Report from lab of Kathryn Lilley, Cambridge (2002)

Proteome analysis of sub-cellular fractions from Cauliflower curd (*Brassica oleracea*)

Summary

Cauliflower belongs to the Brassica genus and is a variety of the *Brassica oleracea* species. Diploid Brassicas are believed to have evolved through triplication and extensive rearrangement of an Arabidopsis like genome. Arabidopsis was chosen as the 'model plant' due to the compact nature of its genome and currently has one of the best characterised genomes available. The evolutionary related Brassicas have a more expanded genome making characterisation more difficult. Brassica species represent a number of economoically important crops, including cauliflower, oilseed rape, mustard and cabbage. Information obtained from the evolutionary relationship between these crop plants and Arabidopsis is extremely valuable in the discovery and manipulation of agrnomically important traits. Although there has been comparative work carried out at the gene level in Brassicas and Arabidopsis, very little work has been carried out at the protein level. It is the proteins that control cellular function, therefore it is important to understand the Brassica proteome. This study has shown that it is possible to gain insight into *B.oleracea* proteins using their evolutionary relationship with Arabidopsis.

Aims

The aim of the project was to investigate the cauliflower proteome using two-dimensional polyacrylamide electrophoresis (2D PAGE) and MALDI-Tof (Matrix assisted laser desorption time of flight) mass spectrometry. A comparison of the data from the cauliflower callus proteome and data from the equivalent Arabidopsis callus proteome were used to assess the relative success of identification of the isolated cauliflower proteins using the Arabidopsis translated genome sequence.
MALDI-Tof mass spectrometry is a relatively simple high throughput technique which can be used to generate peptide mass fingerprints of proteins. If this data can be used to identify a Brassica protein by virtue of its similarity to a protein in the extensive Arabidopsis databases, a wealth of information could be generated in a simple and potentially high throughput manner.

Work Done

In order to obtain the proteins to carry out 2D PAGE, the following preparation was undertaken. Cauliflower proteins were isolated by homogenisation of callus (non-photosynthetic edible portion of the cauliflower plant). Sucrose density centrifugation was then employed to separate soluble and membrane proteins into different fractions. The proteins were further purified from various plant metabolites using phenol precipitation. The extracted soluble proteins were dissolved in a buffer containing CHAPS detergent and urea. The extracted membrane proteins were dissolved in a series of detergents with the aim of isolating increasingly hydrophobic proteins. The initial solubilisation buffer used contained a mixture of CHAPS, thiourea and urea. Material insoluble in this mixture was then subjected to a stronger zwitterionic detergent, ASB14 (amidosulphobetaine-14). Material insoluble in this detergent was boiled in the presence of 2% SDS. This method of fractionation was used to simply the mixtures of proteins applied to individual 2D PAGE gels improving the chance to resolving individual proteins by this technique. The final five fractions obtained were:

1. S1 = Soluble proteins in CHAPS / urea

2. S2 = Soluble organelle proteins inCHAPS / urea (the supernatant from the ultracentrifugation to pellet membrane protein)
3. Membrane and membrane associated protein in CHAPS / urea / thiourea.
4. Membrane and membrane associated protein in ASB14 / thiorea / urea
5. Membrane and membrane associated protein in 2% SDS

A protein assay was performed in order to determine protein concentration in the fractions which did contain thiourea. The amount of protein in fractions containing thiourea was carried out by1D PAGE. 2D PAGE was used to separate proteins based on their isoelectric point and denatured molecular weight. Resulting gels were silver stained and protein spots visualised excised. Peptides were generated from the proteins within excised spots using trypsin and peptide mass fingerprinting carried by MALDI-Tof MS. Not all samples generated could be analysed due to time restrictions. Only fraction S3 was chosen for the focus of the investigation.Figure 1 shows an image of a silver stained gel of the S3 fraction The excised spots underwent silver destaining, reduction, alkylation and tryptic digestion on an automated Mass Prep station. The digested samples were desalted using zip-tips (Millipore). MALDI-ToF mass spectrometry was used to determine the masses of the peptides generated by tryptic digestion. The resulting fragmentation data was used to search virtual fragmentation databases created from the translated Arabidopsis genome sequence using the Mascot suite of protein identification programmes (Matrix Science Ltd.).

