since they are isobaric.

c) A signal at the corresponding *m/z* value was detected.

d) The sequence was confirmed by MS/MS.

e) This peptide contains a deletion in positions 111 or 112.

ND: peptides with *m/z* < 720 were not detected in MALDI-MS due to the mass range selected.

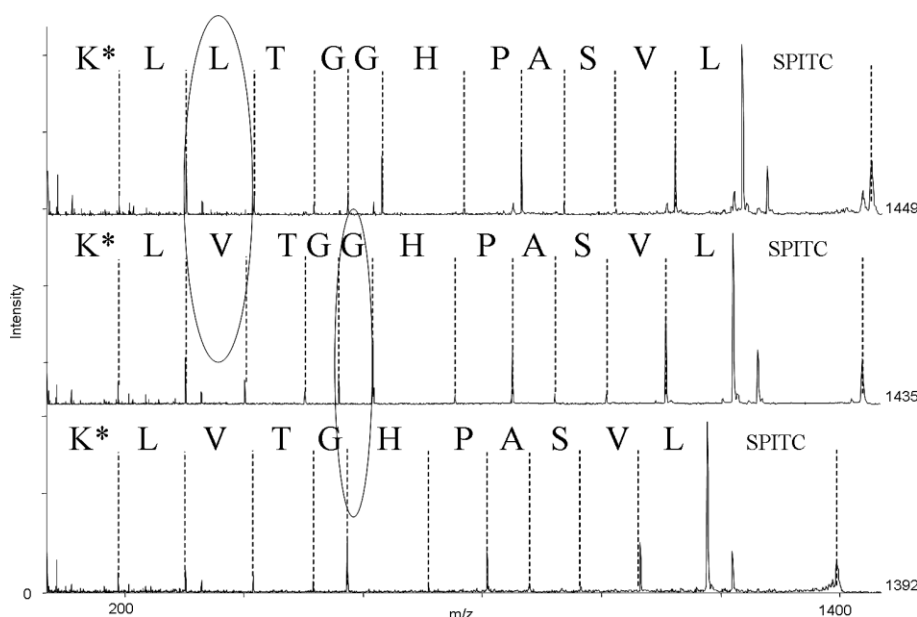


Figure 7. Peptide heterogeneity detected in MALDI-MS/MS mass spectra indicating the presence of different isoforms of the Bet v 1-homologous strawberry allergen within one single spot. Underivatized precursor ion masses are 1192.8, 1178.8, and 1135.8 (Table 2), respectively, which due to the double derivatization, as explained in legend to Fig. 5, were observed in mass spectra as 1449, 1435, and 1392, respectively. Deduced amino acid sequences are LVSAPHGGTLLK, LVSAPHGGTVLK, and LVSAPHGTVLK corresponding to positions 105–116 in alignments in Fig. 6.

subunits (two spots). These proteins are associated with oxidative stress and protein degradation, suggesting that the white strawberries experience stress under normal growth conditions. This seems likely if they, like the white bilberry mutant [38], have reduced levels of flavonoids, which are photoprotective screeners and antioxidants [43].

In summary, the present study shows that it is possible to derive meaningful information by proteomics applied to a species without known genome by the use of manual interpretation of MS/MS data combined with peptide derivatization and bioinformatics. By this approach it has been possible to identify more than half of the up- or down-regulated proteins, to shed light of the biological involvement of one of the most important plant allergen families and to almost fully sequence the strawberry allergen.

We thank Elma Salentijn and Asap Aharoni at Plant Research International for providing us with the unpublished EST-sequences from *Elsanta*, and Åke Truedsson for sharing his observations concerning adverse reactions to strawberry and tolerance to white strawberries. The staff at the Lund University Swegene Proteome Center and the Protein Research group in Odense is acknowledged for skilful technical assistance. This study was supported by a grant from the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (FORMAS, 2002–0042) to C.S.E. Funding for instrumentation by grants from the Danish Research Agency to the Danish Biotechnology Instrument Center (DABIC) and salary to K.H. from the Center for Experimental Bioinformatics (CEBI) founded by the Danish National Research Foundation are acknowledged.