The whole method was repeated with Arabidopsis callus tissue (artificially cultured, non-photosynthetic tissue) and the equivalent CHAPS/thiourea dissolved membrane protein fraction A3, was analysed in the same way by MALDI-ToF mass spectrometry.

Results

Out of 187 spots analysed (106 from the gel of S3, an additional 9 from the gels of S4 and S5 and 72 from the gel of S1), 60 appeared to contain sample material worth considering. Peptide mass fingerprinting positively identified 11 spots as having sequences consistent with Arabidopsis proteins from the *B.oleracea* peptides, and 7 as Brassica proteins. This was a surprisingly low level of identification. Many of the samples appeared to contain very little or no sample material. This may have been as a result of one or a combination of, not enough sample material being taken from the gel in the first place or sample material being lost during the peptide preparation stages. There also appeared to be some contamination in the gel of S3, revealed by a recurring pattern of masses visible in the fragmentation data. This could have lead to suppression of the real peptides by the contaminant being preferentially ionised. There was also some keratin contamination which is a common problem and arises from sources such as hair, clothing and dust.

The 9 spots from gels of S4 and S5 were analysed to see if they too showed the same recurring pattern of masses as seen in the contaminated samples in S3. The pattern was not visible in any of the 9 samples. A second gel of the S1 fraction was run, again to check for contamination of which none was seen. In light of problems encountered with the gel of S3, samples that gave no hits on the MALDI but looked as if they may contain peptides were analysed on the ESI-Q-ToF (Electrospray Ionisation - Quadrupole -Time of Flight). This technique generates fragment ions from which identifications of peptides which may be homologous, but not necessarily 100%

identical, can be made and also from which identification of a mixture of peptides from different proteins can be made. A further 10 spots were identified as having sequences consistent with Arabidopsis proteins or as Brassica proteins on the Q-ToF, making a total of 28 positive hits. In the cases where no protein was identified either there was insufficient sample material for this type of analysis or the *B.oleracea* protein isolated had diverged from the common ancestor to such an extent that it was no longer identifiable as its Arabidopsis homologue. A specific example in this category was spot E12. The original MASCOT search carried out using the MALDI data identified the protein as a *B.oleracea* protein (score 79). The Arabidopsis homologue was identified with a score of only 36. This hit alone would not have been considered as significant. In the case of D2 from gel S4, identified as a *B.napus* homologue, no Arabidopsis homologue was isolated at all.

On analysis of the gel of A3, containing Arabidopsis callus proteins from the CHAPS membrane fraction, 58 out of 72 spots appeared to have sample material worth considering. 21 were positively identified as Arabidopsis proteins using MALDI-ToF mass spectrometry. There was no apparent contamination of this gel as in the case of gel S3. Possible reasons for some proteins not being identified are insufficient sample material or the presence of a mixture of peptides from different proteins.

Summary of results

21 out of 58 Arabidopsis spots identified (36%) all by peptide mass fingerprinting

28 out of 60 Brassica spots identified (47%)

- 18/28 identified by peptide mass fingerprinting (MALDI-MS)
- 10/28 could not be identified by peptide mass fingerprinting, but were identified by ESI-Q-ToF

Of the 28 out of 60 Brassica spots identified:

- 8/28 were identified as Brassica gene products
  - 7 identified by peptide mass fingerprinting
  - 1 identified by ESI-Q-ToF
- 20/28 were not identified as Brassica gene products, but were identified as having sequences homologous to Arabidopsis gene products.
  - 11 identified by peptide mass fingerprinting
  - 9 identified by ESI-Q-ToF

Conclusions.

In conclusion 18/28 (64%) of peptides generated from Brassica 2D PAGE spots of sufficient quality could be identified either as Brassica gene products or Arabidopsis homologues. Due to a restricted time period, sufficient results from which a valid statistical comparison could be made were not produced. Although, in general, the number of positive identifications from Arabidopsis gels and Brassica gels is comparable. Preliminary data would therefore suggest there is some value in further pursuing this method.

Additional Bioinformatics Work

The 20 cases in which positive identifications of *B.oleracea* proteins as their Arabidopsis homologues were made, further analysis was carried out using other bioinformatic tools. Each Arabidopsis protein has an individual At identification number. The At number of each protein identified was used to search the MIPS (Munich Information Centre for Protein Sequences) database to obtain the full protein sequence. The protein sequence was then used to BLAST search a specific Brassica database, developed and maintained at the John Innes Centre and funded by the BBSRC UK CropNet initiative. Currently the Brassica genome has only been partially sequenced and an even smaller proportion of the sequenced genome has been annotated. Out of 16 different protein sequences obtained from MIPS and used to BLAST search the Brassica database, only two were shown to be present (spot H10 from the S3 fraction and spot D7 from the S1 fraction). In this case the peptide fragments identified in the MALDI were also used to search the Brassica database to confirm the correct protein had been identified. It is likely that the other Brassica proteins, which have been identified as being significantly similar to their Arabidopsis homologue, have not yet been correctly annotated. As part of the Brassica Investigating Genome Function initiative, a project is currently being run to integrate maps of the Brassica genome with those of the Arabidopsis genome. It is hoped that the information from this study will be of use in this project.
http://www.mips.biochem.mpg.de/
http://ukcrop.net/perl/ace/ncbi_blast/BrassicaDB
http://www.matrixscience.com

"My Experience"

I started my undergraduate degree with a view to pursuing a career in scientific research. This project has provided me with an excellent introduction to working in a research laboratory. I have been introduced to a new field of biochemistry in which I had no previous experience and have had the opportunity to learn several new practical techniques. I now feel confident in running 2D gels, in obtaining data from the gels by MALDI-ToF mass spectrometry and in analysing the obtained data. The experience I have gained in using the MALDI-ToF mass spectrometer is of particular value due to its importance as a current and future analytical technique and the shortage of experienced users in the UK. My supervisor, Kathryn Lilley, and the laboratory staff have been extremely helpful in teaching me and allowing me to perfect these new techniques. I have enjoyed working on this project and feel the skills I have learnt will be valuable during the rest of my undergraduate course and beyond. My time spent working here has definitely inspired me to continue in scientific research.

Report from lab of Kathryn Lilley, Cambridge (2002)

Table of Results

	Spot Number	MALDI/Q-ToF identified arabidopsis protein (score)	UK CropNet Brassica Protein DB Result	Molecular weight (kDa); pI value of Arabidopsis protein
S3 maldi results Score >73 significant	**C1**	vacuolar-type H+-ATPase subunit E3 (VHA-E3) (90)	***** No hits found ******	(M_r): **27068**; pI value: **5.82**
		disease resistance protein, putative (i)	**Brassica napus** Disease resistance-like protein (Fragment).	(Mr): **35814**; pI value: **6.05**
	C2	vacuolar-type H+-ATPase subunit E1 (81)	***** No hits found *****	(M_r): **26044**; pI value: **6.04**
	C5	putative protein (89)	****No significant hits****	(M_r): **54168**; pI value: **8.56**
	E12	**Brassica oleracea**, Cytochrome B5 (79)	Search not necessary	(M_r): **15053**; pI value: **4.90**
		cytochrome b5 (36)	**Brassica oleracea**, Cytochrome b5.	(M_r): **15075**; pI value: **5.11**
	H9	putative kinesin heavy chain (74)	****No significant hits****	(M_r): **121475**; pI value: **5.91**
	H10 (ii)	H+-ATP synthase beta chain-like protein (108)	**Brassica napus**, ATP synthase beta subunit. (iii)	(M_r): **59676**; pI value: **6.18**
S3 Q-ToF results Score ~>50 significant	**A8**	putative thioredoxin (86)	****No significant hits****	(M_r): **56614**; pI value: **4.90**
		putative protein disulfide isomerase precursor (81) (iv)	****No significant hits****	(M_r): **55852**; pI value: **4.81**
	B3	endomembrane-associated protein (76)	***** No hits found ******	(M_r): **24569**; pI value: **4.99**
	B4	endomembrane-associated protein (81)	***** No hits found ******	(M_r): **24569**; pI value: **4.99**
	B10	endomembrane-associated protein (54)	***** No hits found ******	(M_r): **24569**; pI value: **4.99**
	D6	putative protein (234)	***** No hits found *****	(M_r): **20173**; pI value: **5.63**
	E1	mitochondrial F0 ATP synthase D chain (208)	***** No hits found *****	(M_r): **17169**; pI value: **4.97**
	E2	mitochondrial F0 ATP synthase D chain (175	***** No hits found *****	M_r): **17169**; pI value: **4.97**
	E3	mitochondrial F0 ATP synthase D chain (152)	***** No hits found *****	(M_r): **17169**; pI value: **4.97**
	H11	putative protein (111)	***** No hits found *****	(M_r): **49496**; pI value: **5.99**
S4 maldi results	**D1 (v)**	H+-ATP synthase beta chain-like protein (119)	**Brassica napus**, ATP synthase beta subunit.	(M_r): **59676**; pI value: **6.18**
	D2	**Brassica napus**, myrosinase-binding protein (202)	Search not necessary	(M_r): **104268**; pI value: **5.39**
		Brassica napus, probable myrosinase-binding protein (187)	Search not necessary	(M_r): **96853**; pI value: **5.35**