5 References

- Vieths, S., Scheurer, S., Ballmer-Weber, B., *Ann. NY Acad. Sci.* 2002, **964**, 47–68.
- Aalberse, R. C., Akkerdaas, J. H., van Ree, R., *Allergy* 2001, **56**, 478–490.
- Kazemi-Shirazi, L., Pauli, G., Purohit, A., Spitzauer, S. *et al.*, *J. Allergy Clin. Immunol.* 2000, **105**, 116–125.
- Breiteneder, H., Ebner, C., *J. Allergy Clin. Immunol.* 2000, **106**, 27–36.
- Mirza, O., Henriksen, A., Ipsen, H., Larsen, J. N. *et al.*, *J. Immunol.* 2000, **165**, 331–338.
- Vieths, S., Schoning, B., Petersen, A., *Int. Arch. Allergy Immunol.* 1994, **104**, 399–404.
- Son, D. Y., Scheurer, S., Hoffmann, A., Hausteiner, D. *et al.*, *Eur. J. Nutr.* 1999, **38**, 201–215.
- Reuter, A., Fortunato, D., Garoffo, L. P., Napolitano, L. *et al.*, *Proteomics* 2005, **5**, 282–289.
- Holm, J., Baerentzen, G., Gajhede, M., Ipsen, H. *et al.*, *J. Chromatogr. B* 2001, **756**, 307–313.
- Ferreira, F., Hirtenlehner, K., Jilek, A., Godnik-Cvar, J. *et al.*, *J. Exp. Med.* 1996, **183**, 599–609.
- Karlsson, A. L., Alm, R., Ekstrand, B., Fjellkner-Modig, S. *et al.*, *Allergy* 2004, **59**, 1277–1284.
- Lamanda, A., Zahn, A., Roder, D., Langen, H., *Proteomics* 2004, **4**, 599–608.
- Alban, A., David, S. O., Bjorkestén, L., Andersson, C. *et al.*, *Proteomics* 2003, **3**, 36–44.
- Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J. S., *Electrophoresis* 1999, **20**, 3551–3567.
- Levander, F., Rognvaldsson, T., Samuelsson, J., James, P., *Proteomics* 2004, **4**, 2594–2601.
- Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R. *et al.*, *J. Mass Spectrom.* 1999, **34**, 105–116.
- Marekov, L. N., Steinert, P. M., *J. Mass Spectrom.* 2003, **38**, 373–377.
- Wang, D. X., Kalb, S. R., Cotter, R. J., *Rapid Commun. Mass Spectrom.* 2004, **18**, 96–102.
- Chen, P., Nie, S., Mi, W., Wang, X. C. *et al.*, *Rapid Commun. Mass Spectrom.* 2004, **18**, 191–198.
- Peri, S., Steen, H., Pandey, A., *Trends Biochem. Sci.* 2001, **26**, 687–689.
- Hjerno, K., Hojrup, P., in: Kamp, R. M., Calvete, J. J., Choli-Papadopoulou, T. (Eds.), *Methods in Proteome and Protein Analysis*, Springer-Verlag 2004, pp. 359–369.
- Matthiesen, R., Lundsgaard, M., Welinder, K. G., Bauw, G., *Bioinformatics* 2003, **19**, 792–793.
- Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A. *et al.*, *Anal. Chem.* 2001, **73**, 1917–1926.
- Thompson, J. D., Plewniak, F., Thierry, J. C., Poch, O., *Nucleic Acids Res.* 2000, **28**, 2919–2926.
- Huang, X. Q., Miller, W., *Adv. Appl. Math.* 1991, **12**, 337–357.
- Boguski, M. S., Lowe, T. M. J., Tolstoshev, C. M., *Nat. Genet.* 1993, **4**, 332–333.
- Schiex, T., Gouzy, J., Moisan, A., de Oliveira, Y., *Nucleic Acids Res.* 2003, **31**, 3738–3741.
- Kulikova, T., Aldebert, P., Althorpe, N., Baker, W. *et al.*, *Nucleic Acids Res.* 2004, **32**, D27–D30.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. *et al.*, *J. Mol. Biol.* 1990, **215**, 403–410.
- Huang, X. Q., Chao, K. M., *Bioinformatics* 2003, **19**, 228–233.
- Keough, T., Lacey, M. P., Youngquist, R. S., *Rapid Commun. Mass Spectrom.* 2000, **14**, 2348–2356.
- Larsen, M. R., Cordwell, S. J., Roepstorff, P., *Proteomics* 2002, **2**, 1277–1287.
- Larsen, M. R., Roepstorff, P., *Fresenius J. Anal. Chem.* 2000, **366**, 677–690.
- Mann, M., Jensen, O. N., *Nat. Biotechnol.* 2003, **21**, 255–261.
- Jensen, O. N., *Curr. Opin. Chem. Biol.* 2004, **8**, 33–41.
- Laugesen, S., Roepstorff, P., *J. Am. Soc. Mass Spectrom.* 2003, **14**, 992–1002.
- Winkel-Shirley, B., *Plant Physiol.* 2001, **126**, 485–493.
- Jaakola, L., Maatta, K., Pirttila, A. M., Torronen, R. *et al.*, *Plant Physiol.* 2002, **130**, 729–739.
- Spelt, C., Quattrocchio, F., Mol, J., Koes, R., *Plant Cell* 2002, **14**, 2121–2135.
- Park, K. I., Choi, J. D., Hoshino, A., Morita, Y. *et al.*, *Plant J.* 2004, **38**, 840–849.
- Mogensen, J. E., Wimmer, R., Larsen, J. N., Spangfort, M. D. *et al.*, *J. Biol. Chem.* 2002, **277**, 23684–23692.
- Bais, H. P., Vepachedu, R., Lawrence, C. B., Stermitz, F. R. *et al.*, *J. Biol. Chem.* 2003, **278**, 32413–32422.
- Winkel-Shirley, B., *Curr. Opin. Plant Biol.* 2002, **5**, 218–223.