Report from lab of Kathryn Lilley, Cambridge (2002)

	Spot Number	MALDI/Q-ToF identified arabidopsis protein (score)	UK CropNet Brassica Protein DB Result	Molecular weight (kDa); pI value of Arabidopsis protein
S1 MALDI Results	**C1**	**Brassica oleracea** ascorbate peroxidase (175)	Search not necessary	(M_r): **27658**; pI value: **5.58**
	C2	**Brassica oleracea** ascorbate peroxidase (92)	Search not necessary	(M_r): **27543**; pI value: **5.49**
	C3	**Brassica oleracea** ascorbate peroxidase (123)	Search not necessary	(M_r): **27658**; pI value: **5.58**
	D1	**Brassica napus** ferritin precursor (122)	Search not necessary	(M_r): **28151**; pI value: **5.49**
	D6	RUBISCO (vi)	Search not necessary	
	D7	mitochondrial NAD-dependent malate dehydrogenase (80)	**Brassica napus** Malate dehydrogenase, mitochondrial precursor	(M_r): **35782**; pI value: **8.54**
	E1	**Brassica juncea** glutathione-S-transferase (76) (vii)	Search not necessary	(M_r): **15426**; pI value: **5.41**
	E4	cellulose synthase catalytic subunit, putative (75)	***** No hits found ******	(M_r): **82294;** pI value: **6.19**
	E6	putative protein	***** No hits found ******	(M_r): **59952**; pI value: **9.53**
	G1	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase (99)	***** No hits found ******	(M_r): **84304**; pI value: **6.09**
S1 Q-ToF results	**B2**	**Brassica rapa** nucleoside diphosphate kinase 1(81) (viii)	Search not necessary	(M_r): **16442**; pI value: **6.30**

i) Proteins both match the same set of peptides. Second hit suspected to be a poor annotation in the arabidopsis database – not used in further analysis.
ii) Example in which the peptide sequences gave a positive hit from the Brassica protein DB.
iii) *B.napus* is a hybrid species containing the A (*B.rapa*) genome and the C (*B.oleracea*) genome.
iv) Ambiguous hits, possibly the result of poor annotations.
v) Same protein as found on the S3 gel. Both found in the same place on the two gels.
vi) RUBISCO is a well conserved, photosynthetic protein found abundantly in all plants. Hits came up for many plants, therefore no one in particular isolated.
vii) *B.juncea* is a hybrid with the A (*B.rapa*) and B *(B.nigra)* genomes. Isolated protein likely to be a homologue of that in the C (*B.oleracea*) genome.
viii) Protein likely to be a homologue of that in the A (*B.rapa*) genome.